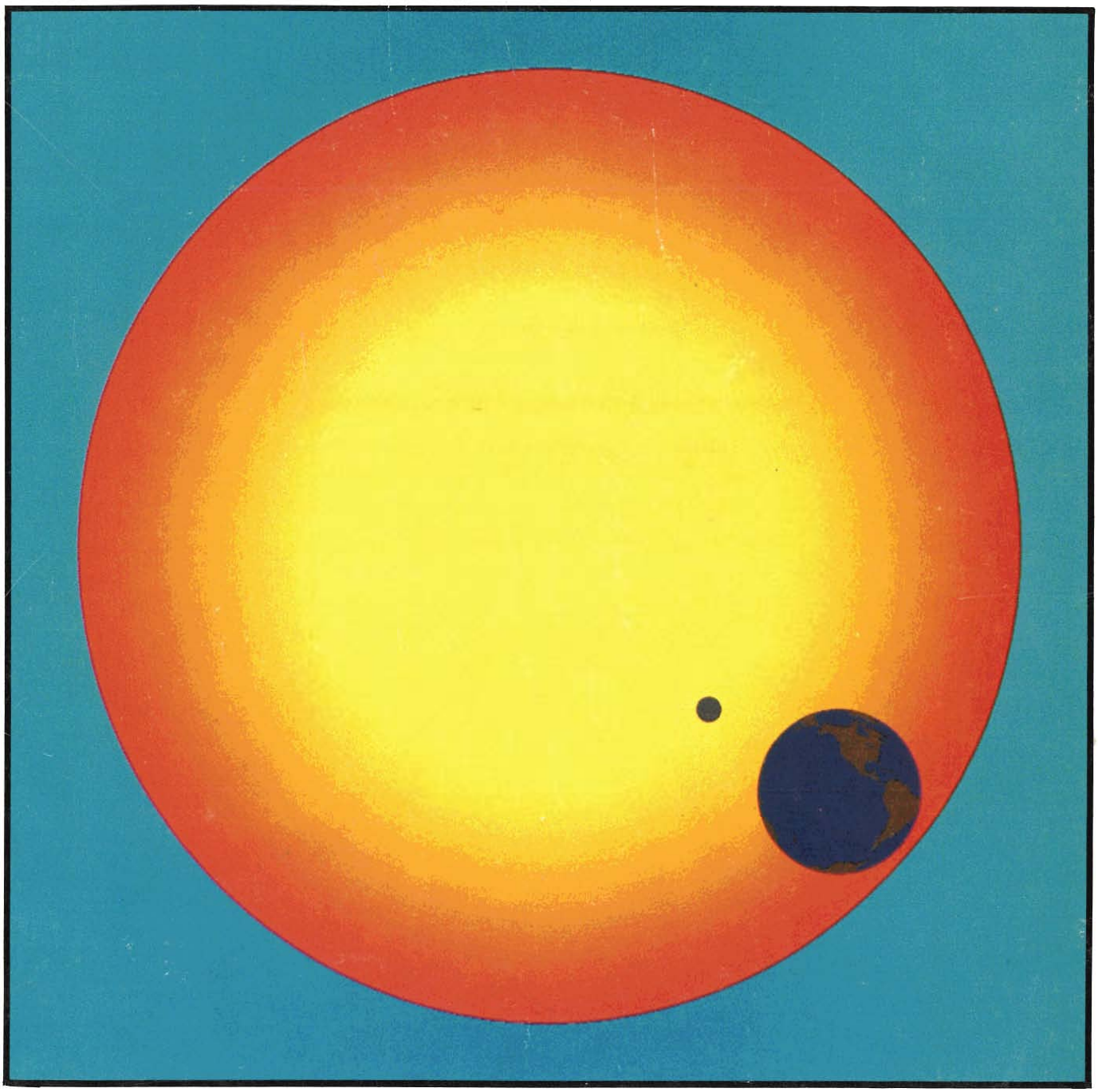
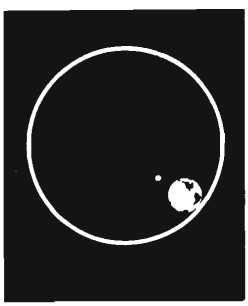
Volume 2, Number 1 WINTER/SPRING 1981

Los Alamos Science

LOS ALAMOS NATIONAL LABORATORY





On the cover: Artist's conception of the sun, earth, and moon before a solar eclipse. This computer-graphic illustration was created using a color-mixing subroutine developed by Rongriego. The subroutine compensates for the nonlinear response characteristics of an FR80 microfilm recorder to provide a perceptibly linear intensity gradation. Saturation problems are eliminated and the additive colors, red, green, and blue may be mixed uniformly to produce a full spectrum of color and intensity.

Overstriking concentric circles of full-intensity red and green delineates the sun's yellow center; the outer regions of the sun were created by selectively removing green with a cosine distribution. Selective blanking protected the emulsion where the earth and moon were to be exposed on the film. These images, generated in separate files, were then projected into the protected areas. The resultant slide was produced using MAPPER, a graphics software package developed at Los Alamos.

Volume 2, Number 1, 1981

Los Alamos Science

LOS ALAMOS NATIONAL LABORATORY

CONTENTS

Los Alamos National Laboratory, an Equal Opportunity Employer/Operated by the University of California for the United States Department of Energy under contract W-7405-ENG-36

Published by the University of California, Los Alamos National Laboratory, Bldg. SM-43, Casa Grande Drive. Second class mailing permit pending at the Los Alamos, New Mexico Post Office.

Address mail to: Los Alamos Science
Los Alamos National Laboratory
P.O. Box 1663, MS 399
Los Alamos, New Mexico 87545

dialogue

HIGH-TEMPERATURE SUPERCONDUCTIVITY

In the first issue of LASL Science, we reported an enormous lattice softening in Ir-Y alloys, which reached a minimum near the eutectic composition. This softening had implications for superconductivity which were thoroughly discussed in our article.

We have now turned our attention toward the metastable alpha prime phase of plutonium, in which up to 1.7 at.% aluminum (normally insoluble in the alpha lattice) has been entrapped by a pressure induced martensitic phase transformation. This entrapment of aluminum into the alpha-plutonium lattice causes profound alterations in the thermal, elastic and mechanical properties of the material. Two of these results—an increased compressibility (measured by sound speed and by mechanical compression tests) and a concomitant anomalous expansion of the lattice—had led previous investigators to believe that the lattice was probably softened through an electronic change in the plutonium 5f electron bonding characteristics.

We felt we could quantitatively investigate (and perhaps further understand) this softening using again the technique of small sample calorimetry discussed in the superconductivity article, (LASL Science 1980). Therefore a 15 mg foil of Pu-1.7 at.% Al was prepared and measured at low temperatures to determine the Debye temperature, a measure of the lattice stiffness. Our results show that the entrapment of 1.7 at.% aluminum into the plutonium alpha lattice reduces the Debye temperature of that lattice by an amazing 30%.

Thus, we have been able to verify in a quantitative manner the degree of softening of the plutonium alpha-phase lattice into which foreign aluminum atoms have been trapped. The question remains however: Why does aluminum weaken the bonds in the alpha-plutonium lattice? We plan to continue our work to dig deeper into this problem.

G. R. Stewart, R. O. Elliott,
and B. T. Matthias

ON THE FIRST ISSUE

As an editor of a number of international scientific journals and someone who has been associated with scientific publishing for a long time, I write to congratulate you on Volume 1, No. 1 of your new publication. It is clearly a much superior publication to any of those put out as yet by the other national laboratories, and will clearly hold its own with any publication in the world market.

This is a major accomplishment for the first issue of any journal, and I can only congratulate you and wish you every success in maintaining this excellent standard in the issues that lie ahead.

Your mix of articles and the obvious careful editorial work that shows in them is both a tribute to the scientific work at Los Alamos and to the care that went into the production of this first issue.

Having said this about the magazine as a whole, let me in particular congratulate Mitchell J. Feigenbaum on his article entitled, "Universal Behavior in Non-Linear Systems." This is by a large measure the clearest exposition of the most recent work leading toward an understanding of turbulence that I have yet found, and I much appreciated it.

D. Allan Bromley, Director
A.W. Wright
Nuclear Structure Laboratory
Yale University

I found the first issue of "Los Alamos Science" both interesting and useful and look forward to future issues. At first I was slightly disturbed by what seemed to be a conflict between "pop feature" layout and almost academic text—as if the illustrations were both out of style and scale. But, hats off to you, the content is good. I like it.

Anthony Tucker
Science Correspondent
The Guardian,
Manchester, England

I wish to congratulate you on the publication of the first issue of "Los Alamos Science."

I enjoyed reading this copy of your excellent magazine and look forward to receiving future issues.

David A. Shirley, Director LBL

Got your first issue of Los Alamos Science. It is great! Congratulations!!

Harold M. Agnew
President, General Atomic Company

Your comments on articles appearing in Los Alamos Science are welcome. Please address c/o The Editor, Los Alamos Science, Mail Stop 399, P. O. Box 1663, Los Alamos, NM 87545.

Inside this Issue

"The endeavor is not, as is sometimes thought, a way of building a solid, indestructible body of immutable truth, fact laid precisely upon fact in the manner of twigs in an ant hill. Science is not like this at all: it keeps changing, shifting, revising, discovering that it was wrong and then heaving itself explosively apart to redesign everything. It is a living thing, a celebration of human fallibility. At its very best, it is rather like an embryo."

Science, a celebration of human fallibility—how refreshing is this phrase in an age when scientists are asked for sure answers and dare not admit the virtue of their uncertainty. But uncertainty is surely a theme of science, and a theme of this issue as well. From nonlinear studies, which are just in their infancy, to the age-old search for fundamental building blocks of nature, from exploring the solar corona to unfolding its influence on the earth's atmosphere, scientists struggle to enlarge the fragile tissue of interconnected ideas. We are gratified to see these efforts encouraged at a time when pragmatic concerns weigh so heavily.

One embryo in this struggle will develop at the Los Alamos Center for Nonlinear Studies. The Center will coordinate research on erratic and complex phenomena that appear in different forms in almost all areas of science and present major stumbling blocks to the success of advanced energy technologies. One of the Center's first projects may be to explore these phenomena as they relate to detonation of high explosives, a topic of continuing study at this Laboratory but, as described in this issue, still more of an art than a science. Another embryo will be fostered by the Los Alamos branch of the Institute of Geophysics and Planetary Physics; here scientists from University of California and Los Alamos National Laboratory can form new collaborations on problems in solar-terrestrial physics.

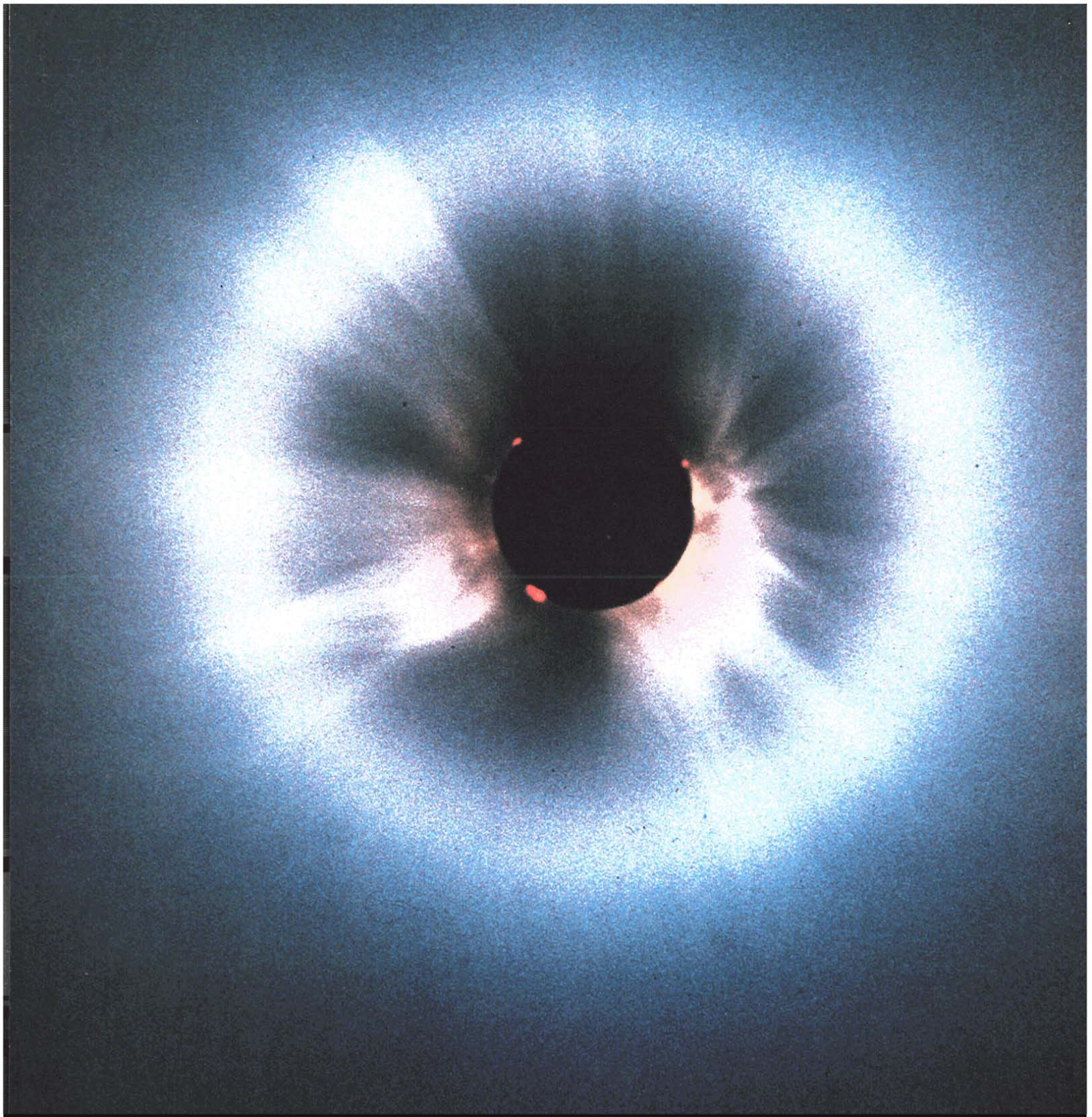
These efforts can flourish only in a supportive environment created by strong leaders who know the nature of the endeavor and all of its risks. Peter Carruthers, former Theoretical Division Leader and currently Senior Fellow of the Laboratory, discusses the subtle role of leadership in science in an unusual and personally revealing interview.

Finally, we focus on neutrino experiments that may reshape our understanding of the universe and the fundamental forces within it.

Altogether, this second issue of *Los Alamos Science* is in itself a celebration, a celebration of the continuing human endeavor we call science.

Necia Grant Cooper

Lewis Thomas, "On the Uncertainty of Science," Harvard Magazine *September-October 1980. Reprinted with permission. Lewis Thomas, author of* The Lives of a Cell *and* The Medusa and the Snail, *is now chancellor of the Memorial Sloan-Kettering Cancer Center in New York.*



by Charles F. Keller and Donald H. Liebenberg

THE SOLAR CORONA AT TOTALITY

Airborne experiments probe the dynamic conditions of the sun's atmosphere during a total solar eclipse.

On February 16, 1980, the moon's shadow raced across our planet from the Atlantic Ocean to China at about 1600 kilometers per hour (Fig. 1). Thousands of spectators and scientists scattered along the shadow's path lifted their heads to glimpse the pearly glow of the solar corona, the sun's outermost atmosphere. Near the midpoint of this path where sun, moon, and earth were aligned for the longest time, a high-altitude United States Air Force jet made its carefully programmed rendezvous with the eclipse shadow. Aboard the aircraft were 18 Los Alamos scientists, observers from Indiana University and Kitt Peak National Observatory, a 12-man Air Force flight and support crew, a Kenyan observer, and two media representatives.

We were intent on recording the conditions in the solar corona during the second highest peak in solar activity in the past century. Five experiments, planned months ahead, were mounted inside the aircraft to make high-resolution measurements of coronal morphology, electron densities, electron temperatures, ion temperatures, and circumsolar dust rings. Our expedition was the seventh in a series of Los Alamos airborne missions that began in 1965 as an outgrowth of the national Test Readiness Program.*

During a total eclipse the aircraft provides near ideal conditions for observing the solar corona. At an altitude of 11 kilometers (36,000 feet) we enjoy freedom from cloud cover and reduced sky brightness and, by flying along the eclipse path, we increase the time of totality by 50% or more. In 1980, we had an unobstructed overhead view of the eclipse for 7 minutes 7 seconds while observers at the main ground site in India saw the event between clouds at a low angle in the afternoon sky for only 2 minutes 9 seconds.

This exceptional high-altitude platform outfitted with state-of-the-art instruments for tracking control and data acquisition (many of which derived from and contributed to the weapons testing program) has enabled Los Alamos to achieve many firsts in observation of coronal temperatures and densities. Such measurements are needed to resolve two major questions in coronal physics. How is the corona heated? And how is the solar wind generated?

We will look at what some of these measurements are and how they fit into study of the corona. We will also describe the technology developed at Los Alamos and how it has been used in carrying out these missions.

Early Coronal Studies

Total solar eclipses have been ob-

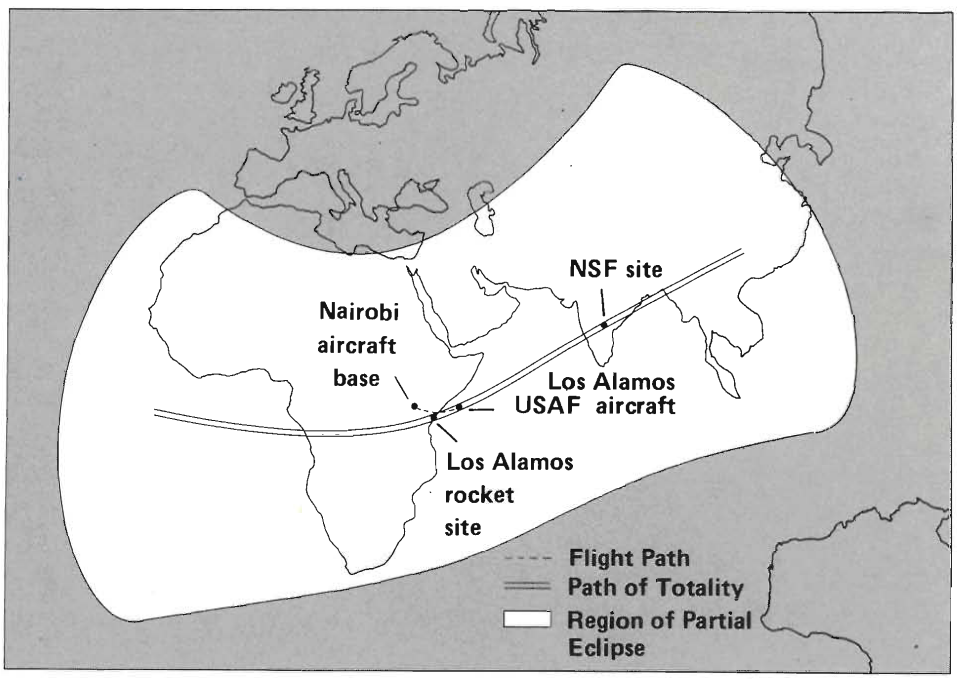


Fig. 1. The path of the moon's shadow first touched the earth in the Atlantic Ocean, then crossed Africa, the Indian Ocean, and India, and finally left the earth over China. Los Alamos observers launched rockets near Malindi, Kenya. The Los Alamos airborne expedition was based in Nairobi and flew to the point of longest totality duration just south of the equator over the Indian Ocean. The main ground-based scientific site was near Hyderabad, India.

*See "In Flight: The Story of Los Alamos Eclipse Missions" in this issue.

CORONAL MORPHOLOGY

Coronal morphology (depicted here out to 3 solar radii) is most likely shaped by the sun's magnetic field, which changes over an 11-year cycle of activity.

During a period of quiescence, coronal structure suggests that open lines of the sun's magnetic dipole field spread out from almost all solar regions except those near the equator, where the corona is brightest and fairly uniform. These regions of open field lines tend to be less bright and may qualify as coronal holes, features characterized by drastically reduced x-ray emission, low particle density, and possibly low temperature. It is known that, during periods of solar inactivity, coronal holes extend from the poles toward the equator and cover much of the sun's surface. Presumably solar wind flows strongly from coronal holes, thus depleting particle and energy densities. Polar plumes extend from the coronal holes at the sun's north and south poles.

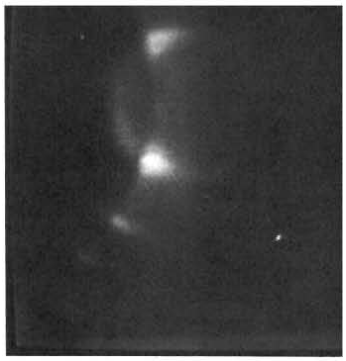
During a period of maximum solar activity, as complex local magnetic fields develop over much of the sun's surface, the open dipole field lines and, with them, coronal holes appear to be restricted primarily to the polar regions. Many distinctive coronal features are formed as a result of this increased activity.

Perhaps the most striking feature is the helmet structure that extends outward into a long radial streamer at about 2 solar radii. The helmet is probably formed by magnetic field loops that rise from active regions of the photosphere. Charged particles are trapped along these field lines, but near 2 solar radii magnetic field strengths decrease sufficiently so that particles can drag the field lines radially outward to form a streamer.

Sunspots, prominences, filaments, and flares also form in active regions of the sun's surface. These regions seem to be confined to well-defined belts that move from high latitudes toward the equator as solar activity declines.

Coronal enhancements and condensations, bright areas near the sun's limb, are also associated with active regions. Their brightness is due to increased particle densities. An enhancement may persist for months, whereas a condensation may appear and disappear over a period of hours or days.

Changes in coronal morphology during the solar cycle are best portrayed by our image-enhanced photographs of the corona in the note "The Changing Corona" ■



served for all of recorded time and probably since humans began to look at the sky, but the origin and nature of the sun's magnificent halo has baffled even the most astute observers. In 1605 Kepler attributed the display to a lunar atmosphere. Others associated it with the atmosphere of the earth or the sun. Not until the 1842 solar eclipse, when a spectacular display of corona and prominences was observed by astronomers in several European countries, was there general agreement on its solar origin.

The corona presented many puzzles, the most important being the absence of emission lines from hydrogen and helium and the presence of the coronal green line, a mysterious emission line that could not be identified with any known element on the sun. The green line was observed with a spectroscope by Harkness and Young in 1869. Originally this and other unidentified spectral lines were attributed to a new element named coronium. But there seemed to be no place for coronium in the periodic table, so the mystery of the spectral lines remained. Then in the late 1930s the lines were identified in the spectrum of Nova Pictoris, an exploding star whose atmosphere was known to be very hot. Finally in 1942, the great Swedish spectroscopist B. Edlen made laboratory measurements of hot plasmas and identified nearly all coronal emission lines as due to "forbidden" atomic transitions from highly ionized elements. The green line was produced by a transition in Fe XIV, an iron atom with 13 of its 26 electrons stripped away.* Other lines were attributed to other ionization states of iron, calcium, and nickel.

Tremendous energy is required to produce such highly ionized states. Also, for atoms to remain highly ionized long enough to deexcite through these forbidden transitions, the density must be very low. Thus the corona is a very hot plasma of ionized hydrogen, helium, and heavier elements with a temperature of a million degrees Kelvin or more (~1 MK) and a density one-trillionth the density of the earth's atmosphere. That the corona should be so hot was quite unexpected because it overlies the relatively cool surface of the sun (photosphere), which is about 6000 K.

What are the dynamics of this hot, rarefied atmosphere? How is it heated? How does it sustain itself when it is bounded below by the cool photosphere and above by the cold vacuum of interstellar space? The answers to these questions still remain quite tentative.

The Solar Wind

In the late 1950s the gross dynamics of the outer corona began to be understood. In 1957 Chapman noted that a simple model of the corona in hydrostatic equilibrium leads to a contradiction. According to such a model, the corona must extend to the earth's orbit and beyond. Its pressure must fall so slowly that at infinity the coronal pressure would be much too high to be balanced by the pressure of interstellar gas.

In 1958 E. Parker solved this problem by introducing the revolutionary and initially unpopular idea that the corona is not static at all, but instead is in a constant state of expansion. His model predicted a supersonic flow of particles out from the corona beginning at a few solar radii. This flow came to be known as the solar wind. Its existence was confirmed in the early 1960s when satellites first began to penetrate beyond the earth's magnetic field and detected particles traveling away from the sun at velocities of 400 kilometers per second. This variable stream of charged particles (mostly electrons and protons) is responsible for the sharp discontinuity of comet tails as they pass near the sun, a phenomenon first noted by L. Biermann in 1951. The solar wind also causes a variety of phenomena as it "blows" past the earth—from geomagnetic storms that interfere with radio communications to possible subtle effects on our weather causing climate variations that appear to be synchronized with the sun's magnetic activity cycles.

It is interesting that until now measurements of solar wind flow, interpreted through solar wind theory, have probably given as good an idea of the average conditions in the inner corona as direct observations. To match solar wind fluxes measured near the earth, these models predict that temperatures in the corona are maximum at about 2 $R_\odot$* and must decrease very slowly thereafter. In particular, if the temperature T falls off with radial distance from the sun as $(r/R_\odot)^{-b}$, then for all $0 \leq b < 1$, supersonic wind flow begins at a critical distance from the solar surface. Values of b may be estimated from observed particle fluxes. Beyond a few solar radii, collisions among particles and photons are so rare that ionization levels are "frozen in." In this collisionless plasma, temperatures refer to an average velocity distribution and do not imply local thermodynamic

Physicists refer to unionized iron as Fe I.

Distances from the center of the sun are commonly measured in units of the solar radius $R_\odot$ (~700,000 km); the limb of the sun is at 1 $R_\odot$

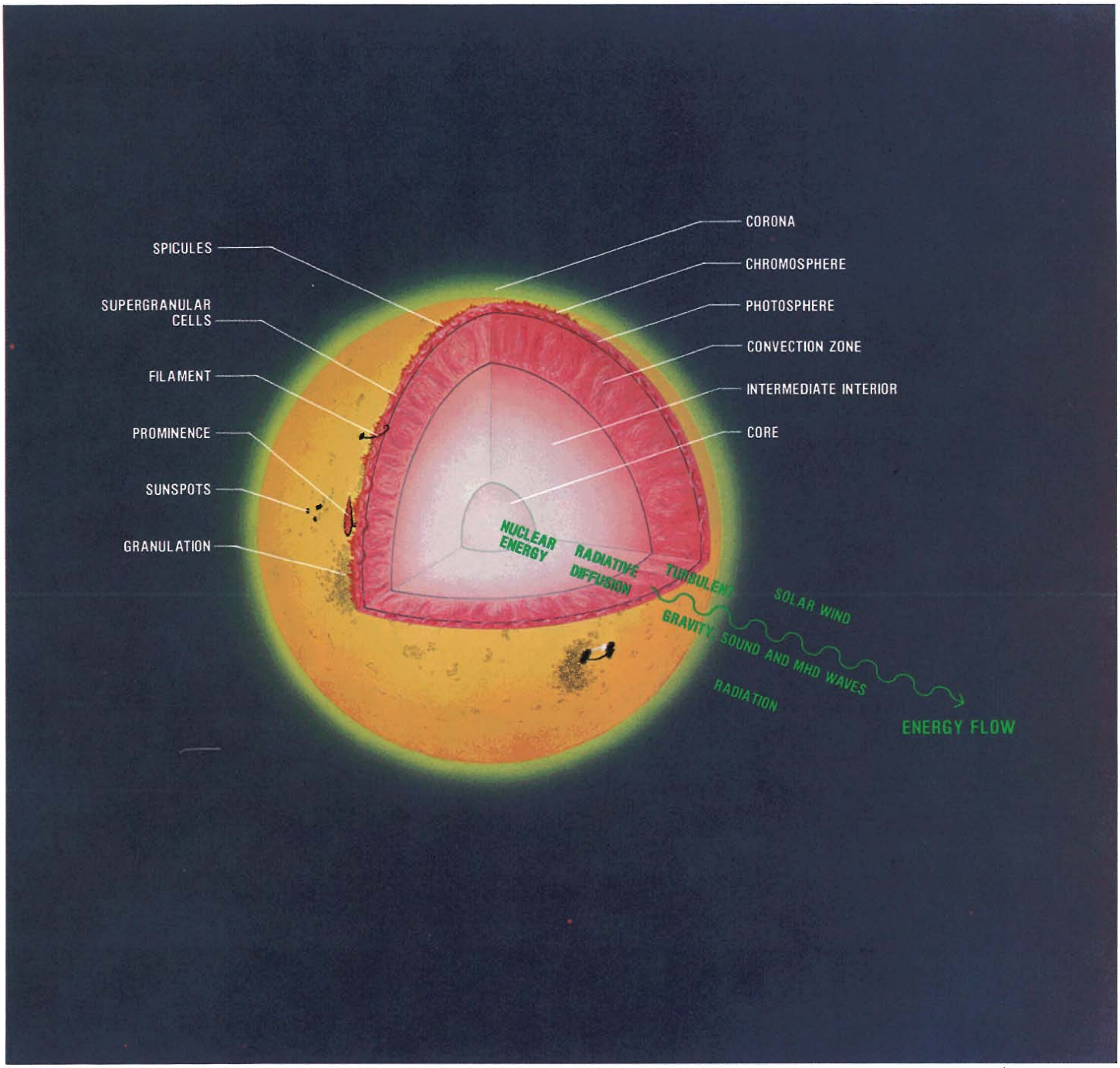
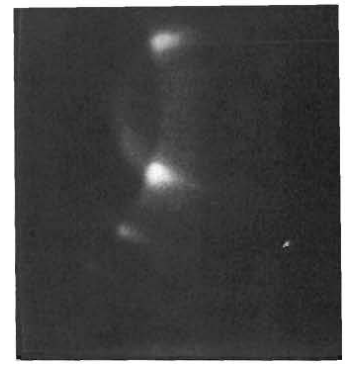


Fig. 2. Schematic depiction of the sun's radial structure and its atmosphere as well as modes of energy efflux. Radiative diffusion transports energy from the sun's core through the turbulent convection zone beginning at about 0.86 $R_{\odot}$. Convection cells in the turbulent layer are visible as a granulation pattern in the photosphere. The cells (about 1000 km in extent) form and disperse with lifetimes of a few minutes, and thereby produce 5-minute oscillations of the granulation layer that may be the source of acoustic waves. Another type of convection pattern called supergranulation is also observed in the photosphere. Large cells (10,000-100,000 km in extent) rise and spread out over a day or so. Supergranulation may produce a network of concentrated magnetic fields at the boundaries of supergranulation cells. This network exerts a strong influence on the upper atmosphere. It certainly shapes the corona and perhaps generates magnetic waves that heat the corona. The network of magnetic field lines is particularly complex in active regions where sunspots, filaments, flares, and prominences are formed.



equilibrium.

Temperature measurements needed to confirm simple fluid models of solar wind flow have been in a state of confusion. Not only are these measurements very difficult to make but the corona is a complex, inhomogeneous structure that changes with time.

During 1973 and 1974, observations of the x-ray and extreme ultraviolet coronal spectrum from the Apollo Telescope Mount aboard Skylab revealed many facts about the inhomogeneous nature of the solar corona. One of the most important for solar wind theory was the existence of coronal holes, regions of low density, perhaps low temperature, and open magnetic field lines that are almost devoid of x-ray emission at the solar surface. At the time of the Skylab observations, when the sun was in a quiet phase, the hole regions extended down toward the equator. The very high velocity component of the solar wind observed during that time evidently emanated from the equatorial coronal hole region where particles stream out at high velocities along open field lines. However, the mechanism for accelerating these particles as well as the contributions to the solar wind from other regions of the corona are still not well known.

High-resolution data from our 1980 mission will help to determine the temperature structure in coronal holes and other coronal features and thus help to explain how the solar wind is generated in each of these areas.

Coronal Heating

Our 1980 data may also shed light on a problem that has puzzled solar physicists since the discovery of the hot corona. By what mechanism is energy from the photosphere deposited in the coronal plasma?

Radiation heating by the enormous flux of visible light emanating from the photosphere is ruled out because the coronal plasma is essentially transparent to visible wavelengths. We must look instead to mechanisms that could be produced by the interaction of mass motions and magnetic fields in the turbulent convection layers of the photosphere (Fig. 2).

Large volumes of gas constantly emerge from the surface, spread out and cool, and finally sink back into deeper layers of the photosphere. This convective energy could generate acoustic waves that propagate into the chromosphere and the extreme lower corona. Alternatively, magnetic waves generated by the interaction of convective cells with magnetic fields might provide a heating mechanism that extends far out into the corona. (Magnetic fields extend outward from the photosphere through the chromosphere and outward with the corona.)

This process might take place in very active regions of the solar surface. In these regions of high magnetic field strengths, we see eruptive prominences, filaments, and sun spots. Coronal condensations overlie these active regions. Perhaps as tangled loops of magnetic fields move around, currents flow that untangle field lines by reconnecting them in simpler configurations, thereby releasing large amounts of magnetic field energy in the form of magnetic waves. A similar but more extreme process is proposed for solar flares. Or perhaps magnetic waves are generated all over the sun's surface.

The details of these phenomena are too complicated to be modeled at present. But there are some questions we can hope to answer through observations. First, how far up in the corona is heat deposited? Is it all deposited in the transition zone between the chromosphere and the corona where the temperature rises from 0.01 to 1 MK in less than 1000 km? Or does the heating mechanism extend out to several solar radii, as suggested by solar wind theory? If so, how is energy transported through the corona?

Is energy distributed equally among electrons, or are protons heated preferentially by nonlinear processes, thus explaining why solar wind protons observed near the earth have higher temperatures? Finally, is solar wind flow a result of temperature gradients alone or are there mechanisms that accelerate the wind without heating it?

If acoustic waves are the major heat source, the lower coronal temperature would be maximum at approximately 1.1 $R_{\odot}$ because acoustic waves would be damped in the chromosphere or extreme lower corona. If magnetic waves are the major source, the temperature maximum would move farther out in the corona, possibly beyond 2 $R_{\odot}$, because magnetic waves would lose their energy much more gradually, possibly out as far as 25 $R_{\odot}$. Direct observations of this temperature maximum would obviously help to determine the mechanism of coronal heating. Also, direct observations of periodic variations in brightness would help to establish the existence of extended wave heating.

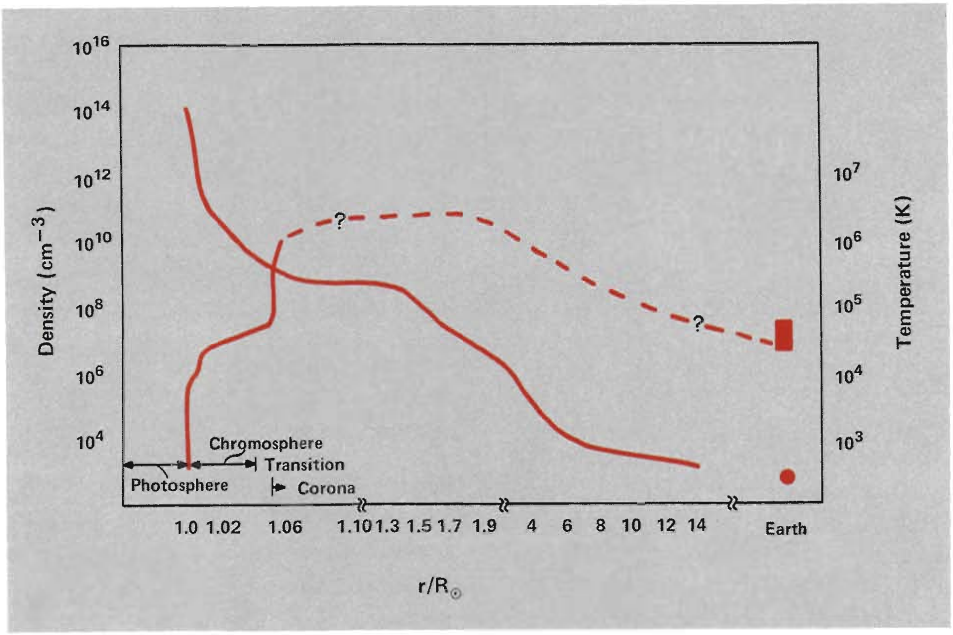


Fig. 3. Approximate variation of particle density and temperature as a function of radial distance. (Note radial distance scale changes.) The steep density decrease above the photosphere and the hundredfold increase in temperature from the base of the chromosphere to the corona are fairly well established but temperatures and gradients in the chromosphere and above are still the subject of research. The exact location of the steep temperature rise in the chromosphere-corona transition region is uncertain, although this region is known to be thin. Temperature and density values in the corona may vary from one feature to another and may be modified by the presence of prominences, flares, and active regions in the photosphere below. Temperature and density of the solar wind at the earth are shown for comparison.

The Problem of Coronal Temperatures

The key parameters for modeling energy transport in the corona are the distribution of electrons and ions in the corona and their temperature, or, more precisely, their velocity distribution.

The average values of the particle distribution are fairly well known, but the temperature distribution is not (Fig. 3). We would like to know not only gross averages but details of temperature distributions in various structural features of the corona since the energy transport will vary from one to another. At present the results from various methods gathered at various times are in disagreement. Although these discrepancies are understandable, given the fact that the various methods for measuring temperature involve different assumptions, the methods are complementary; if done in a coordinated fashion, the separate measurements can be compared to yield more information than any single one can yield alone.

HYDROSTATIC TEMPERATURES. Upper and lower limits on average coronal temperatures can be obtained from electron density measurements. The derivations for both limits assume that electrons and ions are in local thermodynamic equilibrium and electron density equals ion density at each point in the corona. The lower limit estimate is based on a model of the corona in hydrostatic equilibrium (neither expanding nor contracting). The pressure gradient exactly balances the acceleration of gravity,

$$\frac{dP}{dr} = -\rho \frac{g_\odot}{r^2} \quad , \qquad (1)$$

where P is pressure, r is radial distance from the center of the sun in solar radii, $g_\odot$ is acceleration of gravity at the sun's surface, and the material density ρ is directly proportional to the particle density n. Since coronal densities are low, we assume that the ionized gases obey the ideal gas law,

$$P = n k T \quad , \qquad (2)$$

which relates the pressure to the particle number n and temperature T. If we assume that the temperature is constant, we can derive the density variation with radial distance by differentiating Eq. (2) and substituting the result into Eq. (1). We obtain

$$\frac{dn}{n} \propto \frac{1}{T} \frac{dr}{r^2} \quad . \qquad (3)$$

Integrating Eq. (3), we find that a plot of log n versus 1/r is a straight line with a slope proportional to 1/T. Since the corona is nearly all hydrogen we may replace particle density with electron density, $n_e \sim n$. Consequently, if the experimental values of log n_e lie on a straight line, then the so-called hydrostatic temperature may be deduced from its slope.

Actual measurements of log n_e versus 1/r yield a straight line for $r \lesssim 2\ R_\odot$ with a slope corresponding to 1.4 to 1.6 MK. However in equatorial regions of the sun the slope changes between 3 and 5 $R_\odot$ and then beyond 5 $R_\odot$ it has a constant value corresponding to slightly less than 1 MK. This change in slope indicates that the corona does not have a constant temperature. Instead the temperature decreases with distance from the sun, a result consistent with an expanding rather than a static corona. Nevertheless, hydrostatic temperatures are useful as lower limits on average coronal temperature.

TEMPERATURES FOR AN EXPANDING CORONA. To determine temperatures for an expanding corona in, for example, a coronal hole, we add a velocity dependent term to Eq. (1).

$$\frac{dP}{dr} = -\rho \frac{g_\odot}{r^2} - \rho v \frac{dv}{dr} \quad , \qquad (4)$$

where v(r) is the radial expansion velocity. If we assume that the average solar wind mass flux κ is constant and equal to the value observed near the earth, we can obtain a rough estimate for the expansion velocity v from the relation

$$n_e(r)\ v(r)\ A(r) = \kappa\ , \qquad (5)$$

where A is the surface area. We then use the expansion velocity to determine the pressure gradient, dP/dr, from Eq. (4). Assuming that the pressure gradient is due to wave heating of the corona, we determine the temperature gradient from Eq. (2). Then, with a good value for the temperature from an independent measurement, we can use these temperature gradients to determine the variation of temperature with radius. When applied to the 1973 Skylab electron density data, this method yielded a steeply rising temperature curve for the coronal hole above the solar north pole.

OTHER CONSIDERATIONS. This result must be questioned for two reasons. First, since observed densities in coronal holes are very low, collisions may be so infrequent that local thermodynamic equilibrium cannot be assumed. A consequence would be that electron and ion temperatures are not the same. Second, the method assumes that pressure gradients, and thus the acceleration of the solar wind, are due to temperature gradients created by wave heating of the coronal plasma. However magnetic waves moving radially outward may accelerate the solar wind without heating it. Since Eq. (4) does not explicitly separate this nonthermal contribution to the velocity v(r), it yields effective pressures, and through Eq. (2), effective temperatures that are higher than the kinetic pressures and temperatures we are seeking. Temperatures derived from Eqs. (2), (4), and (5) must therefore be interpreted as upper limits. By considering these limits in conjunction with ion temperatures and velocity distributions determined from emission line measurements, it may be possible to separate thermal and nonthermal sources of solar wind acceleration.

EUV ION TEMPERATURES. Ion temperatures have been obtained from emission line measurements by two methods. One method is based on comparing measured and calculated intensities of resonance lines from allowed atomic transitions. These resonance lines are at extreme ultraviolet (EUV) wavelengths (Fig. 4). EUV data from 1 to 1.4 $R_\odot$ have been measured from satellites. Calculations assume an equilibrium model of coronal temperatures and densities and use known atomic parameters to determine the probability for resonance line emission. The assumed temperatures are adjusted to match measured and calculated line intensities. Errors are introduced by the crudeness of the assumed coronal model and by uncertainties in atomic parameters.

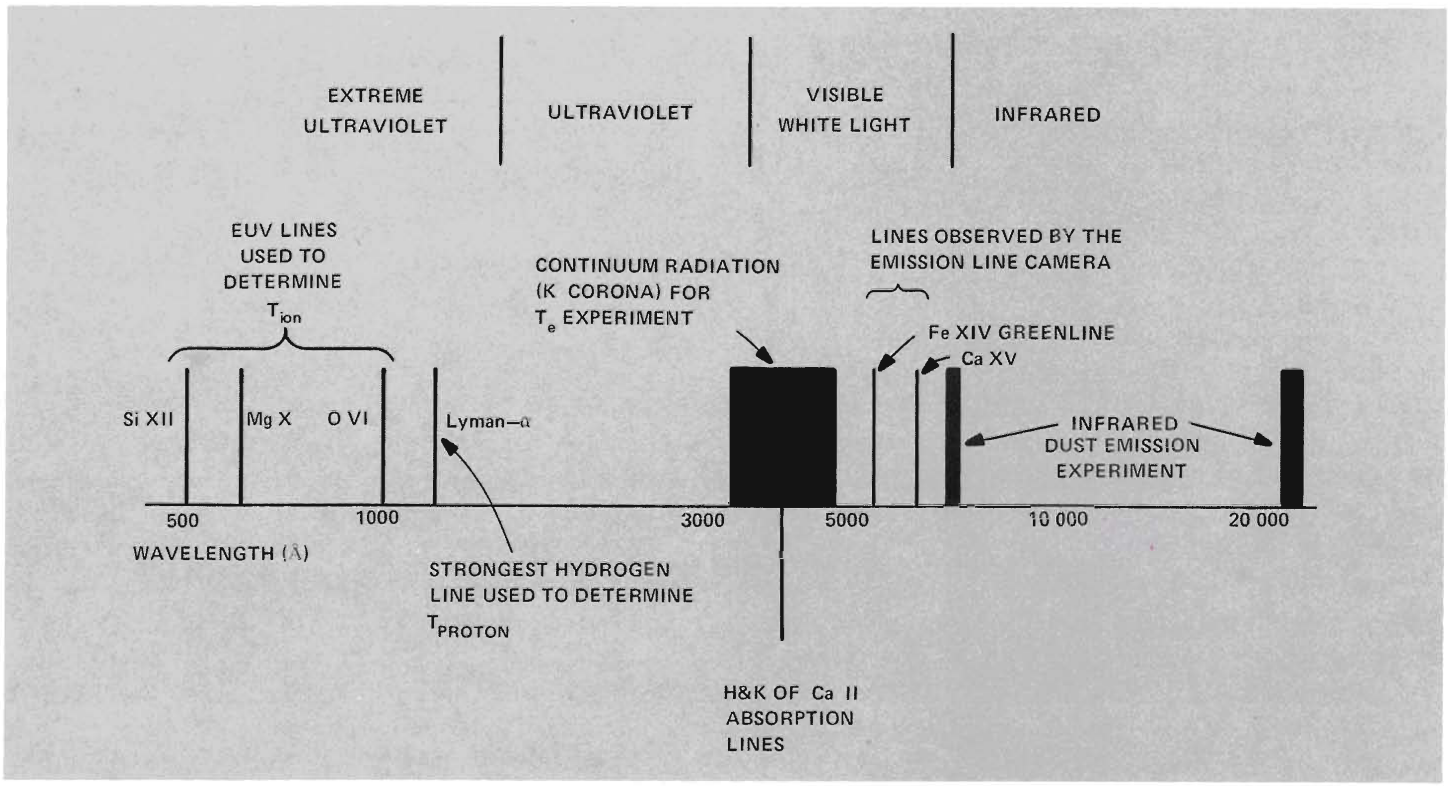
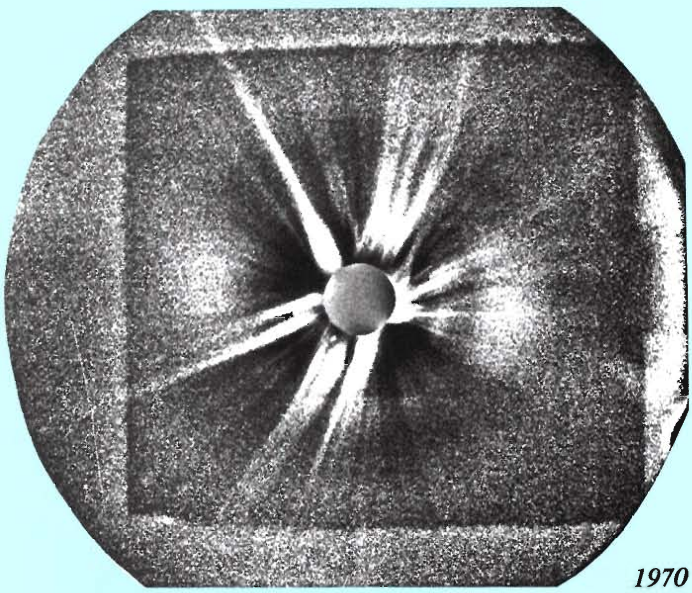
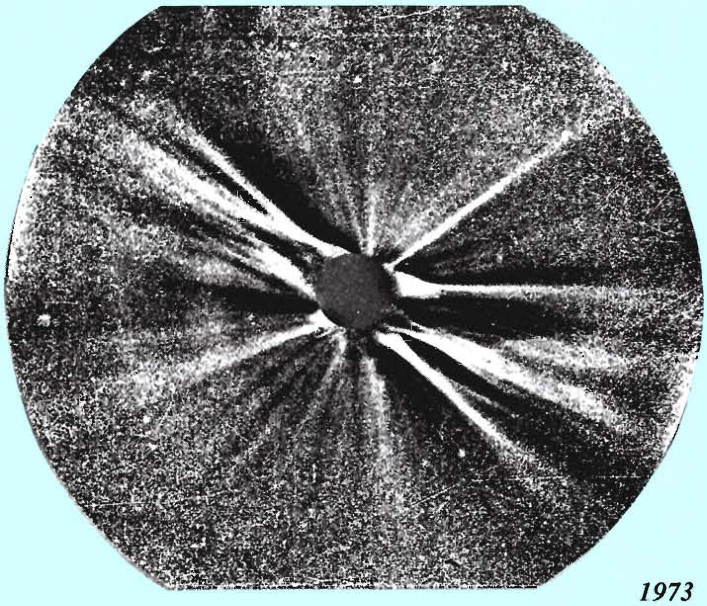


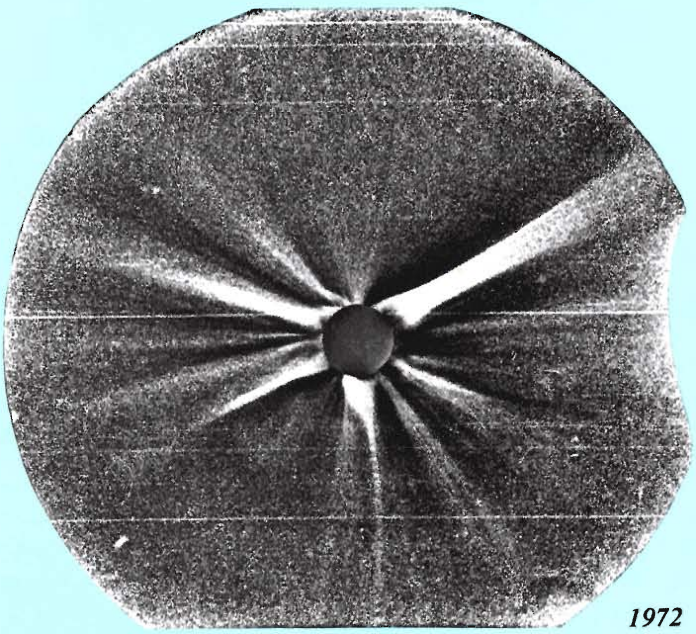
Fig. 4. Relative positions in the electromagnetic spectrum of observations made to determine coronal conditions. We do not attempt here to reproduce the actual coronal spectrum but merely indicate the wavelengths at which observations are made.



1970

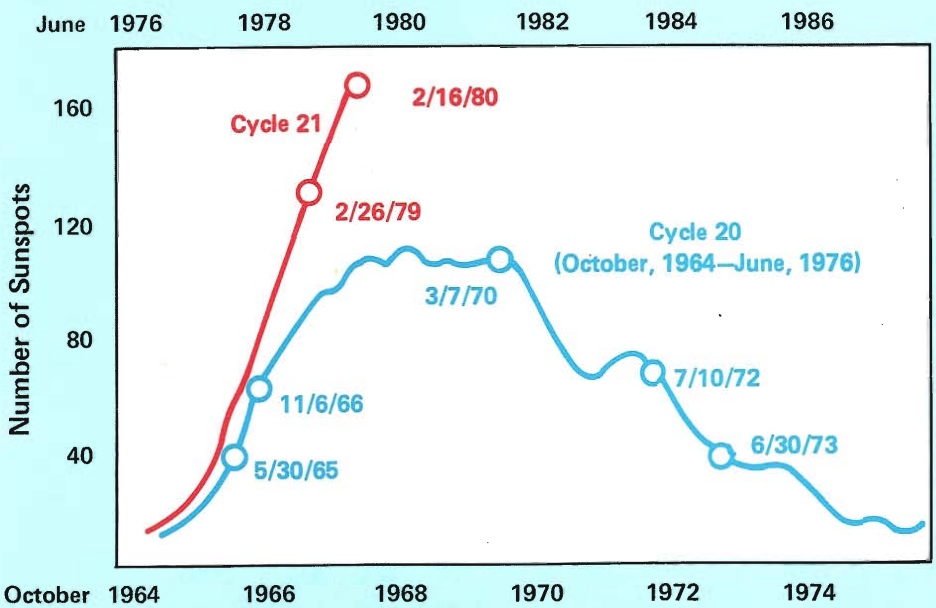


1973



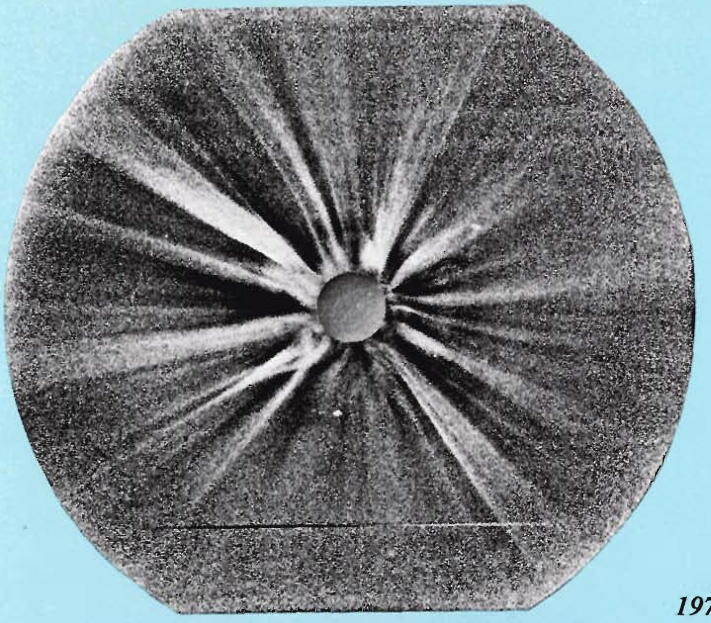
1972

THE CHANGING CORONA

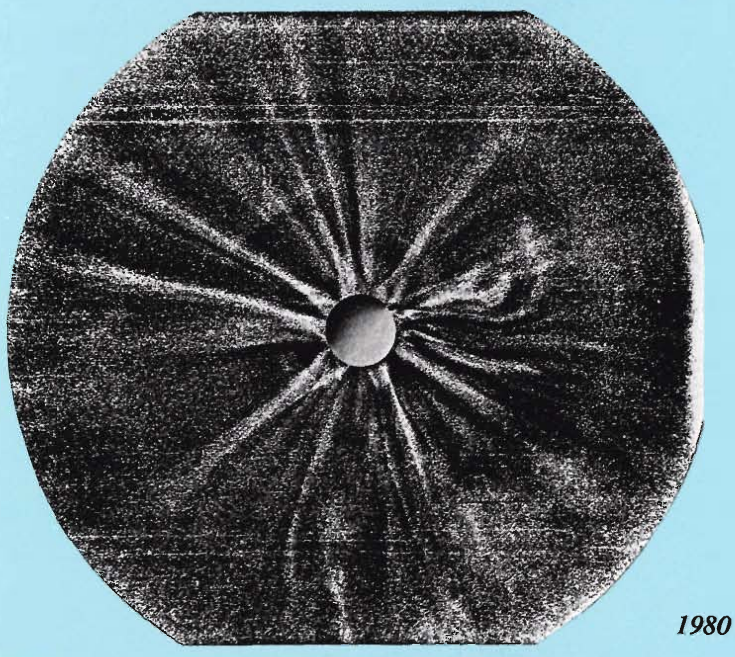


Portraits of the corona during five eclipses display its changing appearance as the sun passes through its cycle of magnetic activity. The eclipse dates are given on the accompanying graph of sunspot activity, which covers the present (red) and previous (blue) solar cycles. Each portrait is a composite of 20 to 30 photographic images taken from the NC-135 jet aircraft with the Laboratory-designed camera-polarimeter.

These portraits of the corona extending from the sun's limb to 12 solar radii are unparalleled. Photographs taken from the ground record only the inner corona to about 4 solar radii. Even Skylab's superb coronagraph can view



1979



1980

only the region between 1.5 and 6 solar radii.

From this unique set of portraits, we can follow the evolution of streamers. The marked variation of the streamer pattern over the sun's 11-year activity cycle gives us a clear picture of the effect of solar activity on coronal structure and solar magnetic field.

The most striking structural change in the corona is the lack of bright streamers over polar regions during periods of low solar activity. At such times (1972 and 1973), polar coronal holes extend to lower solar latitudes. Solar wind flows from coronal holes and apparently bends the sun's magnetic dipole field strongly toward equatorial regions. During periods of high activity (1970, 1979, and 1980), streamers approach and often cover the polar regions and the dipole field is bent only slightly away from the poles.

Individual portraits bear careful study. Close to the sun, streamers are obviously dominated by magnetic fields. We see them twisted and curved in all directions. At larger distances, however, many streamers are seen to bend gradually back toward a radial direction, possibly under the influence of the expanding solar wind. Others remain nonradial even at the farthest distances. The information on streamer dynamics contained in these images awaits theoretical study ■

EMISSION LINE PROFILE ION TEMPERATURES. A somewhat more direct method involves the analysis of measured intensity versus wavelength profiles of emission lines from highly ionized iron and calcium. (Several emission lines have wavelengths in the visible spectrum.) The width of these Doppler-broadened emission lines is a direct measurement of ion kinetic temperatures, provided all the broadening is due to thermal motions. However mass motions and turbulence can also contribute to the broadening.

The Los Alamos line profile data out to 3 $R_{\odot}$ are the most extensive available. Our analysis suggests that ion temperatures deduced directly from the widths of heavy ion emission lines may be larger than kinetic temperatures by as much as 30% because of nonthermal contributions to the line broadening.

Kinetic temperatures may be deduced directly from the width of hydrogen's Lyman-α emission line at 1216 Å because nonthermal contributions to the line broadening are proportionally much less than for heavier mass ions. Since Lyman-α radiation cannot penetrate our atmosphere, such measurements must be made at altitudes of 100 km or greater. Data between 1.5 and 3 $R_{\odot}$ are now available from rocket-borne experiments performed by scientists from Harvard University and the High Altitude Observatory of the National Center for Atmospheric Research.

The results of recent temperature measurements shown in Fig. 5 are in obvious disagreement. In coronal holes (Fig. 5a), we see upper and lower limits on kinetic temperatures based on electron density data. The lower limit is ~1

MK based on the assumption of a static, isothermal corona. The Skylab electron density data, interpreted through a model that accounts for expansion velocities as described above, results in a steeply rising curve that reaches 3.5 MK by 3 $R_\odot$. The other three observations fall within these bounds but are too sparse to allow determination of the position of the temperature maximum. The single EUV temperature and the single Lyman-α proton temperature suggest extended magnetic heating. When the Los Alamos iron ion temperatures are included, the temperature maximum moves below 1.3 $R_\odot$.

Figure 5b shows similar results for relatively quiet regions of the corona (regions that are devoid of obvious streamers but are not coronal holes). Neglecting the iron ion temperature puts the temperature maximum beyond 1.4 $R_\odot$; including it moves the maximum below 1.2 $R_\odot$. Preliminary iron ion temperatures from the Los Alamos 1973 eclipse observations show ever increasing temperatures down to 1.1 $R_\odot$. If the electron density and the EUV results in the lower corona can be believed, then the iron ion results must include a large turbulent contribution to line broadening.

Interpretation of the results shown in Fig. 5 is further complicated by the fact that the data were obtained at different times and at different places in the corona. To resolve the discrepancies, we need simultaneous observations of ion temperatures, electron temperatures, and electron densities. One of the primary goals of the Laboratory's 1980 solar eclipse expedition was to do just this.

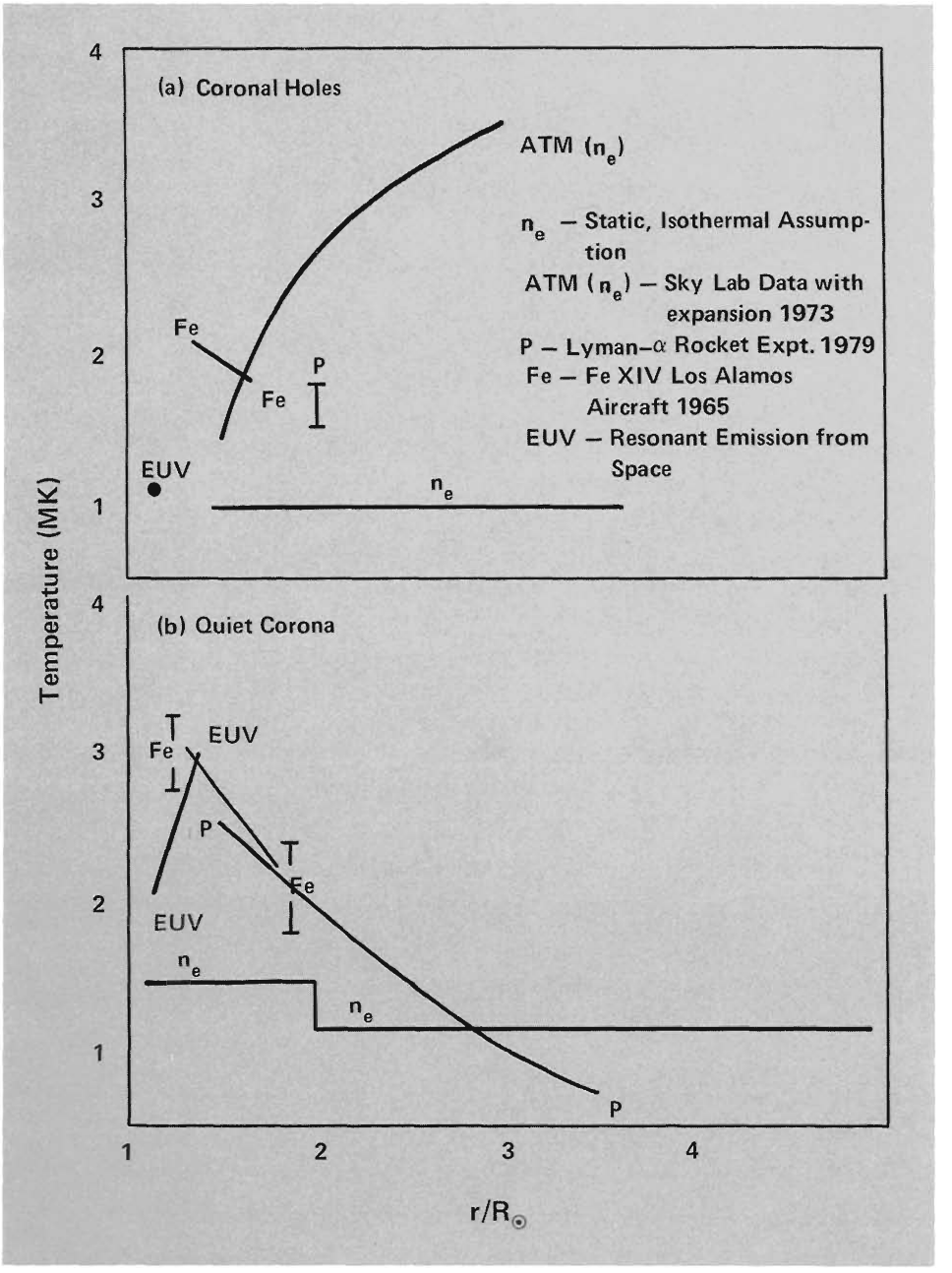


Fig. 5. Plots of coronal temperature versus radial distance for (a) coronal holes and (b) a quiet coronal region. As discussed in the text, data are sparse and, when compared, confusing. The main purpose of the Los Alamos expedition was to attempt to reduce some of the sources of disagreement by making cotemporal and cospatial measurements.

The 1980 Experiments

The sun's magnetic activity follows a well-known 11-year cycle becoming very active, waning to near total inactivity, and then repeating. The amplitude of this variation is never the same, but by 1979 it was obvious that the 1980 maximum would be one of the highest in history and many experiments were planned to observe the solar corona during the eclipse in February, 1980.

The National Aeronautics and Space Administration (NASA) planned to launch the Solar Maximum Mission, a satellite similar to Skylab's observatory. Aboard it would be experiments to determine electron densities and to obtain a crude estimate of ion temperatures from Fe XIV emission intensities. (Unfortunately this experiment was not operating by the time of the eclipse.) NASA also funded two rocket experiments, including one from Los Alamos, to measure proton temperatures from the Lyman-α line. The Naval Research Laboratory's orbiting coronagraph was to make hourly recordings of the white light from the outer corona (3-10 $R_\odot$). The National Science Foundation (NSF) planned to send to India a large contingent of ground-based observers, who would make a variety of measurements.

In planning the Los Alamos airborne

expedition, we were in communication with scientists from NASA projects and some NSF projects. It was apparent that cross-calibration and data comparison would assure maximum confidence in the results.

As discussed above, to answer significant questions about the nature of coronal heating and solar wind acceleration requires simultaneous measurements of the following:

- electron density, n_e
- ion temperature, T_{ion}
- electron temperature, T_e.

All these measurements were attempted by our expedition. All were designed to provide good two-dimensional spatial resolution. We expected to be able to determine all the quantities not only in the relatively uniform portions of the corona, but also within major features such as streamers, condensations, and coronal holes. Table I summarizes the experiments and the information sought.

To determine ion temperatures we measured the Doppler-broadened emission lines from two heavy ions—Fe XIV and Ca XV. We will compare our results with proton temperatures de-

TABLE I

1980 Solar Eclipse Experiments

Experiment	Principal Investigators[a]	Information Obtained or Sought	Resolution
Emission Line Ion Temperature Measurement of coronal Fe XIV and Ca XV emission line profiles with Fabry-Perot interferometer	D. H. Liebenberg E. A. Brown R. N. Kennedy H. S. Murray W. M. Sanders	Temperatures of Fe XIV and Ca XV, variation of emission line intensity with radial distance, and nonthermal contributions to ion temperature from 1 to 3 $R_\odot$	0.005 $R_\odot$ (3500 km)
Electron Temperature Measurement of K coronal spectral intensity with silicon photodiode detector array	M. T. Sandford F. J. Honey R. K. Honeycutt, Indiana University	Electron temperature and coronal spectra at 1.2, 1.6, and 2.0 $R_\odot$	0.067 $R_\odot$ (46,000 km)
Electron Density Measurement of intensity and polarization of coronal light with camera-polarimeter	C. F. Keller J. A. Montoya B. G. Strait	Electron density, K + F corona intensity, and K corona polarization from 1.1 to 5 $R_\odot$ in polar regions and from 1.1 to 12 $R_\odot$ in equatorial regions and image-enhanced photographs of streamers from 1 to 20 $R_\odot$	0.067 $R_\odot$ (46,000 km)
Photography Coronal photography with radially graded filter and internal occulting disc	W. H. Regan C. G. Lilliequist	Detailed coronal structure from 1 to 6 $R_\odot$	0.033 $R_\odot$ (23,000 km)
Infrared Emission of Dust Rings Measurement of coronal infrared emission intensity with InSb detector and charge-injection-device television camera	J. P. Mutschlecner R. R. Brownlee D. N. Hall, Kitt Peak National Observatory	Radial location of dust rings and infrared emission intensity from 2 to 50 $R_\odot$	0.25 $R_\odot$ (175,000 km)

[a] *Unless otherwise indicated, the investigator is affiliated with Los Alamos. Other Laboratory personnel contributing to the 1980 solar eclipse expedition were W. H. Roach (scientific commander), R. A. Jeffries (in charge of liaison with the USAF), D. H. Collins (logistics officer), C. T. Barnett (movie photographer), and J. S. Martinez (still photographer).*

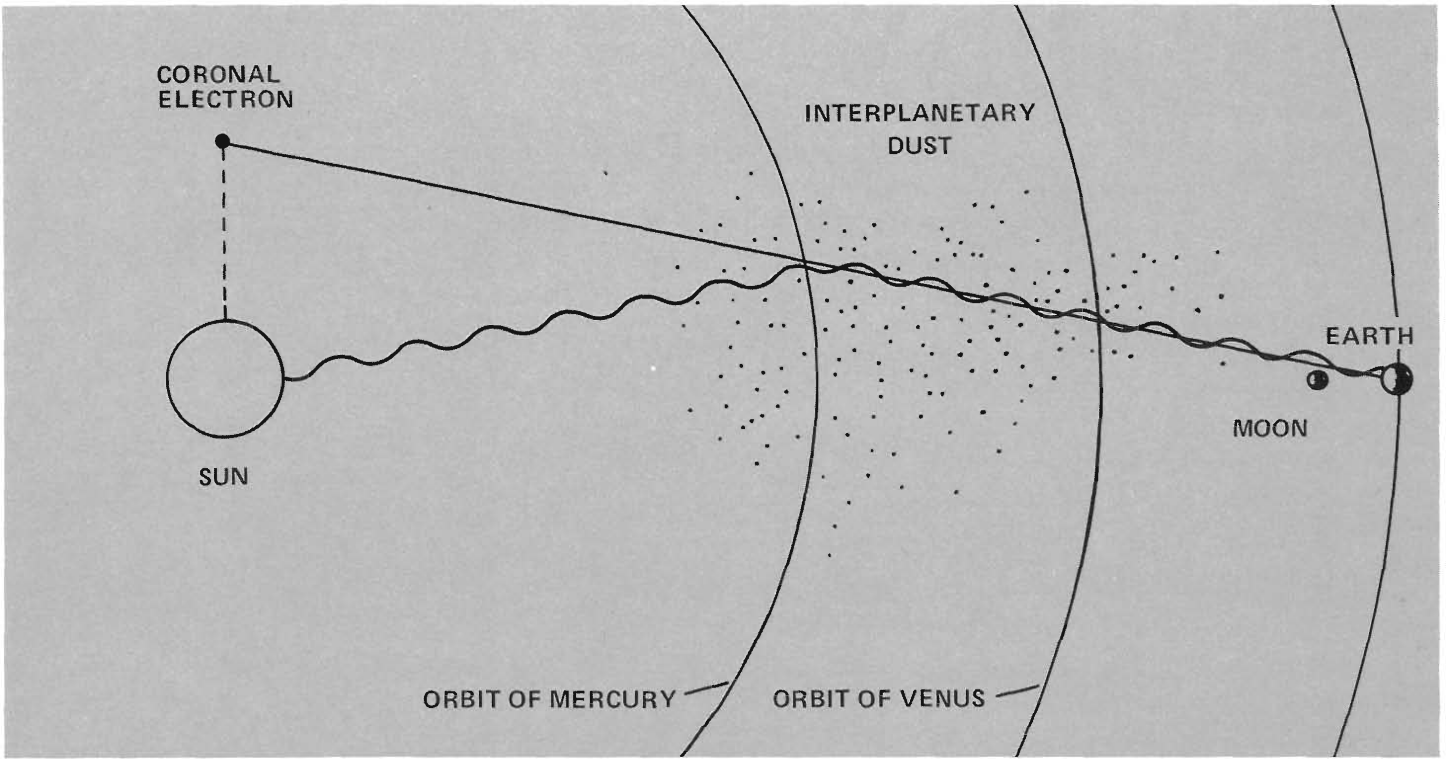


Fig. 6. The white light we observe during a total solar eclipse has two components, light from the photosphere scattered to us by electrons in the corona (straight line) and light from the photosphere scattered to us by interplanetary dust (wavy line).

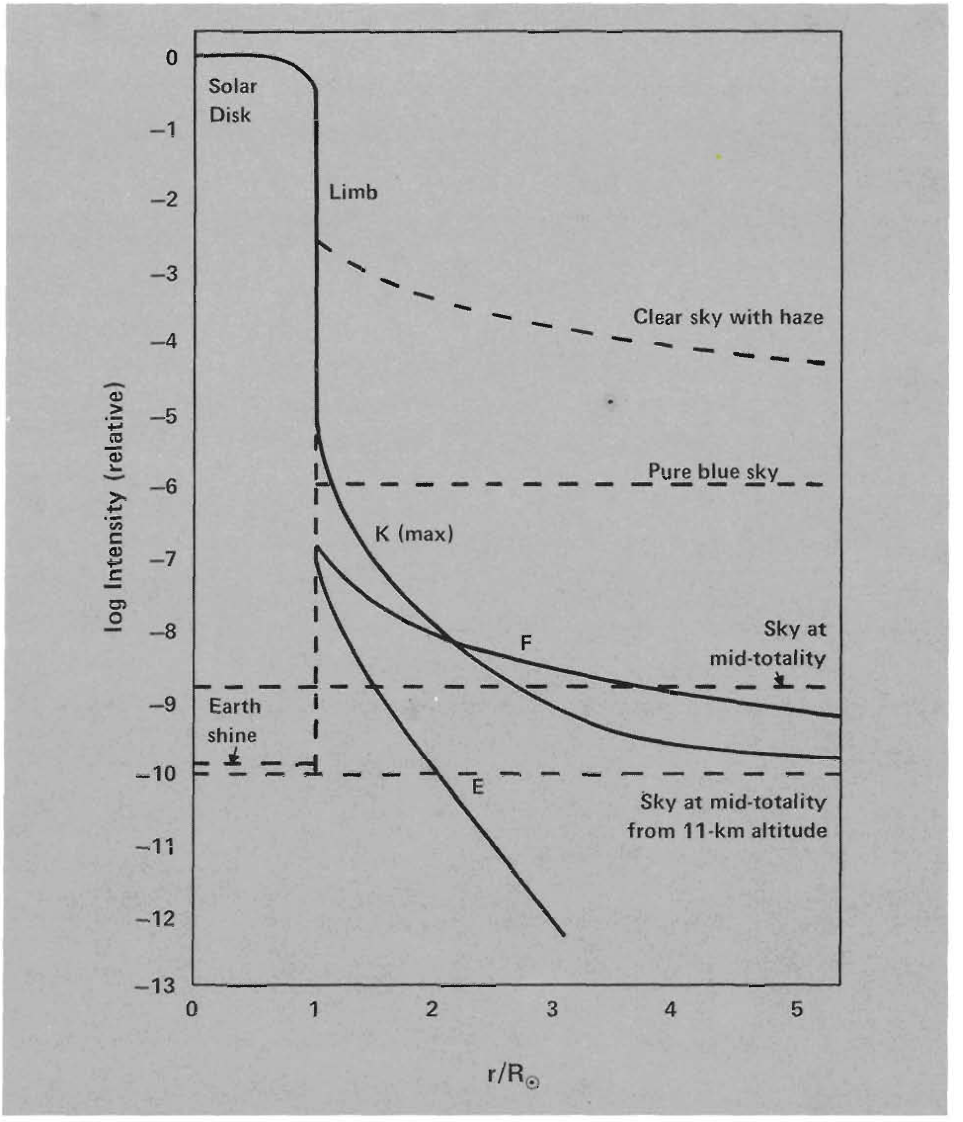


Fig. 7. Relative intensity of components of coronal light as a function of radial distance. The K corona is continuous light from electron scattering. The F corona is light from dust scattering. E is the combined light of emission lines. Relative background intensities under various observation conditions are also shown. Note the tenfold reduction in sky background intensity for high-altitude observations over ground-based observations.

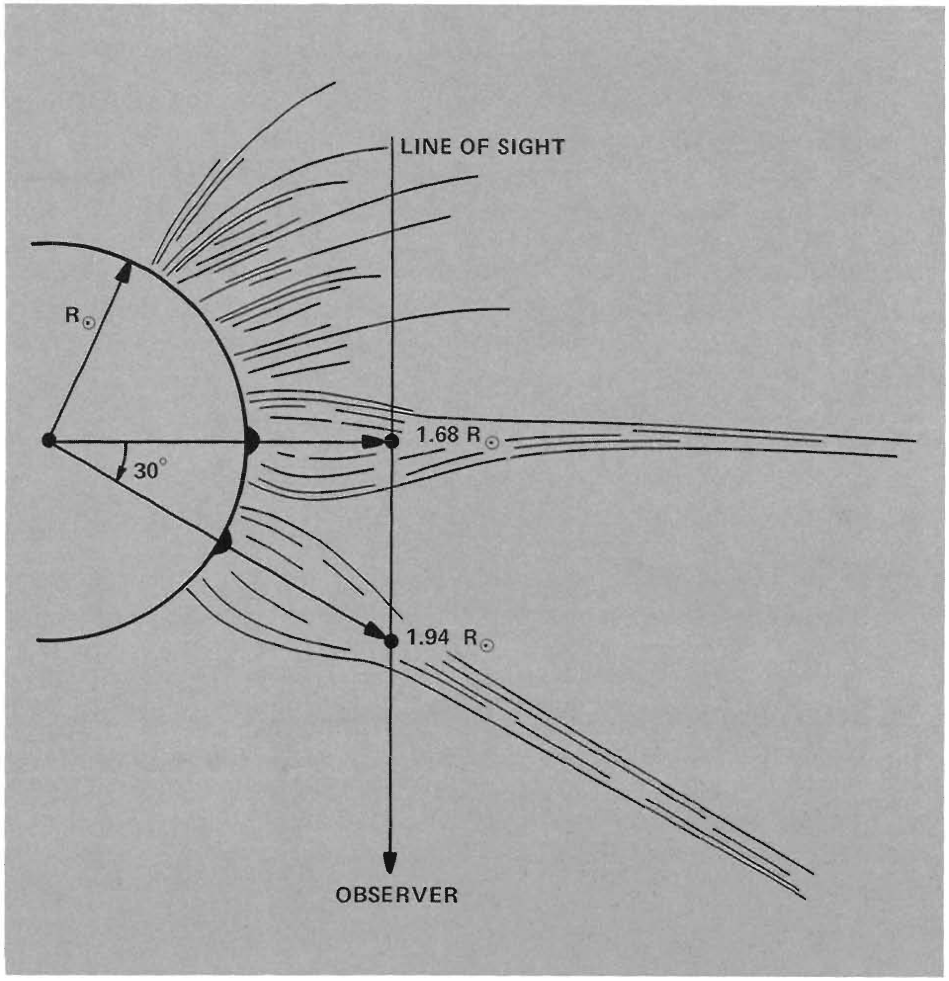


Fig. 8. Eclipse observations of the corona are made along a line of sight, but in many cases single coronal features can be identified reliably. This is because features closer to the limb are more intense and therefore contribute more light to the measured signal. For example, the line of sight shown here intercepts a streamer in the plane of the sky at 1.68 $R_\odot$ and an identical streamer 30° out of the plane of the sky at 1.94 $R_\odot$. The streamer in the plane of the sky contributes about 2.5 times as much white light and about 5.5 times as much emission line intensity to the observed signal as does the streamer out of the plane of the sky.

duced from rocket measurements of Lyman-α emission at several points in the corona beyond 1.5 $R_\odot$ to distinguish thermal from nonthermal contributions to line broadening. We are particularly interested in determining whether the extremely high ion temperatures at the base of the corona deduced from our 1973 results are correct or should be attributed to the effects of turbulence, but we must await data from these low altitudes. We also measured electron densities, recorded high-resolution images of the corona, and attempted to observe electron temperatures by a new, untried method. Finally, we attempted to determine the location of hot dust rings around the sun from their infrared emissions.

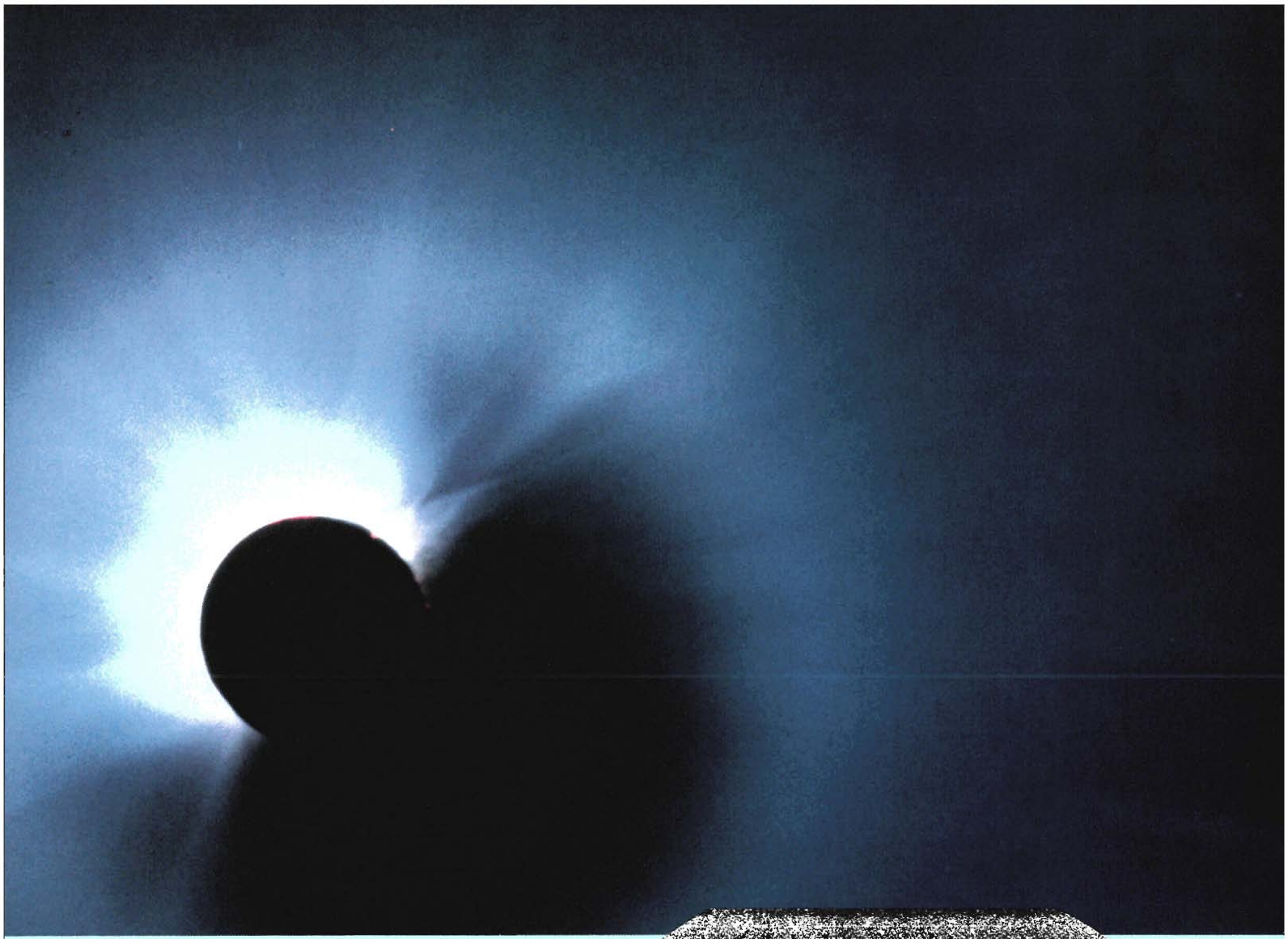
Except for the data loss from a mechanical failure in the electron temperature experiment, we obtained excellent data in all areas. And photographs of the corona taken from the aircraft will enable us to associate our detailed observations with particular spatial structures.

Data reduction and comparison is a long and difficult process that is only beginning. Before we discuss each experiment and give an early estimate of the data it acquired, let us consider some of the general conditions of coronal observation.

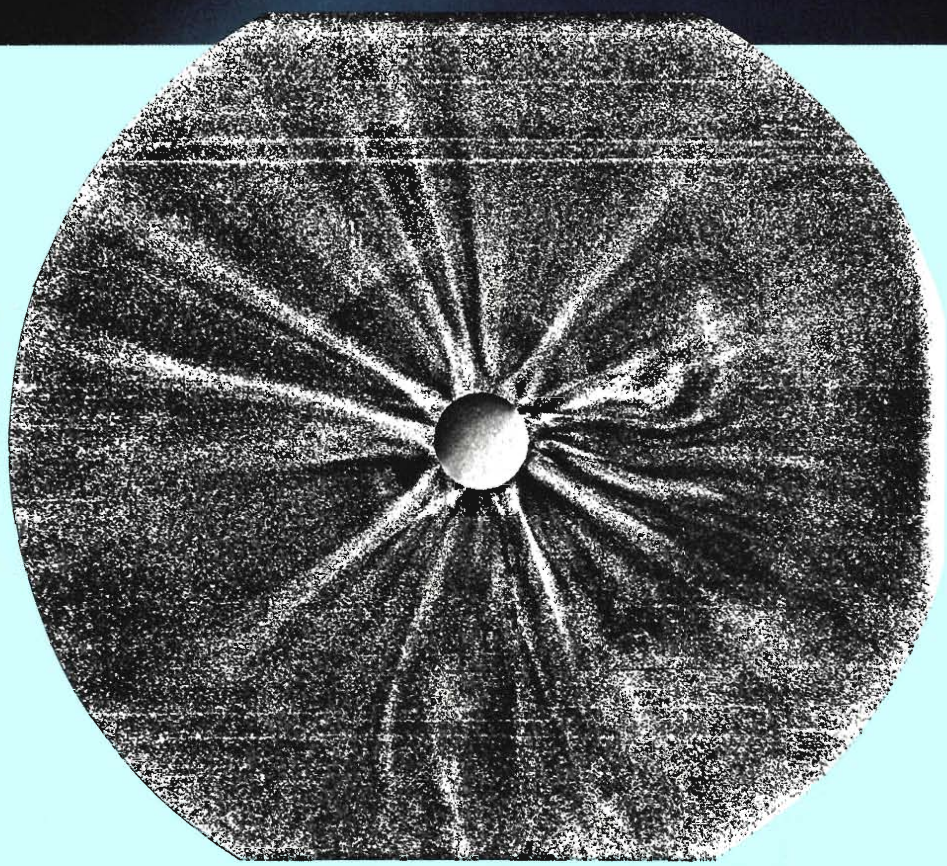
Observing the Corona

The angular sizes of the moon and the sun, as seen by an observer on (or near) the earth are approximately equal. During a total eclipse, this fortuitous circumstance enables us to view the solar corona, a region otherwise obscured by the intensity of the photosphere. Coronal light comes from several sources, the most intense being light from the photosphere that has been scattered by particles in the corona. This white light consists of two components, the K corona—sunlight scattered by electrons—and the F corona—sunlight scattered by dust grains (Fig. 6). Superimposed on these spectral continua are emission lines from highly ionized heavy ions. Figure 7 displays the relative intensities of the various components of coronal light; these intensities are compared with the brightness of the sky under various observational conditions. Note the sharp reduction in sky brightness during a solar eclipse. Although emission line intensities are much less than white light intensities, they are clearly visible above the white light background when observations are limited to very narrow wavelength bands. However, the rapid decrease of emission line intensity with radial distance from the sun makes measurements impossible beyond 3 $R_\odot$ with present instrumentation.

Apart from the signal intensity, the other unalterable limitation on observations is geometry—we can only record two-dimensional information about three-dimensional objects. As shown in Fig. 8, every observed signal is integrated along a line of sight through the corona. If the corona were spherically



A MASS ERUPTION FROM THE SUN



The color photograph of a huge bubble-like structure extending from the sun's limb to 7 solar radii was a most exciting result of our 1980 expedition. The structure is formed by a hydrodynamic eruptive disturbance that ejects large quantities of mass into interplanetary space. The eruption is very clearly visible on the computer-enhanced image processed from 32 digitized photographs. These large and frequent mass eruptions have been recorded since the early 1970s by Naval Research Laboratory satellites and by Skylab, but the bases of the eruptions were always obscured by the oversized occulting discs used in satellite coronagraphs. Thus ours is the first complete record of the phenomenon.

At first glance the structure looks like a tennis racket with its handle in the sun. Closer examination reveals that it is not entirely symmetric about an axis extending radially from the sun. Its polar side is markedly flattened and a density enhancement, which is possibly due to a shock wave ahead of the eruption, is very prominent above and on the equatorial side, but is nearly absent on the polar side. Nearby major streamers that usually extend in a radial direction from the sun are bent toward the disturbance, more so on the equatorial than on the polar side. In addition, the eruption may possibly be bent slightly toward the solar equator. Characteristics similar to these have been reported for eruptions observed from Skylab.

The position of this eruptive disturbance recorded from the aircraft differs markedly from that recorded by the Laboratory's rocket team in Kenya about 15 minutes earlier; from the difference we infer an expansion velocity of about 500 km/s. This value is in agreement with the eruption's absence from photographs taken from the main scientific site in India, where the eclipse occurred about 90 minutes later. During that time, the eruption would have moved beyond range of observation from the ground. If the velocity of expansion is nearly constant during its lifetime, the disturbance must have begun just as the eclipse was arriving on Africa's west coast, about 90 minutes before we saw it. We are at present trying to obtain good photographs taken from Zaire and Tanzania to study the earlier phases of the event.

Such eruptions are apparently quite common. They are estimated to have occurred at least once a day in 1973 during low solar activity and perhaps three times more frequently during this past year's maximum solar activity. Each one typically ejects a mass of about 10^{13} kg and an energy of 10^{24} J. The apparent rate of occurrence would make them responsible for at least 10% of the entire solar mass efflux!

We also obtained camera-polarimeter images and Fe XIV emission line profiles of the eruption. Its Thompson-scattered light is highly polarized so we will be able to determine electron densities and its base appears to have the most intense green line (Fe XIV) emission in the corona. We anticipate that these measurements and our photographs, which together comprise a wealth of interconnected information, will answer several questions about these important sources of solar wind ■

symmetric, this would pose no problem. But in fact its visible brightness varies markedly from one region to another. Thus interpretation of all measurements requires assumptions, based on detailed photographs, about the three-dimensional structure of the corona.

Direct Photography — Imaging the Corona

Good photographic prints of the solar corona are extremely difficult to make because coronal brightness varies by factors of 1000 from 1-4 $R_\odot$ and 10,000 from 1-10 $R_\odot$, whereas prints can display only a factor of 10. We use three methods to reduce this radial brightness gradient.

1. A radially graded filter placed just in front of the film transmits light in precisely the desired amount as a function of distance from the sun.
2. An internal occulting disc in the converging light path within the camera also uniformly reduces the brightness gradient.
3. And computer processing reduces the gradient in the digitized photographs from the electron density experiment.

All three methods were used during the 1980 expedition. Because of a slight misalignment of the mirror tracking system aboard the aircraft, the first two gave good photographic records of only two-thirds of the corona. Nevertheless these pictures give us a remarkably detailed view of the corona during the sun's maximum activity and a particularly fine look at a large mass ejection extending from the base of the corona out to 7 $R_\odot$.

One of these photographs, together

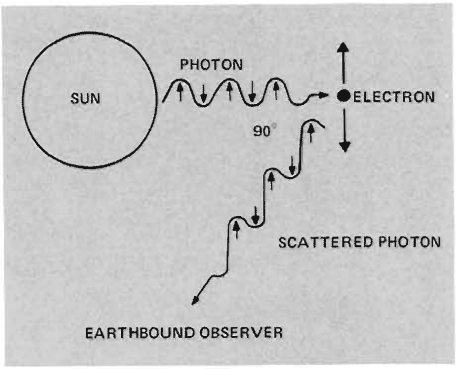


Fig. 9. Schematic of light scattered at 90° by an electron in the plane of the sky to an earthbound observer. This light is 100% linearly polarized along the direction tangent to the sun's radial vector, because its electric field vector (indicated by the arrows under the wavy lines) is always perpendicular to its direction of motion. However, not all the light we observe in our camera-polarimeter has been scattered in the plane of the sky. Each point on our digitized photograph is a sum of contributions that have been scattered from many angles into the line of sight. Consequently, the maximum observable polarization of the K corona from a spherically symmetric corona is only 67%. This theoretical maximum is reduced even further because the corona is not spherically symmetric and, in the inner corona, photospheric light is traveling in nonradial directions before being scattered. Thus the expected polarization as we move toward the sun's limb decreases from 67% to 0%.

with a computer-enhanced image processed from 32 digitized exposures taken with our camera-polarimeter, is shown in the note "A Mass Eruption From the Sun."

Electron Density Experiment

The scattering process that produces the K corona (see Figs. 6 and 7), namely, Thompson scattering of photospheric light by electrons, is independent of wavelength and depends only on electron density. K coronal intensities are thus a direct measure of electron densities. Since we can measure only total white light intensities (K + F corona), we need a method to subtract the unwanted F coronal light.

Much of the K coronal light we observe during an eclipse has been scattered through angles near 90°. The K corona is therefore highly polarized (Fig. 9), whereas the F corona is not. The two components can be separated by measuring both the absolute intensity of the K + F corona (K + F) and the fraction and direction of polarization at each point on digitized images. The measured fractional polarization can be written

$$P_{total} = \frac{P_k K}{K + F} \quad ,$$

where P_k is the fractional polarization of the K corona, and K and (K + F) are the K coronal intensity and total white light intensity, respectively. Data reduction to determine K requires a model of the corona. We assume that the corona has cylindrical symmetry and that its density varies smoothly between polar and equatorial regions. From this model, we calculate P_k and K and compare them with the measured quantities P_{total} and (K + F). Finally we vary electron densities in the model to obtain consistency between the two measured and two calculated quantities.

Photographs of the K + F corona are taken with a camera-polarimeter through plane polarizing filters oriented at three different angles. A fourth photograph taken without any polarizing filter completes a set. During the 1980 eclipse, we made 10 sets of photographs using high-resolution film for the bright inner corona and very fast low-resolution film for the outer corona. This is the first time we have taken special care to get high-resolution data from the inner corona.

The camera and its center-of-mass, three-axis gyrostabilized tracking system were designed and built at Los Alamos (Fig. 10). With this tracking system, motion during a 3-second exposure was less than 20 arc seconds. Our polarization data extend out to 12 $R_\odot$ in the equatorial region of the corona—much

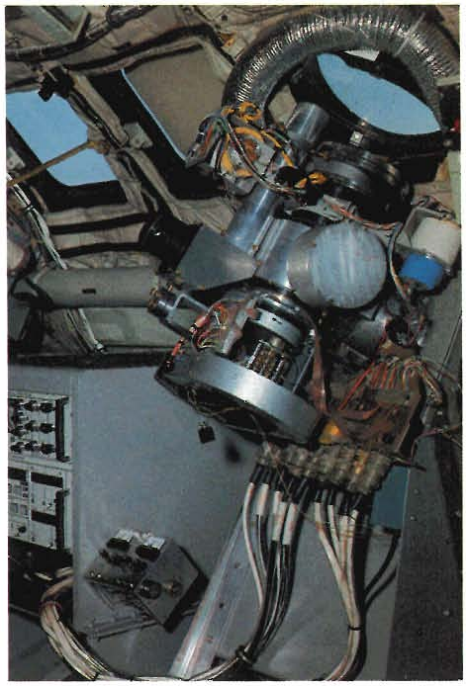


Fig. 10. Los Alamos-designed telescope for photographic polarimetry and electronic control equipment (lower left) mounted in aircraft. This instrument photographs the corona beyond 12 $R_\odot$ and has provided data for determining electron densities and image-enhanced photographs of the outer corona. (See note "Unique Record of the Changing Corona.") The blue cylinder (upper right) is one of three orthogonally mounted gyroscopes obtained from NASA's Apollo program. These provide the inertial references that allow this instrument to point at high accuracy even when the aircraft is in motion.

greater distances than have been achieved from ground observations during an eclipse or from Skylab observations outside of eclipse. In the 1979 and 1980 eclipses we were able to make two exposures out to 20 $R_\odot$.

We have made similar measurements using the same instrument at five times during the sun's 11-year magnetic activity cycle (1970, 1972, 1973, 1979, and 1980). We thus have a very uniform set of data with which to compare variations in coronal structures, brightness, electron density, and material distribution as a function of solar activity. (See note

"The Changing Corona.") For instance, standard observational models of coronal brightness indicate a marked variation between maximum and minimum solar activity. We are unique in being able to verify this variation for the inner corona (out to 3 or 4 $R_\odot$) and to extend it reliably to 12 $R_\odot$.

Results of 1973 Electron Density Experiments

In recent years the 1973 eclipse was the most widely observed both from the ground and in space. Figures 11a and b show plots of absolute intensity of the corona (K + F) as a function of radial distance in the inner and outer corona, respectively. Also included in these plots is the consensus model K + F brightness based on pre-1965 eclipse data. Figure 11a is a comparison of results from five independent observations of the 1973 eclipse: two from the ground, two from space coronagraphs, and one from the Los Alamos airborne expedition. Only the Los Alamos data extend over the entire range shown and beyond.

In comparing our results for the 1973 eclipse with those of Skylab's Apollo Telescope Mount, we were puzzled to find the Skylab values of coronal brightness disturbingly higher than ours. Upon

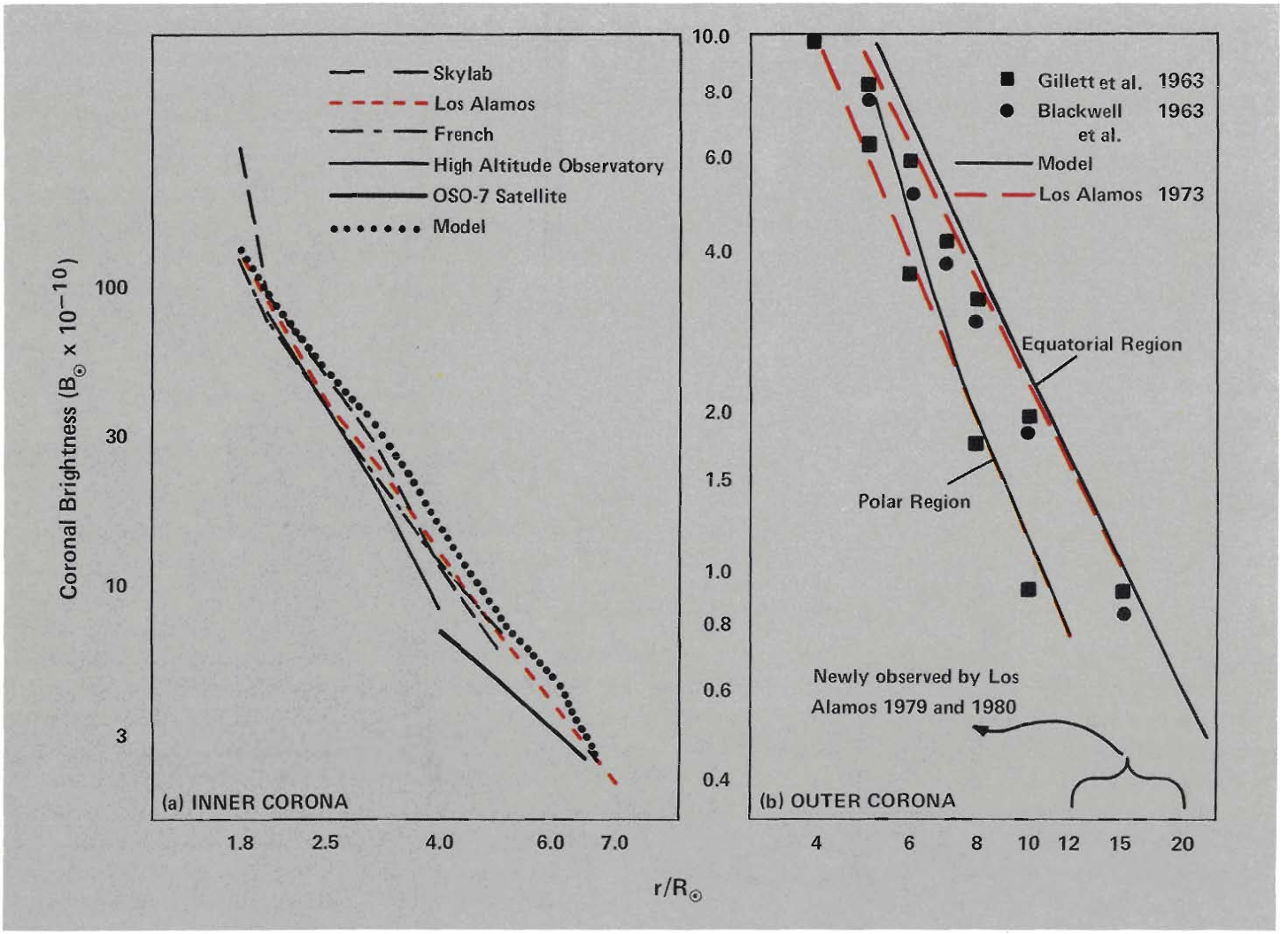


Fig. 11. Various observations of coronal brightness versus intensity for (a) the inner corona and (b) the outer corona. Also shown in both cases is the brightness calculated from a consensus model [D. Blackwell, D. Denhirst, and M. Ingham, "The Zodiacal Light," Advances in Astromony and Astrophysics , 1 (1976)]. Figure 11a is a unique comparison of cospatial, cotemperal observations from two ground-based, one airborne, and two space experiments. Only the Los Alamos data cover the entire range of comparison, and the Los Alamos 1980 data, when reduced, will extend observations to 20 $R_\odot$ for the first time since 1963.

careful examination of Skylab reduction procedure, we discovered an error in absolute calibration that resulted in a 9% reduction of all Skylab values of coronal brightness. This transfers roughly to a similar reduction in all published results prior to 1978. The intensities plotted in Fig. 11a have been corrected for this error.

Figure 11b compares Los Alamos data for the outer corona with the two best recent observations of that region.

Figure 12 shows our 1973 results compared with those from K. Saito's ground-based observations.* Because of sky background brightness, ground-based observations seldom are reliable beyond 5 $R_\odot$ in equatorial regions and 3 $R_\odot$ in polar regions, but the Los Alamos intensity data appear excellent beyond 10 $R_\odot$. Polarization measurements in the equatorial regions are also good to this distance, but over the sun's poles the K corona is so faint that polarization falls below the limit of photographic detectability at 4 $R_\odot$. We measure a fainter corona in polar regions than Saito does and therefore conclude that electron densities in polar coronal holes are lower than previously thought.

In equatorial regions the situation is more complicated. We find coronal intensity to be lower, but we measure higher polarization beyond 4 $R_\odot$. Our measurements suggest that in equatorial regions electron densities beyond 4 $R_\odot$ are higher and, more significantly for solar wind models, do not fall off as fast as we proceed outward from the sun. We also observed this significant result at the 1970 and 1972 eclipses.

We look forward to comparing electron densities from the 1973 eclipse with those of the 1980 eclipse as soon as our 1980 data have been reduced and evaluated.

Preliminary analysis of our 1980 intensity data shows that the corona was about three times brighter than it was in 1973, and that polar regions are as bright as equatorial regions out to $5_\odot$, where the F corona begins to dominate.

*K. Saito, *Annals of the Tokyo Astronomical Observatory* 13, 93 (1972).

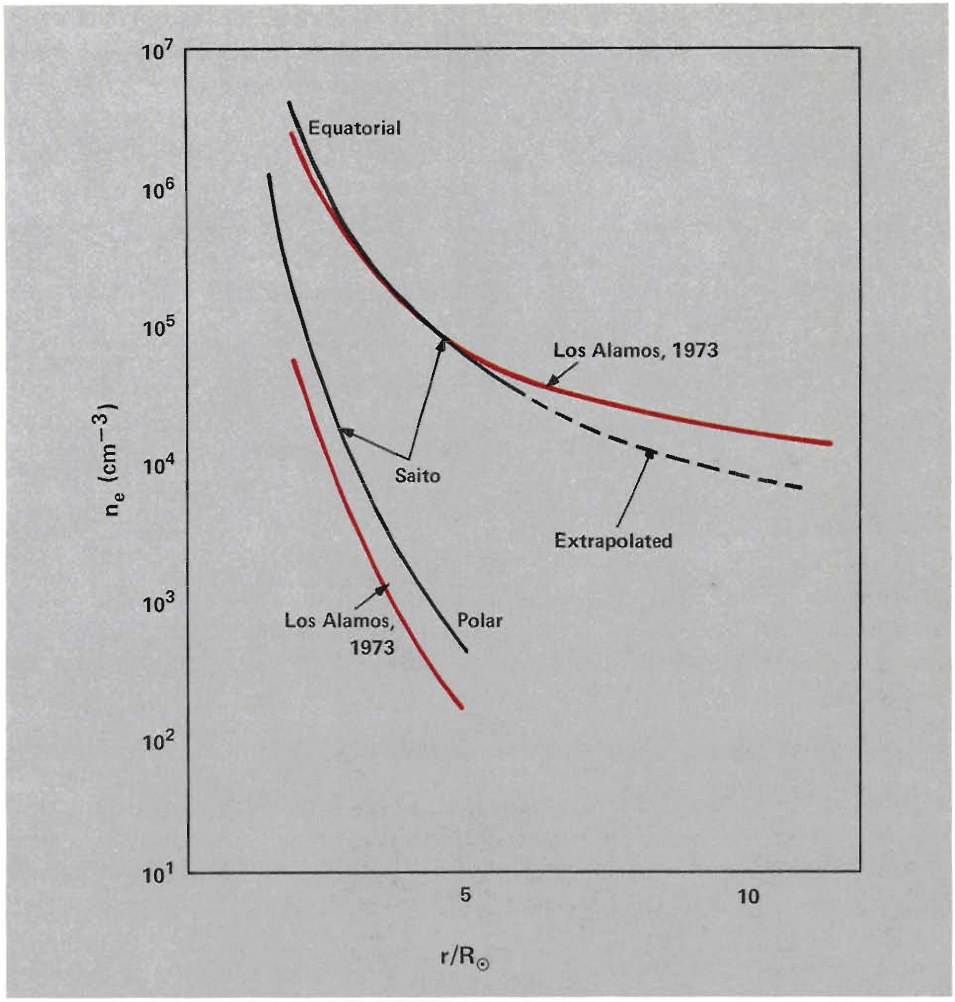


Fig. 12. Electron density as a function of radial distance determined from photographic polarization data gathered during the Los Alamos airborne expedition of 1973. The results of K. Saito, a respected ground-based observer, are shown for comparison. Photographic polarization data for the polar regions are reliable only to about 3-4 $R_\odot$, whereas absolute intensity data (see Fig. 11) are good beyond 10 $R_\odot$. We differ from Saito's results in equatorial regions beyond 4 $R_\odot$ primarily because the reduction in atmospheric contamination realized by our airborne expedition permitted us to acquire good polarization data to 12 $R_\odot$, as compared with the limit of 5 $R_\odot$ for ground-based observations.

Emission Line Experiment

We consider the coronal emission line experiment to be the most important aboard the aircraft because the measured line profiles contain a wealth of information on the state of the coronal plasma. The shape of the line profiles can be analyzed to yield ion temperatures and nonthermal components of the velocity distribution. The variations of line intensity with position, or with time, contain information about the mechanisms that excite the emitting ions. Bright features in or near the plane of the sky (plane through the sun's center perpendicular to the line of sight) easily contribute the major portion of our measured signal along a line of sight because emission line intensities fall off very rapidly with distance from the sun. Therefore we can attribute measured signals to specific coronal features with a fair degree of confidence.

Since 1965, we have made airborne measurements of the Fe XIV green line intensity and wavelength broadening using a Fabry-Perot interferometer to obtain spectral (wavelength) resolution. We chose an interferometer for several reasons. It is a relatively small instrument well suited to the space constraints

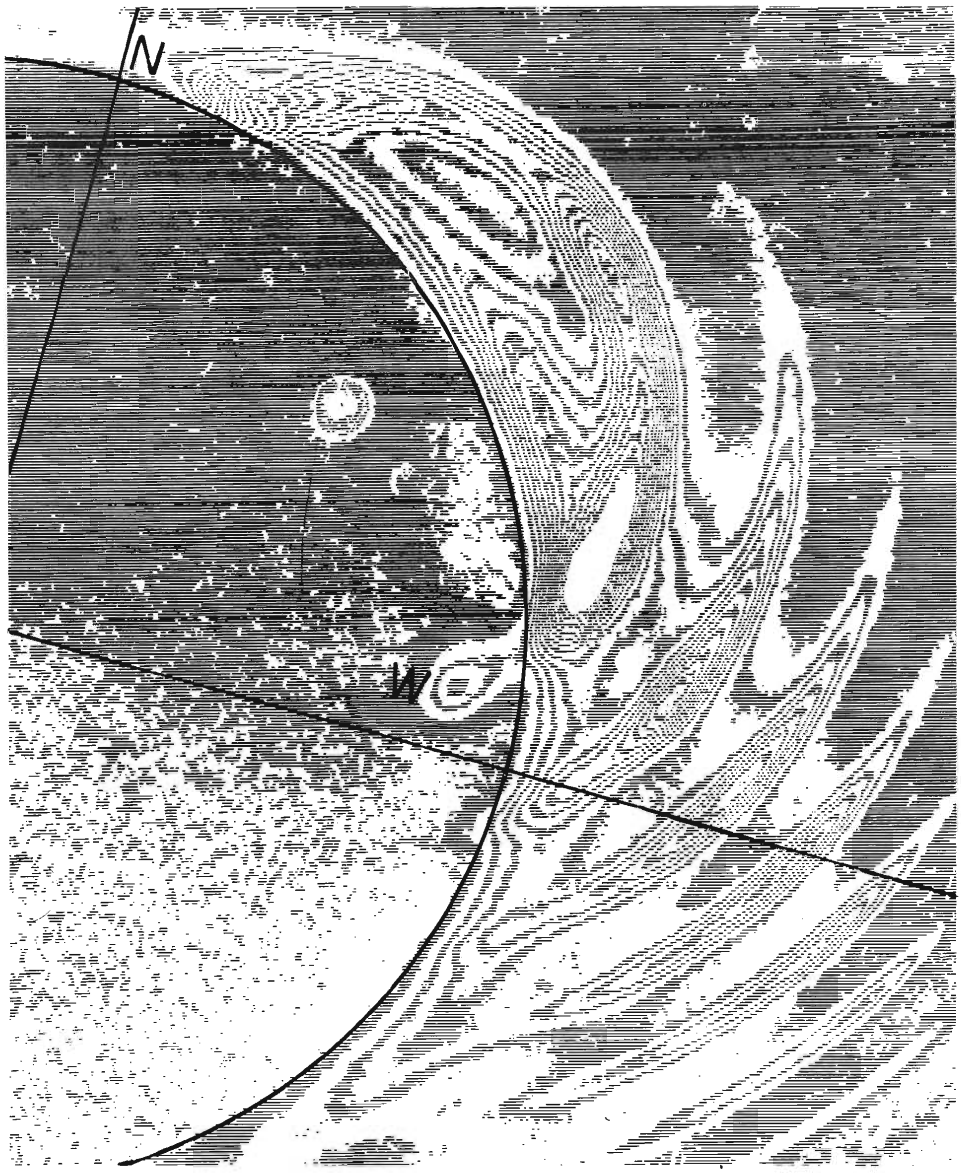


Fig. 13. Isodensity tracing of the interferometer image of the Fe XIV emission line obtained during the 1965 total solar eclipse. The heliocentric coordinates have been overlaid and the lunar limb position is indicated.

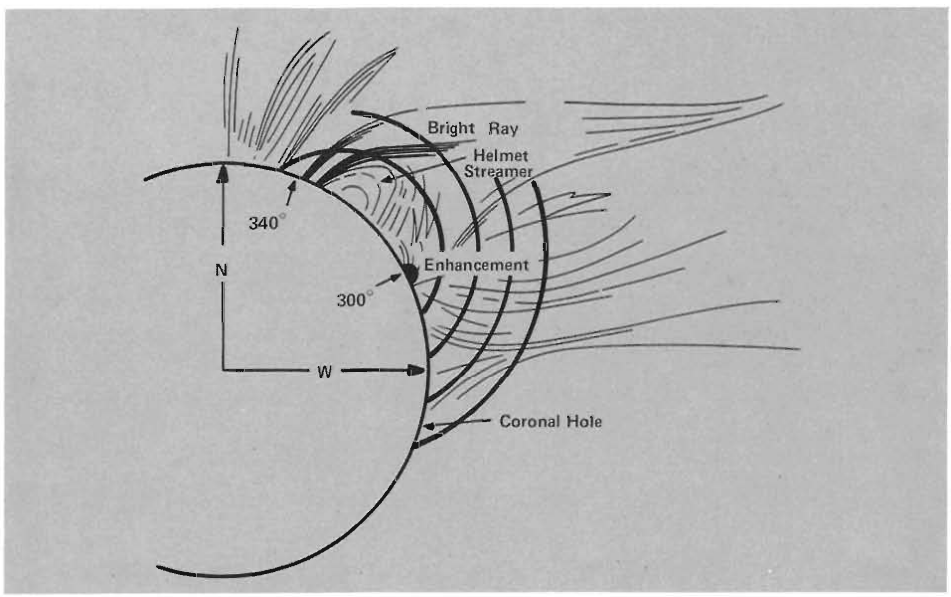


Fig. 14. An artist's sketch of the white light corona in the northwest quadrant during the May 30, 1965, total solar eclipse. The locations of the interferometer fringes are shown together with the position angles of two coronal features.

aboard the aircraft. It also has a high spatial resolution, in part because it preserves a two-dimensional image rather than the one-dimensional images obtained from grating spectrographs and other instruments that pass the incoming light through a slit. Perhaps most important, spectral distortions introduced by the interferometer can be calculated quite accurately and subtracted from the measured line profiles, enabling us to achieve very high spectral resolution.

In 1965 we obtained the first Fe XIV line profiles out to 2 $R_\odot$ from a photographic interferometer image of the corona off the sun's west limb (Fig. 13). A radial trace from the center of the fringe system through each fringe produces an intensity versus wavelength profile emitted by the corresponding region of the corona. Our spatial resolution was limited by tracking to the order of minutes of arc.

As shown in Fig. 14, these fringes passed through a helmet streamer, a coronal hole region, and an enhancement. From these data we estimated ion temperature, temperature gradients, and intensity gradients out to 2 $R_\odot$, and made the first tentative determinations of ion temperatures in helmet streamers and coronal holes at 1.5 and 2.0 $R_\odot$.

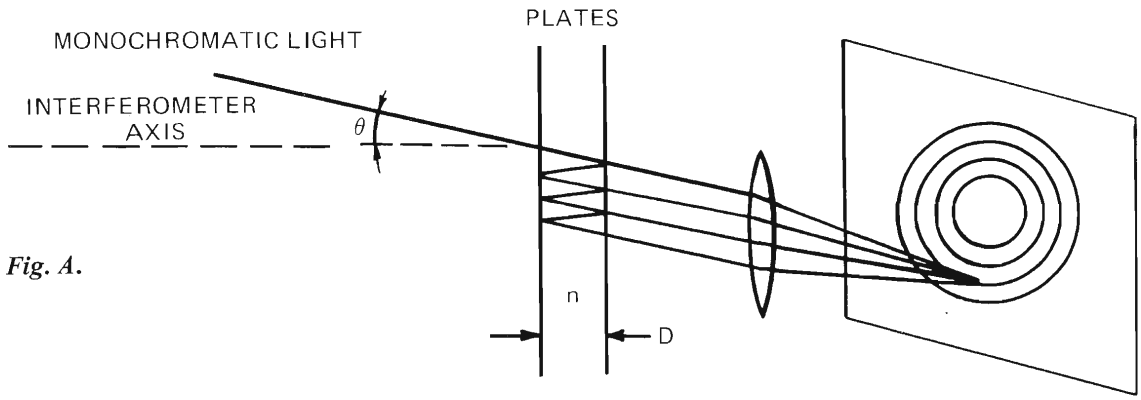


Fig. A.

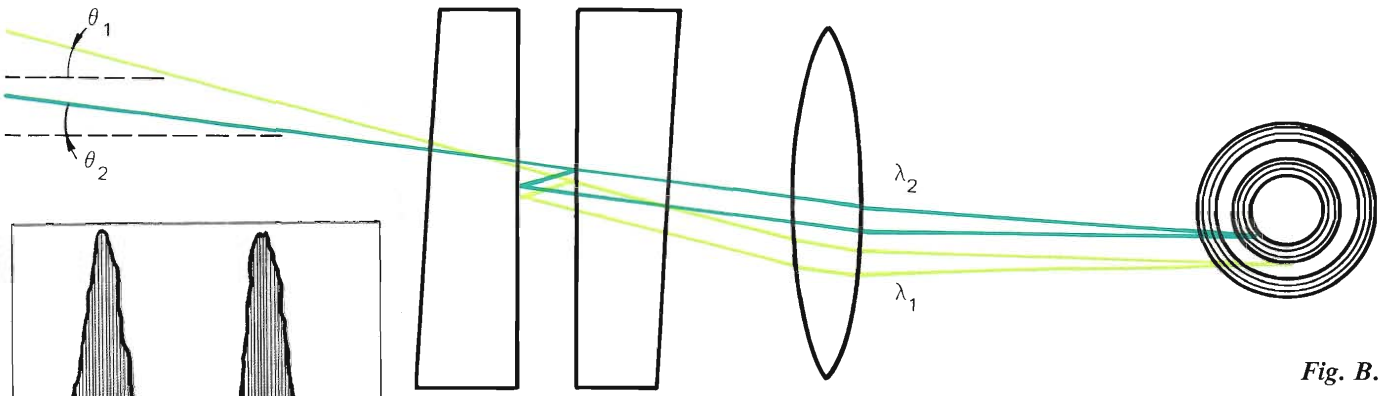


Fig. B.

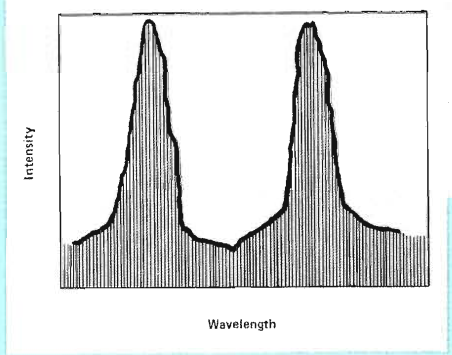


Fig. D.

THE FABRY-PEROT INTERFEROMETER

The heart of the emission line camera is a Fabry-Perot interferometer. It resolves the light from a particular atomic transition occurring in the corona into a high-resolution profile (emission line profile) of intensity versus wavelength.

The basic instrument consists of two accurately plane quartz plates maintained strictly parallel at a constant distance D. The opposing plate surfaces are coated to be highly reflecting. The plates are enclosed in an airtight chamber and the space between them is filled with gas under pressure. The index of refraction n of the gas-filled space can be varied by changing the gas pressure.

A beam of monochromatic light with wavelength λ entering the interferometer at some angle θ to its axis is reflected between the plates many times (Fig. A). The beam's optical path length is $2nD \times \cos\theta$ for a round trip between the reflecting surfaces. When this path length is exactly equal to an integral number m of wavelengths, that is, when

$$m\lambda = 2nD \cos\theta , \qquad (A)$$

then the reflected and incident light waves are exactly in phase and constructive interference (reinforcement) takes place. Successive values of m correspond to successive orders of interference. If the incident angle deviates even slightly from θ, destructive interference quickly reduces the intensity because of the large number of reflections in a highly reflective Fabry-Perot interferometer. Thus monochromatic light from an extended source produces a series of concentric bright rings corresponding to successive orders of interference.

The coronal spectrum includes many widely different wavelengths that can interfere constructively for a given value

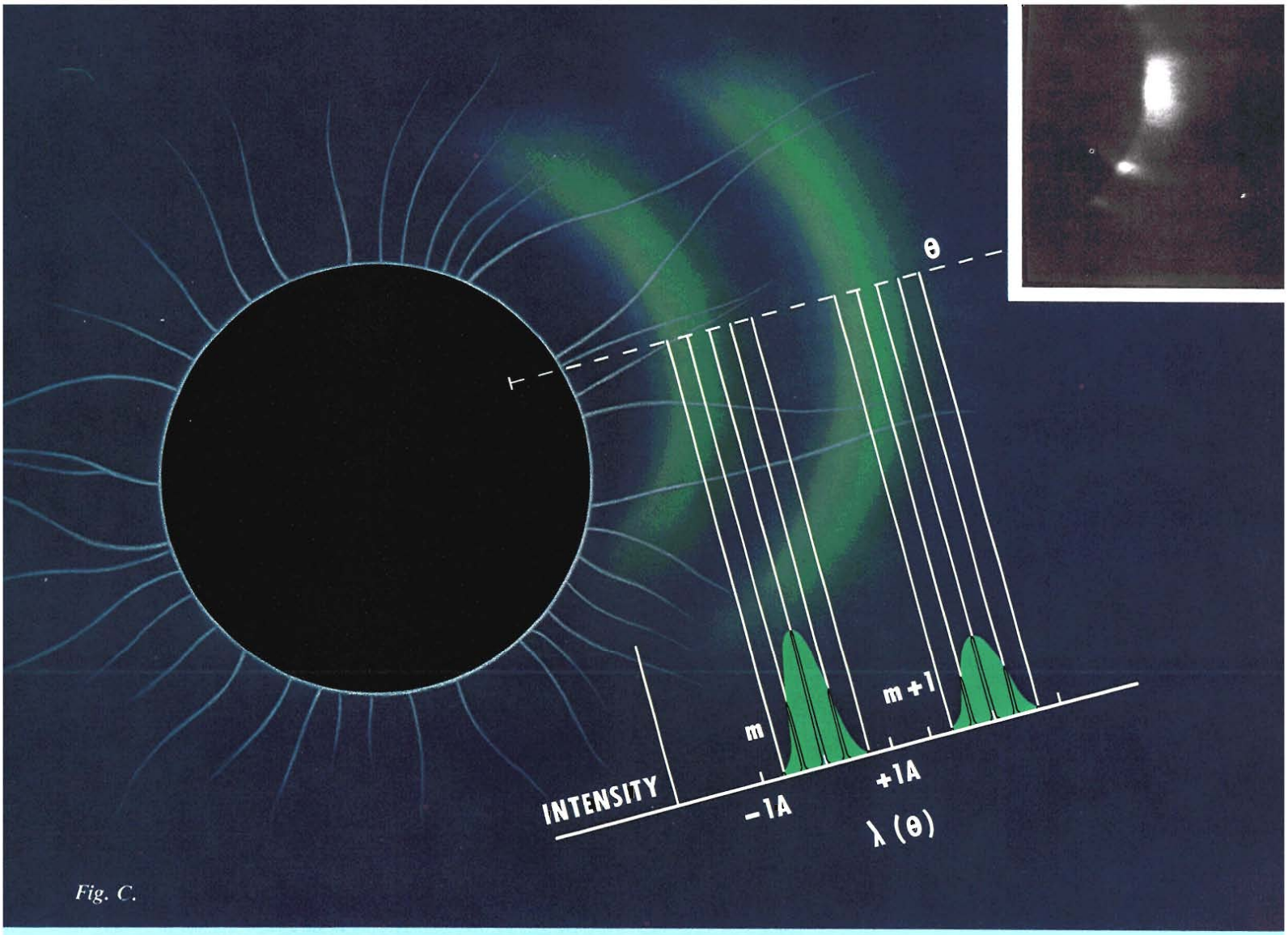


Fig. C.

of 2nD cos θ. This range is restricted by a prefilter to keep the long wavelength end of one order from overlapping the short wavelength end of the next order. The prefilter is centered on the coronal emission line being measured and transmits wavelengths over a 20 Å range.

Coronal emission lines are Doppler-broadened features roughly 1 Å in full width at half-maximum intensity. If an emission line centered at λ_1 constructively interferes for an incident angle θ_1, then another wavelength in the profile, say λ_2, entering the interferometer at a slightly different incident angle θ_2 will also interfere constructively (Fig. B). Thus successive wavelengths in the line profile are imaged by the interferometer as adjacent rings at successive values of θ, and the result is a broad ring where the profile of intensity versus λ is imaged as a profile of intensity versus θ. One carefully makes a trace radially across a ring to extract intensity versus θ, which is easily converted by using Eq. (A) to intensity versus λ (Fig. C). Note that these profiles are deduced for an extended region of the corona corresponding to less than the width of the broad interferometer rings.

The partial rings produced during the 1965 eclipse by a segment of the corona are shown in Fig. 13 of the main text. In 1965 we had time to take data only at a fixed value of 2nD, so the only regions of the corona that could be studied were those corresponding to the bright rings in Fig. 13.

During the 1980 eclipse, our rapid data-acquisition system gave us time to vary the gas pressure inside the interferometer, and thus to change n and cause the rings to traverse the entire coronal image by changing the angle for constructive interference. By thumbing the corner of this journal, on which are printed our 1980 photographs, the reader can observe the movement of the rings caused by the pressure scan. From this series of images corresponding to different values of n, we can plot the intensity change at a single spatial point as the gas pressure is varied with time. This plot of intensity versus n is converted to intensity versus λ.

Our video recording system stored an array of 250,000 spatial data points from each image; during a pressure scan of one order, 120 images were recorded. Each point can produce such a profile. One such profile is shown schematically in Fig. D. The line profile is repeated as the pressure scan moves the ring corresponding to the succeeding order across the point being plotted. The spatial resolution of a line profile deduced from a single point is 4-5 arc seconds, which is an order of magnitude better than can be obtained from a radial trace ■

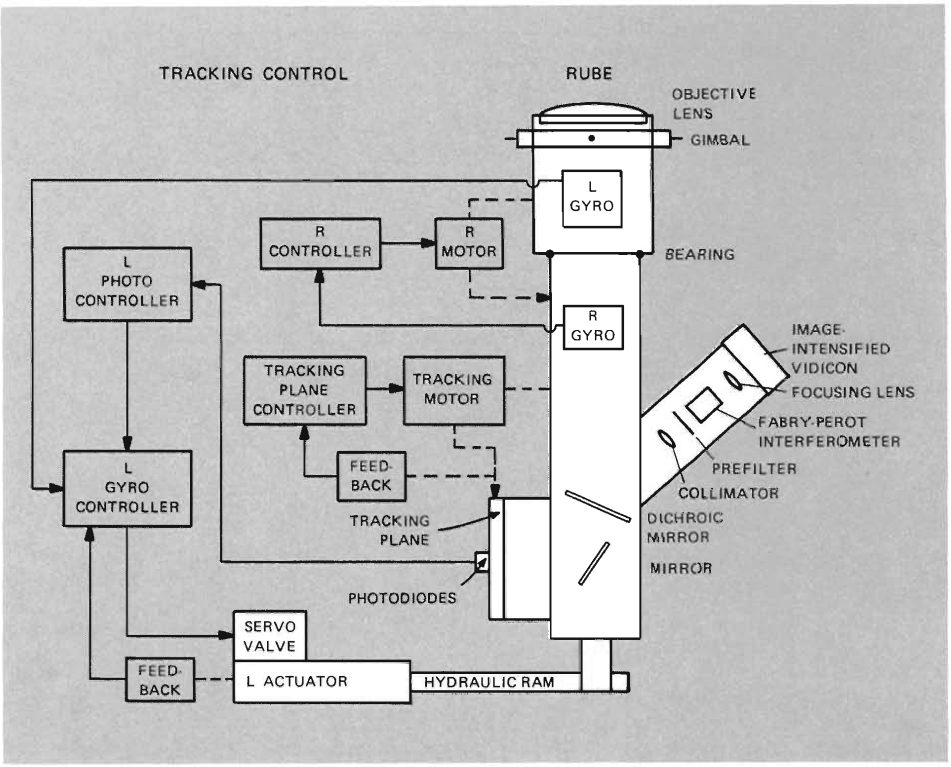


Fig. 15. Main components of the tracking control for the coronal emission line camera. Tracking along two orthogonal axes is controlled directly by an operator or by feedback from gyroscopes mounted on the telescope. Only one of these gyroscopes (L gyro) and its associated control mechanisms are shown here. Gyroscope drift in the two orthogonal axes is compensated by error signals from a set of photodiodes that views an image of the eclipsed sun. The tracking plane controller is preprogrammed with the various tracking patterns to be followed during totality. In addition, rotation of the telescope about its optic axis is required to correct not only for aircraft motion but also for the apparent rotation of the sun during the duration of totality, which amounted to nearly 2° during the 1980 eclipse. This rotation is controlled by feedback from a gyroscope (R gyro) and by preset gyroscope precession in the R controller. Error signals, the difference between a gyroscope axis and the required axis, are about 4-5 arc seconds for the telescope in flight. Similar performance accuracy has been determined from photographs and video recordings of lunar limb motion.

In 1973, we scanned the entire corona out to 2.5 $R_\odot$ with a spatial resolution of 15-20 arc seconds. We measured very weak polarization of the green line along radial vectors from the sun. Weak polarization implies that green line emission is excited by electron collisions rather than absorption of photons. This result confirms expectations that electron collisions are the dominant mechanism by which energy is distributed in the inner corona. We also observed that Ca XV yellow line emission occurred over a much larger region of the lower corona than previously expected. The presence of the yellow line was thought to signify very hot regions of the corona, corresponding to the 4-5 MK ionization potential maximum of the Ca XV ionization state, but the 1973 line widths indicate that the yellow line is produced in cooler regions as well, and thus that coronal temperatures may be much more uniform on a scale of arc minutes than had been guessed from other emission line measurements.

Our 1980 data is much more extensive and has a much higher spatial resolution than that from previous eclipses.

Light from the 1980 experiment was collected by the "Rube Goldberg," a massive 10-inch telescope with an 80-inch focal length that has been used aboard the aircraft since our 1965 expedition. Emission line signals, imaged by the telescope and interferometer, were amplified and recorded on videotape at 16-ms intervals from an image-intensified vidicon detector. This rapid data acquisition system, a dramatic improvement over the 30-second exposure times required by photographic techniques, collected well over 20,000 line profiles from the base of the corona and in some regions out to 3 $R_\odot$. The detector was designed specifically for astronomical measurements and for compatibility with standard computer image analysis equipment by M. T. Sandford. Its progenitor was developed at Los Alamos for diagnostics in the weapons program.

The line resolution and diameter of the detector, combined with the relatively large image size (we view the corona in segments), determine a spatial resolution of 4-5 arc seconds. But to realize this resolution, we must keep the image centered in the interferometer and on the axis of the telescope. We accomplish this by offsetting the entire 210-kg emission line camera (telescope, interferometer, and detector) with large hydraulic actuators that are controlled by a sophisticated tracking system accurate to 4-5 arc seconds (Fig. 15). The hydraulic actuators are able to move the massive instrument at frequencies of up to 10 Hz to correct for higher-frequency aircraft motions.

Data from the 1980 measurements are much easier to reduce than those from previous eclipses. Preliminary results show wide variations in line shapes and are suggestive of the turbulent conditions that may be present during a solar maximum. We also see many de-

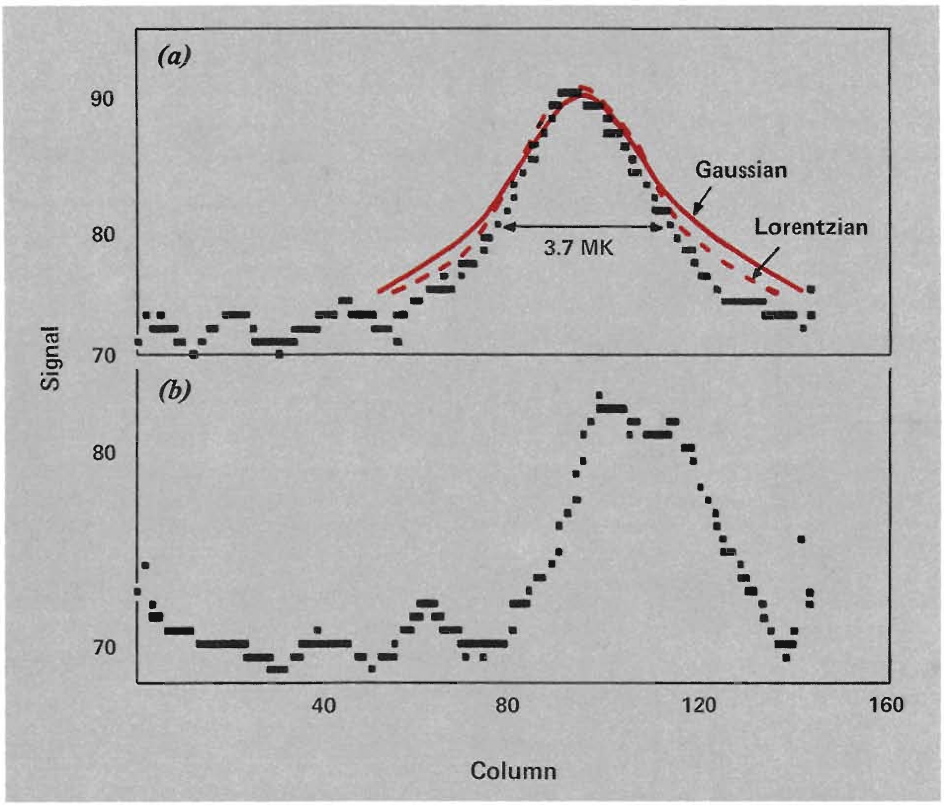


Fig. 16. These preliminary 1980 Fe XIV emission line profiles were obtained by digitizing a sequence of video frames and reordering the data to give a signal versus frame number (or wavelength) profile of the intensity at a single pixel (corresponding to a single spatial location in the corona). In (a) we have fitted both a Gaussian and a Lorentzian profile to the emission line such that the full width at half-maximum intensity and the peak intensity agree with observations. The wings of the observed line are not fitted by either profile. In (b) an example of a more complicated emission line is shown. This profile could represent coronal material with significantly different velocities along the line of sight. A relative velocity of about 15 km/s would fit this profile. Some further corrections for intensity and wavelength calibration will be applied to these data before our analysis is complete.

tails within small regions of high activity. However, data reduction must be completed before we can make definitive interpretations.

Interpreting Emission Line Profiles

Both line shapes and line intensities are analyzed to learn about the conditions in the coronal plasma. Since the intrinsic line widths (from ions at rest) are extremely narrow, the observed Doppler-broadened line profiles are identical in shape to the velocity distribution of the emitting ions.

For ions in thermal equilibrium, both the velocity distribution and the emission line profile would be Gaussian in shape. The width of the line profile would then be proportional to the kinetic temperature of the emitting ions. However, ion temperatures deduced directly from observed Gaussian line profiles (for example, the iron temperatures shown in Fig. 5) may be too high because nonthermal effects, such as macroscopic turbulence and magnetic wave acceleration of the solar wind, may add to the velocities of the ions and, in turn, to the line broadening.

An approximate expression for the line width in terms of the kinetic temperature T and the average turbulent macroscopic velocity v_t is given by

$$\frac{\Delta\lambda}{\lambda} \propto \left(\frac{T}{m} + v_t^2\right)^{1/2} ,$$

where $\Delta\lambda$ is the full width of the line profile at half-maximum intensity and m is the mass of the emitting ion. Note that the kinetic temperature contribution depends on the mass of the ion, whereas the contribution from macroscopic turbulence is mass independent. Therefore it is possible to separate thermal and turbulent contributions to the line broadening by measuring emission line profiles of two ions with widely differing masses. But these measurements must be made at the same time and at the same place in the corona so that we can assume that the two ions have the same kinetic temperature and average turbulent velocity v_t.

Our Ca XV data from the 1980 eclipse may be appropriate for comparison with our Fe XIV measurements, but the emission was weaker than expected and so the profiles may not be accurate enough. However, the Lyman-α data recorded by rocket experiments at several places in the corona will certainly be useful in determining nonthermal contributions to our Fe XIV line profiles.

Variations in line shape are also indications of nonthermal velocities. For example, calculations show that large expansion velocities from solar wind flow tend to flatten and extend what would have been a Gaussian line shape. Although these departures are not significant until expansion velocities are 40 km/s or greater, such velocities are predicted by some models of solar wind flow at distances of 2-3 $R_\odot$. The line profile shapes are also altered by large-scale turbulent motion or significant magnetic wave acceleration of the plasma.

Figure 16 illustrates the variety of line shapes that were actually measured in 1980. We see asymmetries, as well as departures from Gaussian shapes. In-

terpretation of these line shapes depends on theoretical assumptions and supporting data, except perhaps for the case of split lines, which clearly indicate relative bulk motion of the emitting material.

Analysis of line intensities can also be quite revealing. Since emission intensity depends on density and temperature, comparison with independent electron density measurements is helpful in deciding whether bright regions are due to high temperatures or, alternatively, to high densities. We can also analyze intensity variations as a function of radial distance to infer excitation mechanisms. If ions are excited by electron collisions, intensities are proportional to the square of the electron density $(n_e)^2$ and thus should decrease as $1/r^{12}$. On the other hand if photon excitation is the dominant mechanism, then line intensities are proportional to n_e and thus decrease as $1/r^6$.

Finally we can try to find periodic variations in intensity and line shape to identify wave heating of the corona. We have recorded tentative evidence of periodic variations in some of our 1973 data taken aboard the Concorde during a record-setting 74 minutes of totality (Fig. 17).

Emission Line Data Results and Questions

Our 1965 results (see Fig. 5 for average ion temperature values) showed ion temperatures decreasing outward from the base of the corona with little variation in temperature from one coronal feature to another. A more detailed look at the data in each feature is illuminating.

HELMET STREAMER. One interferometer fringe (see Figs. 13 and 14) crossed a helmet streamer near the top of the helmet at 1.4 $R_\odot$ where field lines begin to change from a closed loop to an open-line configuration extending radially from the sun. Since particles in the helmet are presumably trapped along magnetic field lines, the helmet should have above-average temperatures. In particular, along a line through the top of the helmet we should see a temperature increase in the helmet and a temperature decrease on either side. However, our 1965 data (Fig. 18) show that temperatures along this line fluc-

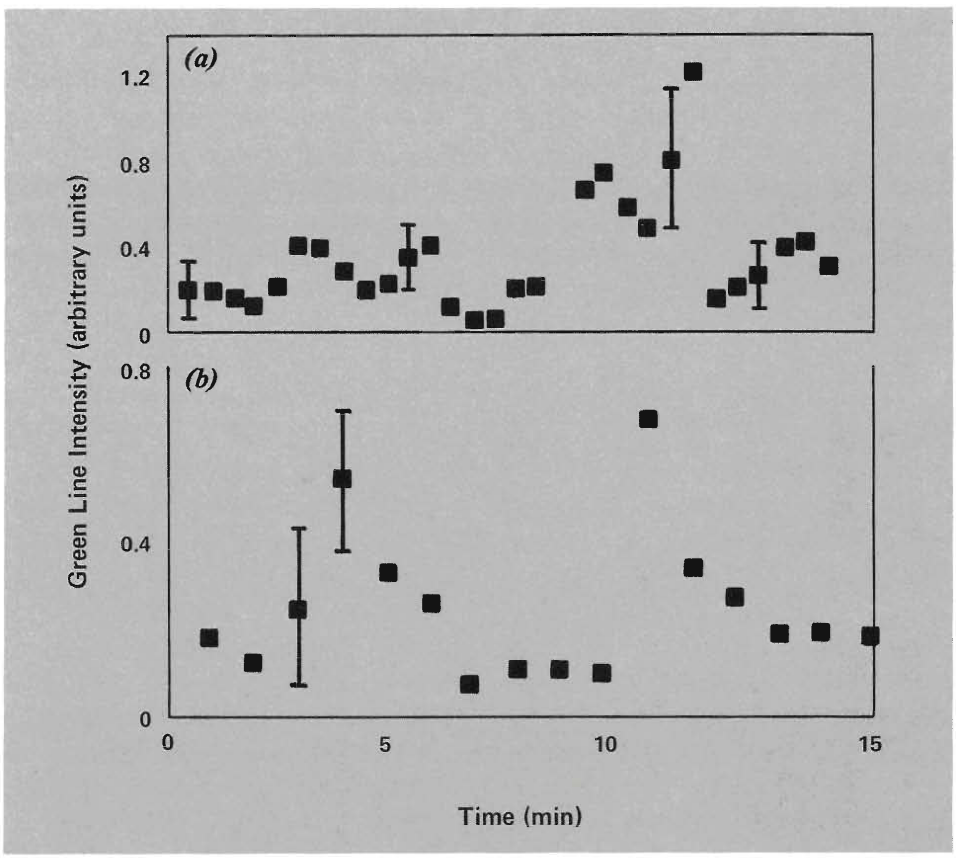


Fig. 17. Peak intensity versus time for two 15-minute segments of Fe XIV emission line data obtained aboard the Concorde 001 at the June 30, 1973, total eclipse. The data in (a), from a region near the base of a helmet streamer at 1.07 $R_\odot$, suggest intensity variations with a periodicity of about 5 minutes corresponding perhaps to a coronal response to the 5-minute oscillations of the photosphere. Tracking problems introduced the errors shown by the bars. The data in (b) are from an equatorial bright coronal region at 1.07 $R_\odot$. The sudden intensity increase followed by a slow decay suggests the presence of a shock disturbance. The variations have a periodicity of about 6 minutes.

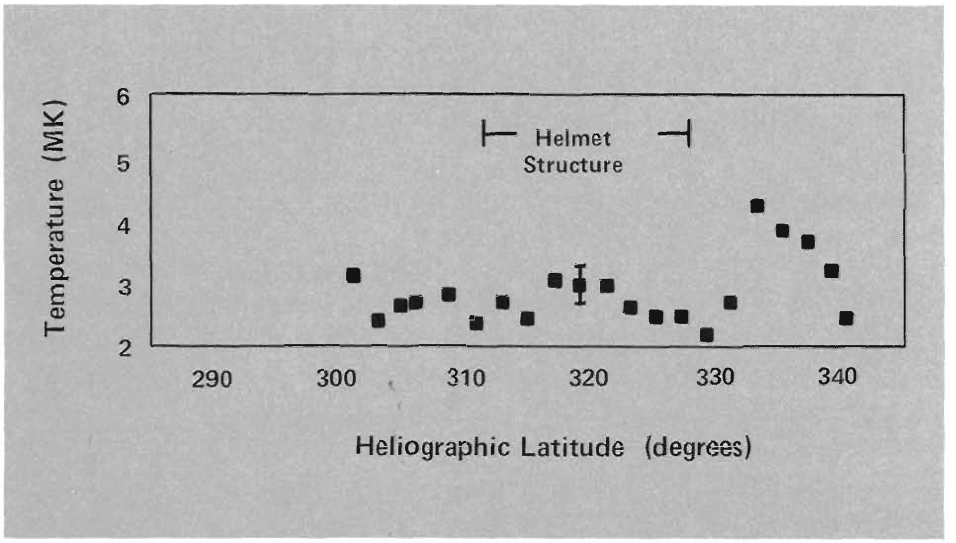


Fig. 18. Temperatures across a helmet structure deduced from Fe XIV emission line data obtained during the May 30, 1965, total solar eclipse.

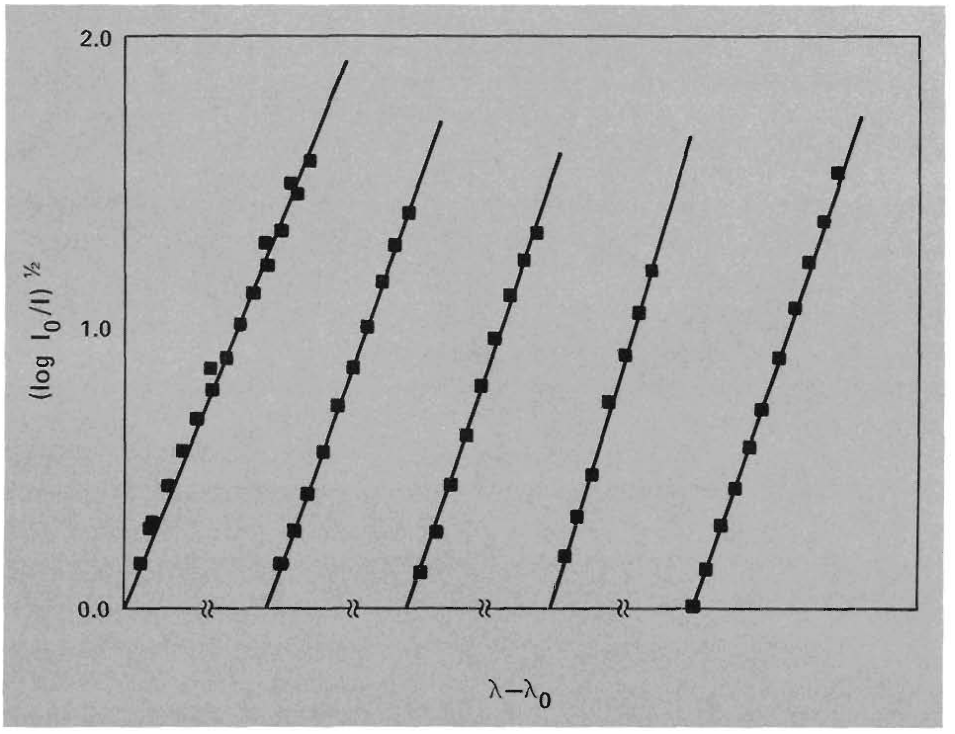


Fig. 19. Intensity versus wavelength profile of the coronal Fe XIV emission line obtained during the May 30, 1965, total solar eclipse. The maximum intensity I_0 is at the central wavelength λ_0. A straight line fit indicates a Gaussian line shape.

tuate from about 2.5-3.0 MK without much systematic change. The Gaussian line shapes from the helmet (Fig. 19) suggest that, as expected, turbulent velocities are small, perhaps less than 25 km/s in this long-lived magnetically confined region.

Our 1980 data will provide a more detailed look at the temperature structure across a helmet and will extend our view of the region above the helmet where the field lines open to form a streamer. We are particularly interested in knowing whether temperature gradients above the helmet are consistent with solar wind flow. At present the contribution of streamer regions to the solar wind is poorly known.

CORONAL HOLES. Several fringes of our 1965 data crossed a coronal hole region, so we were able to estimate temperature gradients as well as temperatures in this region. Coronal holes are expected to have lower temperatures and densities than other regions, and since they are known to be a source of solar wind flow, we expect temperature gradients consistent with solar wind theory.

One surprise was the high temperatures (2.0-2.5 MK) out to 1.7 $R_\odot$, much higher than those determined for the base of the coronal holes by the EUV measurements.

Our temperatures, which we estimated directly from line broadening, would be slightly less but still much higher than those determined from other experiments, if turbulent velocities contribute to the line broadening. In fact, we have invoked turbulent velocities in some parts of this region to deduce temperature gradients consistent with solar wind flow. If the temperature at the lowest observed altitude (1.42 $R_\odot$) is fixed at 2.5 MK, then successive temperatures at 1.62 $R_\odot$ and 1.74 $R_\odot$ are 2.43 MK and 2.39 MK, respectively, provided the turbulent velocities at these three altitudes are 35, 10, and 5 km/s, respectively. The temperature variation is then $T \propto (1/r)^{0.5}$.

We also determined the variation of line intensity with radius in this coronal hole region. Figure 20 shows that intensities fluctuate rapidly near the base of the corona, but at greater heights the intensity is proportional to $1/r^{12}$, or $(n_e)^2$. This rapid falloff is consistent with a collisional excitation mechanism—a rather surprising result given the very low density in this region determined by independent electron density measurements. However, if magnetic field lines are clustered, rather than uniformly dis-

tributed, and if electrons are trapped along these clustered lines, then the average electron density would remain low, as measured along a line of sight, but local densities would be high enough to produce ion excitations by collisional processes. Evidence for density inhomogeneities is seen on white light photographs of the corona, but whether these inhomogeneities correspond to clustering of field lines is not at all certain. More detailed intensity measurements from the 1980 eclipse may clarify the situation.

CORONAL ENHANCEMENT. A coronal enhancement is a long-lived region of higher density that overlies an active region in the photosphere. Consequently, we expect higher turbulent velocities in such a region than in helmet streamers. From our 1965 data we estimated temperatures of ~3.0 MK after subtracting turbulent velocity contributions to the broadening. Turbulent velocities in the enhancement are about 25 km/s and decrease with altitude. Temperature gradients, when corrected for turbulent velocity broadening, suggest that solar wind flow is probably very weak in this region.

We also determined intensity variation with radial distance from 1.1 to 1.6 $R_\odot$. Electron density decreases by a factor of 10 over this region. As seen in Fig. 20 intensity falls off as $1/r^{12}$ or as $(n_e)^2$; collisional excitation, as expected, dominates in this relatively dense region.

THE 1980 CONDENSATION. One of the most exciting observations during the 1980 eclipse was the hydrodynamic eruptive disturbance from the base of the west limb of the sun out to 7 $R_\odot$. From a first look at our Fe XIV emission line data we have been able to identify a very

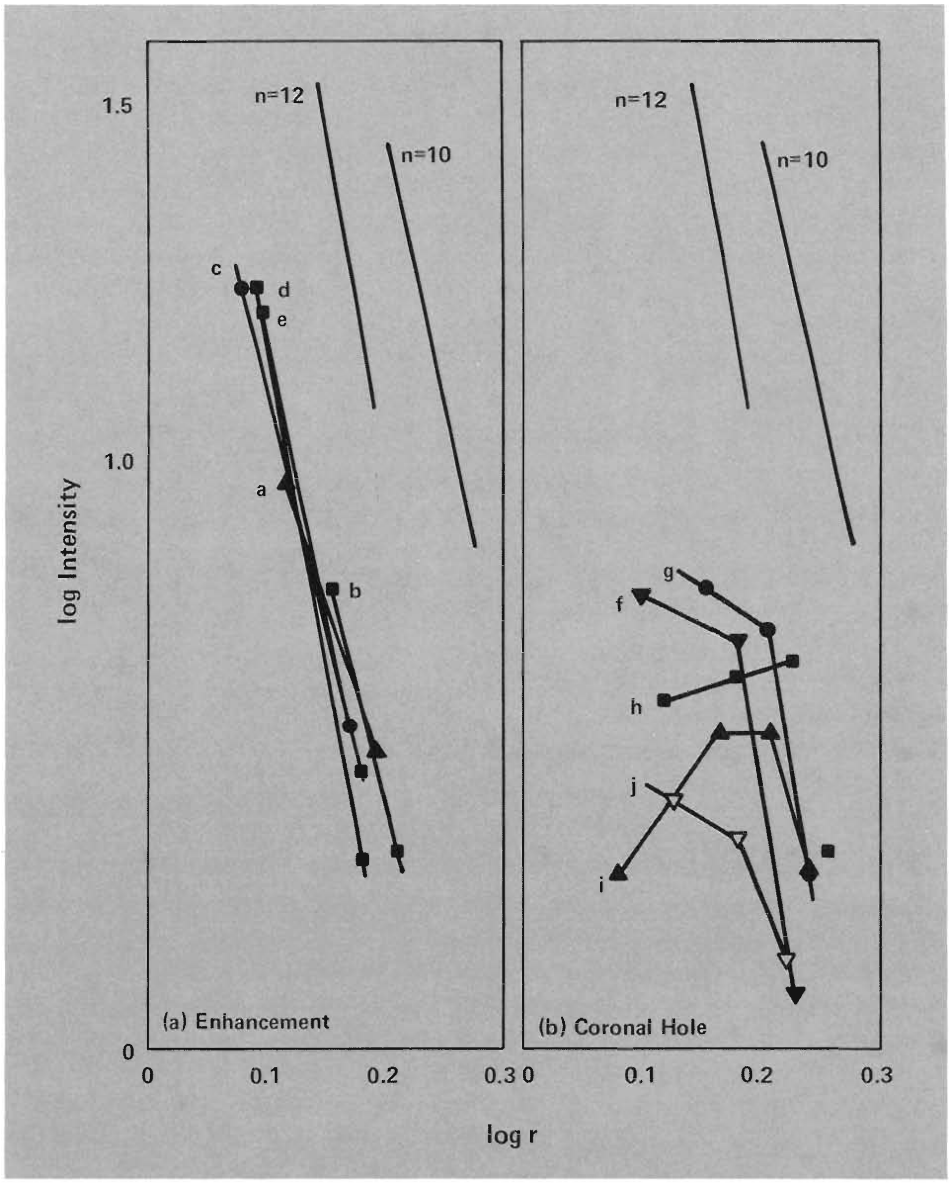
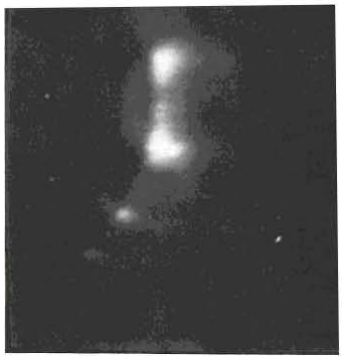
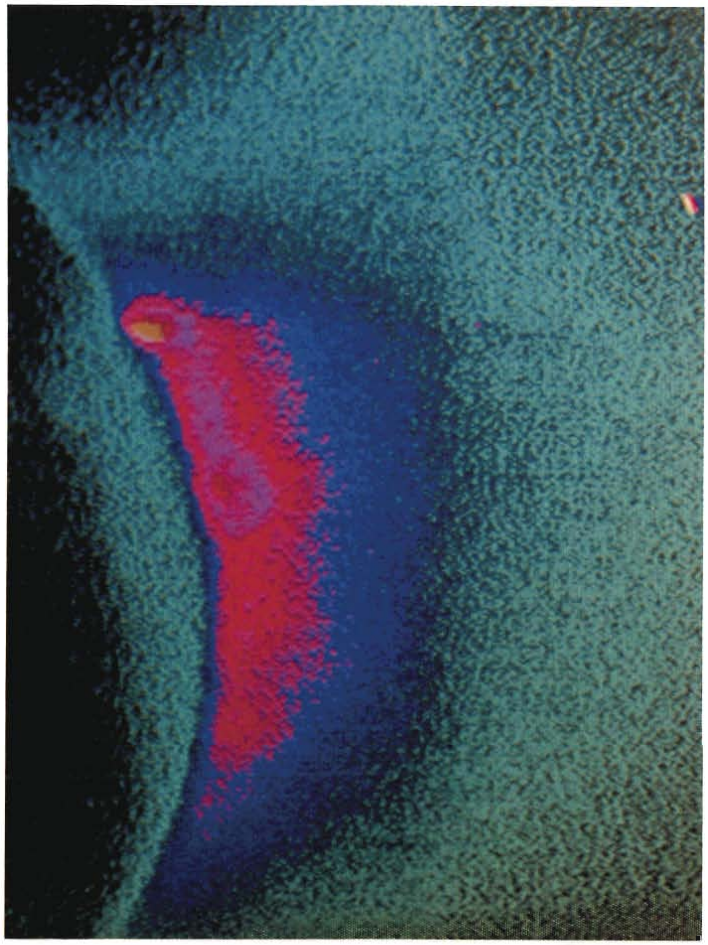


Fig. 20. Emission line intensity versus radial distance for (a) a coronal enhancement and (b) a coronal hole region. Data were obtained during the May 30, 1965, total solar eclipse. Letters label lines through data taken at the same position angle. Curves of $1/r^n$ ($n = 10$ and 12) are shown for comparison.

bright region on the west limb (Fig. 21) whose position angle is near that determined for the eruptive disturbance photographed by the coronal camera. We recorded this region a second time when we were looking for the Ca XV yellow line. The bright knot appears again, this time in the K coronal light from Thompson scattering that is accepted by the interferometer along with the yellow line (Fig. 22). The K coronal brightness in this region indicates high electron densities and so the Fe XIV brightness is probably related to increased density and not just to a temperature close to the Fe XIV ion population maximum. This region has the right size and density enhancement to qualify as a coronal condensation. If other observations taken a day or two later show that this condensation is sporadic or short-lived, we might speculate that perhaps many such condensations are related to the development of hydrodynamic eruptive disturbances. Then previous observations of coronal condensations might be used to determine the frequency of such mass eruptions throughout the solar activity cycle.

NEW CORONAL IMAGING TECHNIQUE. The emission line data from the





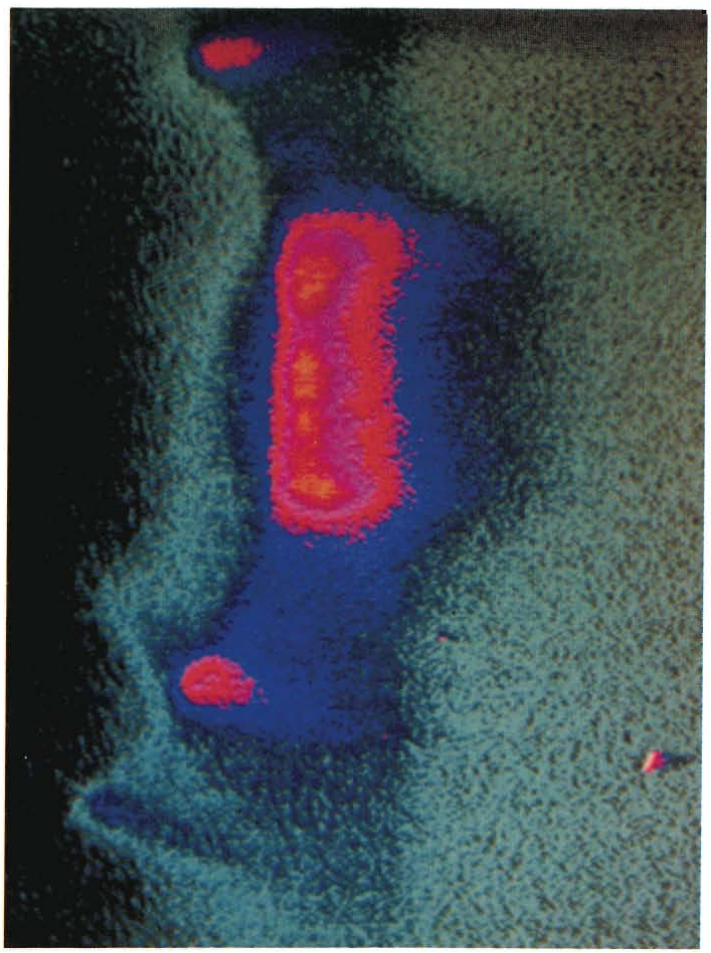


Fig. 21. False colors in this photograph of the Fe XIV emission represent intensity levels, yellow being the most intense. The central fringe was stopped near the west limb. Spatial variations in the emission line intensity are fine scale and regions 20 arc seconds (4 × 4 pixels) in area are observed at various position angles. Data were obtained during the February 16, 1980 total eclipse.

Fig. 22. Intensities of the Ca XV emission line are shown here in false colors, but with enough amplification to bring out the electron-scattered K coronal light. The bright knot is a high-density region and would explain the intense Fe XIV line emission from this same position angle. If the umbilical connection with the hydrodynamic eruptive disturbance (observed in K corona light) is extended back to the solar surface, this bright knot has a corresponding position angle. Data were obtained during the February 16, 1980 total eclipse.

scanning interferometer can be used to obtain a picture of the corona in the light of an emission line. Although such photographs have been obtained with expensive Lyot filters specially designed for operation at one wavelength, the new technique we present here is cheaper and quite adaptable to a variety of wavelengths. A pressure scan over one order of the interferometer is integrated onto a single photograph. As shown in Fig. 23, details of the base of two helmet streamers, the bright knot, and other features on the west limb can be seen on such an image.

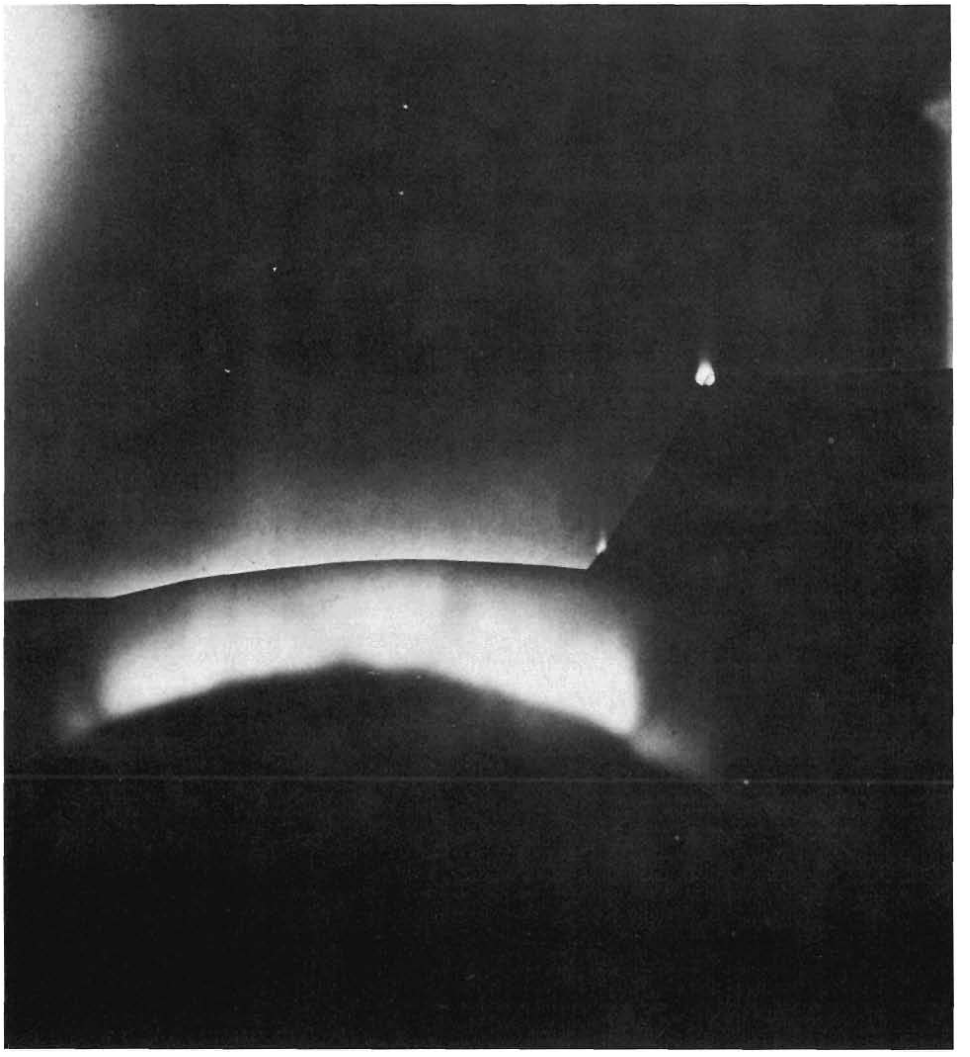


Fig. 23. Photograph of the green line corona off the west limb. This image was processed from a pressure scan of one interferometer fringe. Visible features include two helmet streamers and a bright knot.

Electron Temperature Experiment

Coronal electron temperatures as a function of radial distance are badly needed to determine whether electrons and ions are in thermal equilibrium. In 1976 L. E. Cram* suggested a method to determine these temperatures from spectral intensities of the K corona. Our 1980 attempt to perform the measurement used a much more sensitive technique than has been tried before and, although we failed to record data, we did confirm the feasibility of the technique.

The principle behind the experiment is based on Cram's analysis of the K coronal spectrum. This spectrum is formed as light coming from the photosphere (the Fraunhofer spectrum) is scattered by energetic coronal electrons. The Doppler shifts suffered by these scattered photons broaden the absorption lines in the original Fraunhofer spectrum to produce the diffuse spectrum known as the K corona.

Cram's calculated spectra for a spherically symmetric, isothermal corona at four electron temperatures are shown in Fig. 24. His results display two very interesting features. First, there exist spectral "nodes," that is, wavelengths at which the intensity is independent of the assumed electron temperature. These nodes occur on either side of the H and K absorption lines of Ca II and near other absorption features of the Fraunhofer spectrum. Second, the spectra undulate about these nodes with an amplitude that is related to the electron temperature. Thus, the ratio of two measured intensities, one at a nodal wavelength and one at a wavelength nearby, should determine the electron temperature uniquely.

Although such an experiment is very simple in principle, in practice it is quite difficult because the intensities must be measured with great precision. An error of $\pm 0.2\%$ in the intensity ratio translates to an uncertainty of ± 0.1 MK in coronal electron temperature. The 5-10% errors common in standard photometric techniques would obliterate the information sought (as learned by Menzel and Pasachoff* who searched without success for residual depressions from the H and K absorption lines of CA II in photographic records of absorption spectra from the 1936 total eclipse).

To achieve the required sensitivity, we built our instrument around the only type of detector suitable for the task—an array of silicon photodiodes. These solid-state photosensitive devices have very high quantum efficiency in the spectral range of interest.

A commercially available detector, consisting of a linear array of 512 silicon photodiodes on a single integrated circuit, served as the main component of our K corona spectrograph (Fig. 25). This large array enabled us to look not just at two wavelengths but at

L. E. Cram, "Determination of the Temperature of the Solar Corona From the Spectrum of the Electron-Scattering Continuum," Solar Physics **48**, *3-19 (1976).*

D. H. Menzel and J. M. Pasachoff, "On the Obliteration of Strong Fraunhofer Lines by Electron Scattering in the Solar Corona," Publications of the Astronomical Society of the Pacific **80**, *458 (1968).*

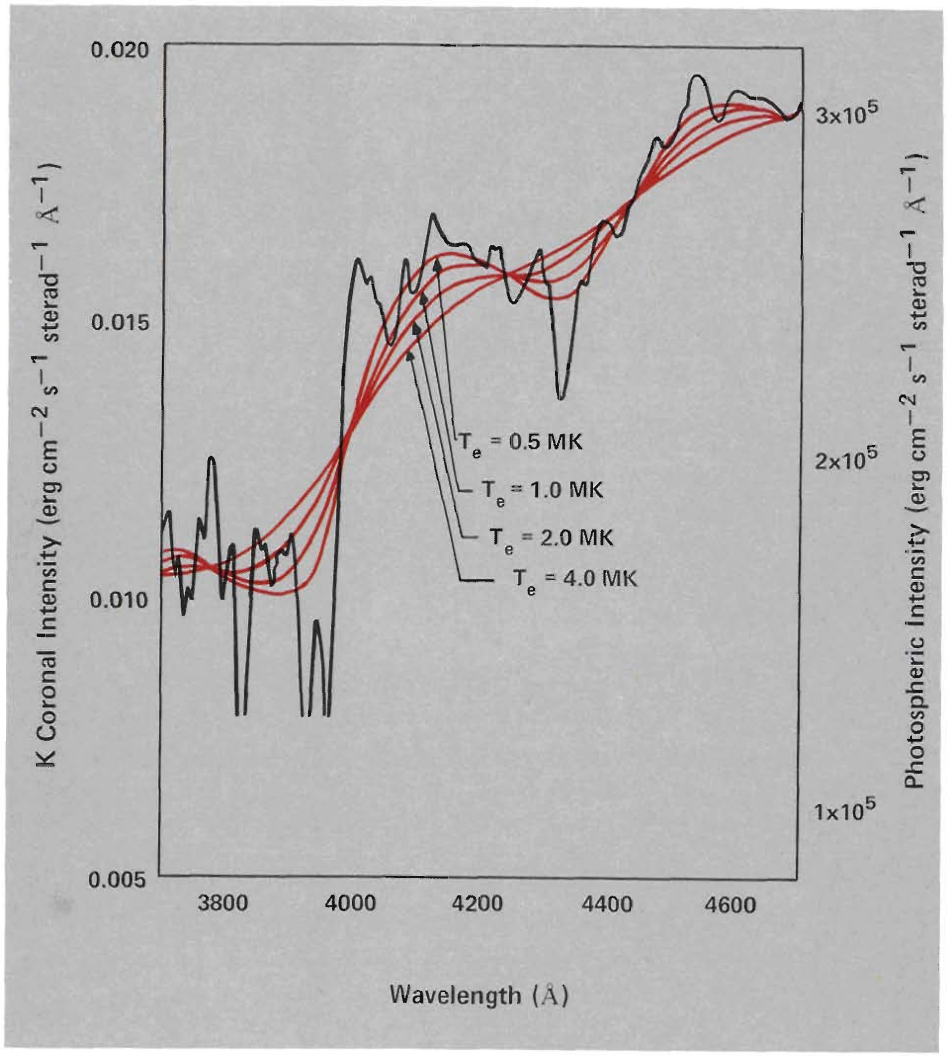


Fig. 24. Absolute intensity of the K coronal spectrum for four coronal electron temperatures, as calculated by Cram. The absolute intensity of the photospheric spectrum from the center of the sun's disc is shown for comparison.

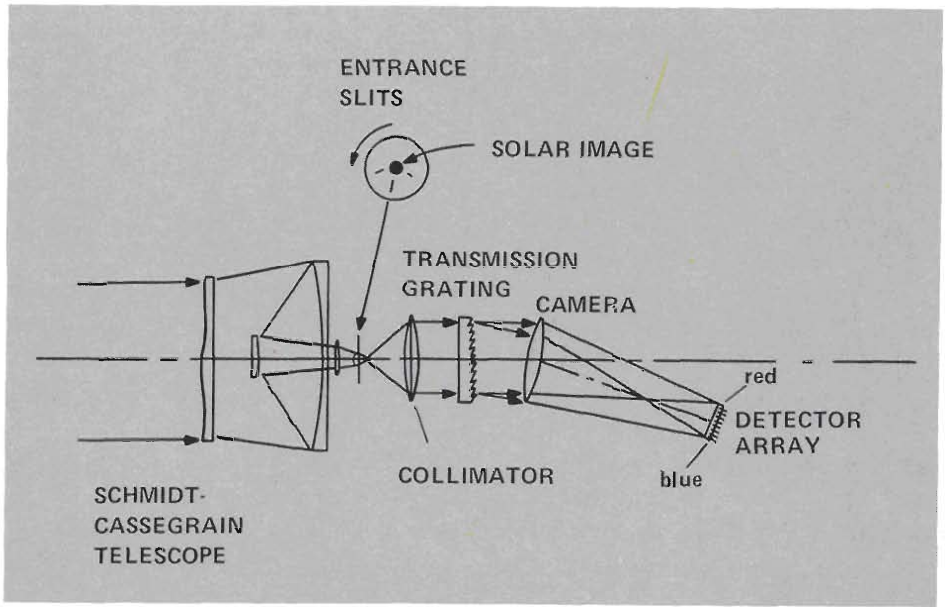


Fig. 25. The K corona spectrograph for the electron temperature experiment consists of a telescope, entrance slit assembly, a transmission grating spectrograph, a silicon photodiode detector array, and electronics for data acquisition and storage. The f/2.0, 400-mm focal length telescope maximizes the light at the detector. Three entrance slits, corresponding to radial distances from the sun of 1.2, 1.6, and 2.0 $R_\odot$, were formed by photolithographic techniques on a glass disc. A computer-controlled lever-arm and motor-drive assembly positions the selected entrance slit in the telescope focal plane. The entrance slit length is magnified by the spectrograph to match and thereby utilize the entire photodiode length. The spectrograph covers 1500 Å and its resolution is 3 Å per channel. Exposure times range from 10 seconds at 1.2 $R_\odot$ to 40 seconds at 2.0 $R_\odot$. Satisfactory data can therefore be obtained only during eclipses of relatively long duration or from aircraft whose flight path lengthens the duration of totality.

wavelengths of the K coronal spectrum over a 1500 Å range with 3 Å resolution.

The experiment was in some sense a fishing expedition. We set out to observe the entire spectrum with high precision and search for Cram's predicted undulation patterns at three radial distances from the sun.

During the eclipse, the spectrograph appeared to function well, and we were able to confirm that its sensitivity is adequate to measure the shape of the K coronal spectrum and, in principle, to determine electron temperatures. The data that appeared on our monitoring screen were not recorded by our computerized data-acquisition system (Fig. 26) because both the computer disk drive and the back-up drive failed to function.

Indeed the equipment failure was a tremendous disappointment, but we did prove that the experiment was feasible, and are now redesigning some of the electronics with hopes of flying again during the 1983 solar eclipse.

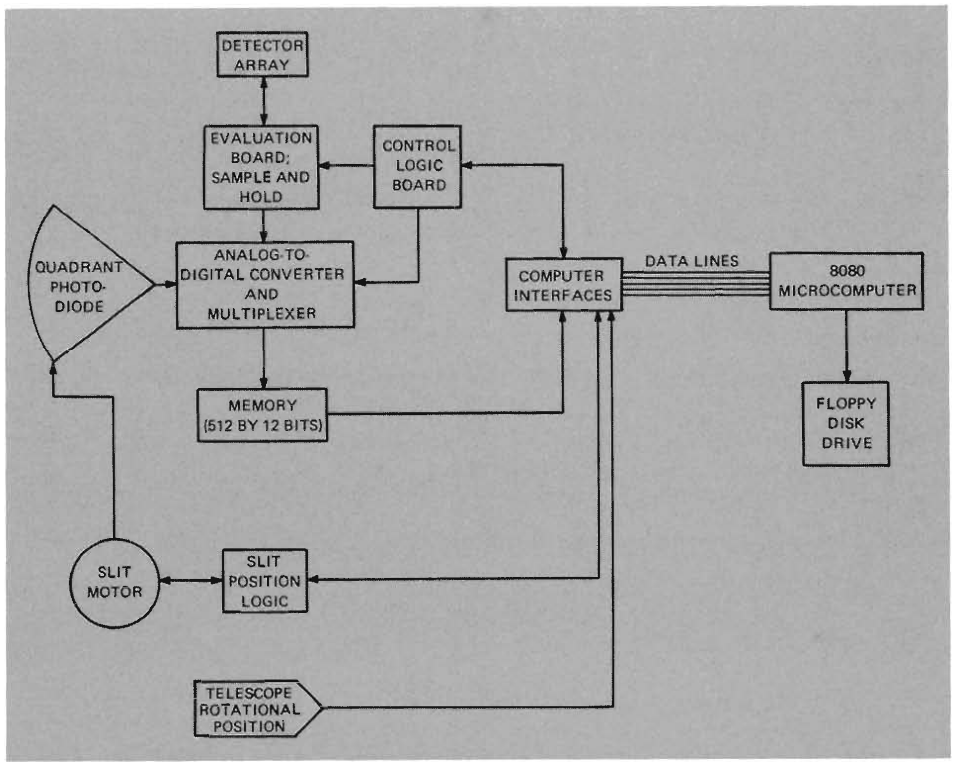


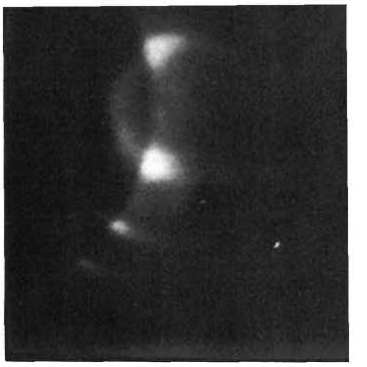
Fig. 26. Block diagram of computer-controlled data-acquisition and storage system for electron temperature experiment. The detector array was read out with Reticon Corporation's evaluation circuitry. Although not ideal for this application, the circuitry provided a signal-to-noise ratio of greater than 100. Data were to be stored in near real time on floppy disks. Computer programs were written in FORTH, a language developed at the University of Rochester.

Infrared Observations of Dust Rings

The F corona and the zodiacal light, a zone of scattered light symmetric about the approximate plane of the planetary orbits, give clear evidence of the presence of dust throughout much of the solar system. These observations also indicate that the dust particles are very fine—only a few micrometers in diameter. Their presence may be a remnant of the formation of the solar system but probably wandering comets have also left a contribution.

Whatever the source, theorists have modeled the fate of the dust once it is deposited in the solar system. The dynamics of the dust particles is influenced by four factors: the sun's gravity, radiation pressure, the Poynting-Robertson effect, and particle evaporation. Calculations including these factors indicate that the dust will spiral in toward the sun but may ultimately settle at particular distances and form broad rings about the sun. Evaporation of the dust may also lead to its being blown outward once again. By determining the radial location of these rings we learn something about the dust's composition because its location depends on its reflectivity and evaporation temperature. Likely compositions include obsidian, silicate-type rock, iron, and perhaps water-ice.

The presence of glowing dust rings at 4, 9, and 20 $R_{\odot}$ and elsewhere has been deduced from observations of infrared emission features (Fig. 27). These emission features are presumably produced as dust particles heat up and evaporate, emitting radiation corresponding to black-body temperatures between 500 and 2000 K. These observations are difficult to make from the



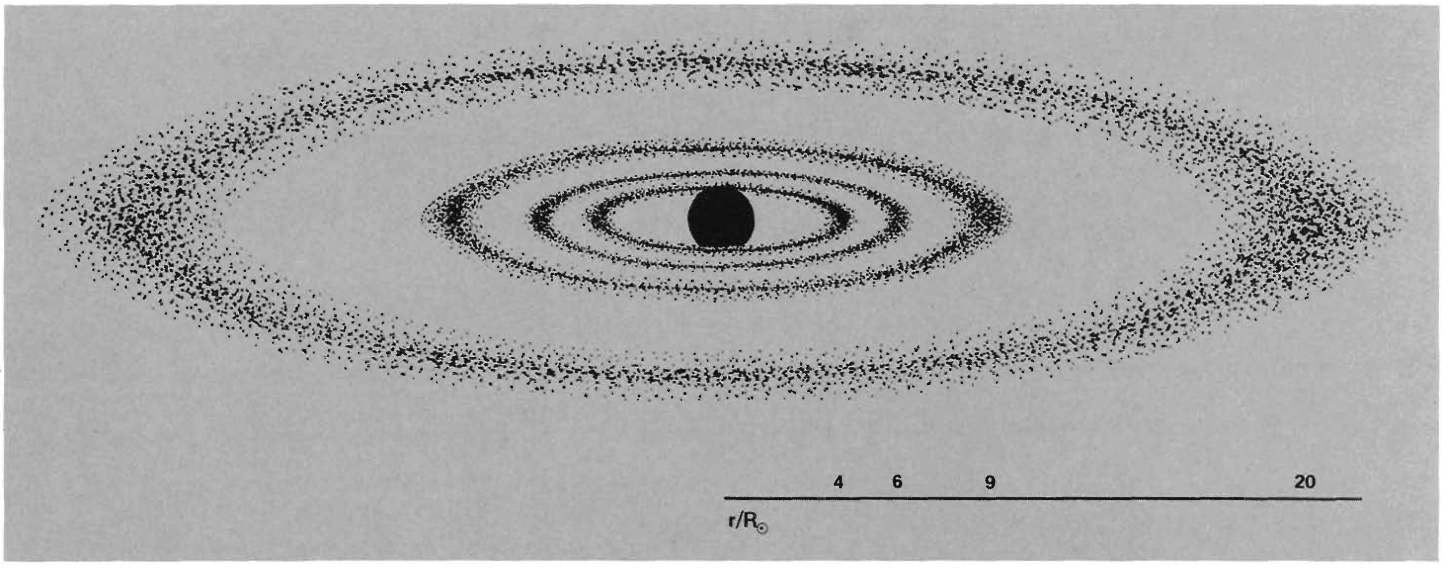


Fig. 27. Approximate radial location and extent of infrared emission features reported by various observers. Only the features at 4 and 9 $R_\odot$ are confirmed.

ground and, with the exception of the feature at 4 $R_\odot$, the existence of these features is in dispute.

From aboard the aircraft during the 1980 eclipse, we made a new attempt to determine the spatial distribution of infrared emission features.

Infrared intensity as a function of distance from the sun was measured at 2.2 μm with an InSb detector, a standard technique, and at ~0.7 μm with a charge-injection-device television camera.

The corona was scanned along the ecliptic (plane of the earth's orbit) out to 50 $R_\odot$ east and west of the sun and at several angles to the ecliptic. The scan pattern was designed to determine the radial location and approximate extent of any dust rings.

Both detection systems worked well and yielded high-quality data. However, tracking errors during totality will severely compromise interpretation of data close to the sun, so we will not be able to confirm the existence of the feature at 4 $R_\odot$. At greater distances, however, these errors have less effect.

Our analyses to date strongly suggest the existence of a dust ring at 9 $R_\odot$, but indicate no other features at greater distances from the sun. Thus we find no feature at about 20 $R_\odot$. A feature at this location, which corresponds to a temperature of ~800 K, had been suggested by E. P. Nye on the basis of ground-level observations at the February 26, 1979, and February 16, 1980, eclipses; however, at both eclipses his observations may have been contaminated by the effects of clouds. Existence of such a feature would be of interest because dust shells with temperatures of ~800 K have been found about some other stars.

Conclusions

Most scientists studying the solar corona believe that a temperature maximum exists near 2 $R_\odot$. Our results indicate that the maximum might be closer to the limb, although some of the Fe XIV emission line broadening can always be attributed to large-scale turbulence or to expansion velocities. However, comparison with proton temperature values from the rocket-borne coronagraph at a few positions in the corona will allow us to determine such velocities if they exist. We can then interpret the rest of our iron emission line data in light of these findings. The fact that many of our 1980 profiles are actually double-peaked indicates large differential mass motions that are probably due to the extremely complex nature of coronal features during maximum solar activity.

Ion temperature gradients measured in 1965 indicated that coronal holes could support solar wind flow. We now have high-resolution data from our 1980 measurements of the residual coronal hole above the sun's south pole to compare with the 1965 results. We will also

be able to analyze temperature gradients and velocity distributions in streamer structures, to determine whether these structures also contribute to solar wind flow.

Our emission line intensity data from several previous solar eclipse observations suggested that collisional mechanisms dominate energy transport in the lower corona out to 2 $R_\odot$. Our high-resolution 1980 data from several streamer structures and one coronal condensation, when combined with electron densities determined from camera-polarimeter measurements, will be a better test of this model.

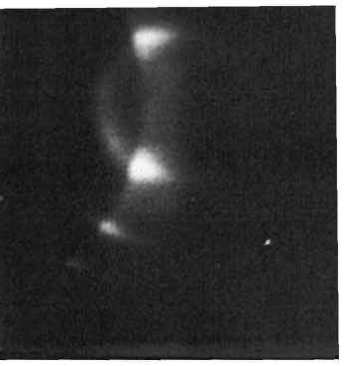
The possibility of extended wave heating in the corona has become increasingly significant. Our emission line intensity data from the 1973 Concorde flight, which gave 74 minutes of totality, provided tentative evidence for periodic temperature variations, and in 1980 we looked for temporal changes on the sun's west limb by taking a sequence of intensity measurements at 30-second intervals over the 7 minutes of totality. But a more extensive search will be required to establish the dynamics and the extent of plasma heating.

Comparison of the computer-enhanced coronal photographs from five eclipses shows that the corona at the 1980 eclipse exhibited features never before seen, including the umbilical connection of the hydrodynamic eruptive disturbance to the solar surface. This observation calls into question previous speculation, based on Skylab photographs down to 1.5 $R_\odot$, that the connection to the surface was quite broad. Evidently the energy source for some mass eruptions is localized and concentrated. We may learn more about the dynamics at the base of this disturbance once our electron density and Fe XIV data have been reduced.

Our photographic record also shows that nearby streamers respond to the eruption by bending around it, an effect also seen on Skylab photographs but only out to 6 $R_\odot$. The curious shape of long streamers above the sun's south pole suggests that a second eruptive disturbance may also have been in progress at the same time.

Electron densities resulting from our observations agree with other studies in the inner corona, but beyond 4 $R_\odot$ in equatorial regions the density decrease with radial distance is much more gradual than indicated previously. This could have a marked effect on theoretical models of the solar wind.

In summary, we now have collected ion temperature and electron density data at all phases of solar activity except during a deep minimum. Our 1980 data combined with proton temperatures from rocket experiments will provide a focal point for all future analysis. For the first time, we will have an accurate determination of material density, ion temperatures, and nonthermal velocities at several places in the corona. We hope these results will be a basis for more meaningful interpretation of our data throughout the rest of the inner corona ■



AUTHORS

Donald H. Liebenberg earned his Bachelor of Science in 1954 and his Ph.D. in physics in 1971, both from the University of Wisconsin. He has been a Staff Member at Los Alamos since 1961. His interest in solar physics, which dates to his undergraduate days, led in 1963 to the Laboratory's involvement in observations of total solar eclipses. This activity has resulted in detailed studies over a solar cycle of the physical properties and energy transfer mechanisms of the coronal plasma. During 1967 and 1968 he was Program Director for Solar Terrestrial Research at the National Science Foundation. In 1973 he received a National Science Foundation Grant for observation of a solar eclipse from the French-British Concorde 001. His other research interests include properties of gases at high pressures and laser optics and technology.

Charles F. Keller, Jr., is well known for his involvement in airborne eclipse expeditions, having served as principal investigator of polarization experiments on missions in 1970, 1972, 1973, 1977, and 1980; he was scientific coordinator for missions mounted in 1979 and 1980. After receiving his Bachelor of Arts from St. Vincent College in 1961, he went on to earn a Bachelor of Science in physics from Pennsylvania State University and a Master of Science and Ph.D. in astronomy from Indiana University in 1967 and 1969, respectively. His thesis, "Computer Models of Pulsating Stars Including Radiative Transfer," was produced jointly at Los Alamos and Indiana University. He joined the Laboratory in 1969 and has served as a Staff Member, Assistant Group Leader, and Group Leader of the Diagnostics Design Group of the Field Testing Division. At present he is with the Laboratory's Computer Modeling Group of the Geosciences Division. Keller served as the Field Testing Division's representative on the Director's Computer Advisory Committee, and as the Laboratory's coordinator of computer modeling for the Department of Energy's Atmospheric Studies over Complex Terrain (ASCOT) Program. He is a member of the American Astronomical Society and its Solar Physics Division.

Further Reading

E. N. Parker, *Interplanetary Dynamical Processes* (Wiley-Interscience, New York, 1963).

D. E. Billings, *A Guide to the Solar Corona* (Academic Press, New York, 1966).

H. Zirin, *The Solar Atmosphere* (Blaisdell Publishing Co., Waltham, Massachusetts, 1966).

J. C. Brandt, *Introduction to the Solar Wind* (W. H. Freeman and Co., San Francisco, 1970).

R. G. Athay, *The Solar Chromosphere and Corona: Quiet Sun* (D. Reidel Publishing Co., Dordrecht, Holland, 1976).

J. B. Zirker, Ed., *Coronal Holes and High Speed Wind Streams* (Colorado Associated University Press, Boulder, 1977).

E. N. Parker, *Cosmical Magnetic Fields: Their Origin and Their Activity* (Clarendon Press, Oxford, 1979).

J. B. Zirker, "Total Eclipses of the Sun," Science **210**, 1313-1319 (1980).

IN FLIGHT:

THE STORY OF LOS ALAMOS ECLIPSE MISSIONS

by Barb Mulkin

Los Alamos can claim an extensive and quite unique history of solar eclipse studies, but little known is that the capability arose from resources made available during the uneasy period climaxed by the signing of the 1963 Atmospheric Test Ban Treaty.

Made cautious by an era of broken nuclear test agreements, Congress required four safeguards before ratifying the Treaty: establishment of an aggressive underground testing program, improvement of our monitoring capability vis-á-vis Sino-Soviet nuclear activity, continuation of progress in America's nuclear technology, and "maintenance of the facilities and resources necessary to institute promptly nuclear tests in the atmosphere should they be deemed essential to our national security or should the Treaty or any of its terms be abrogated by the Soviet Union."

So was born the Test Readiness Program and with it the acquisition by the Atomic Energy Commission of a joint interest in three Air Force NC-135s to be used as flying laboratories by the nation's weapons research facilities of Los Alamos, Sandia, and Lawrence Livermore.

According to the AEC's 1964 annual report to Congress, weapons test diagnostics from airborne platforms had proved so successful in the latter part of Operation Dominic (the last American atmospheric test series in 1962) that extensive ground installations such as those at Christmas Island and Eniwetok were unnecessary. The point was also made, perhaps, that some things can be done more thoroughly from the air. In any event, Air Force crews began flying readiness missions in the modified NC-135 jets with scientific contingents drawn from the weapons labs. For a while, B-52 bombers simulated weapons drops above the Pacific so scientists could hone their data-gathering skills, and the 135s came to be viewed as the mainstays of the Test Readiness Program's diagnostics effort.

However, the rather bland simulation exercises added little joie de vivre to the scientific flight crews. The researchers, quick to see the potential of the aircraft for a different kind of scientific mission, petitioned the AEC for permission to gather data on cosmic rays, electrical and magnetic fields in the ionosphere, and the frequent solar eclipses studied almost exclusively, until then, from the ground. They argued that such missions could replace the simulation exercises without interfering with the readiness status required by the Treaty.

High-flying, long-range jet aircraft, the scientists claimed, could be used as observation stations for extraterrestrial phenomena and would have a tremendous advantage over ground facilities. Jet aircraft capability would place researchers above the clouds that often negate ground observation, and above 80 per cent of the earth's atmosphere, whose dust and water vapor scatters light and degrades data. For instance, they contended, meaningful measurements of the solar corona, which are impossible from the ground because of water vapor that blocks the infrared end of the solar spectrum, could be made from the aircraft.

The obvious advantages of a fast-flying plane in tracking a solar eclipse were presented to the AEC as making it possible to extend the length of totality by typically 50 per cent over a fixed ground location.

Scientific arguments were sufficiently persuasive that the AEC gave its permission for the program within a program that continued, albeit catch-as-catch-can, until 1975.

The first airborne solar eclipse mission in 1965 was to take Laboratory researchers to Pago Pago in American Samoa. Led by astrophysicist Art Cox, the scientists were confronted by several obstacles, for the planes were already fully equipped laboratories and solar observation equipment had to be phased in smoothly, so as not to interfere with the aircraft's primary mission. Cox's crew designed optical instrumentation, provided stabilized instrument platforms that reduced image motion blur, fitted their instruments into the cramped space available, arranged for the fuselage of the plane to be modified at Kirtland Air Force Base in Albuquerque to allow special instruments to be mounted, and plotted a course that would place them at that time and point in three-dimensional space where a few precious minutes of totality could be mined for information.

Pago Pago followed months of frantic preparation and it was a tour de force. The mission was notable for many things, not the least of which was the debut of an ingenious multiple-use telescope dubbed the "Rube Goldberg" for its unusual configuration that was reminiscent of the zany cartoon contraptions by that name. The Rube was much modified for later missions and has long been recognized as the workhorse of the airborne eclipse effort.

Designed by a group headed by Don Liebenberg that included Ken Williamson, Bill Ogle, Walt Wolff, and Paul Rudnick of Los Alamos and Mert Rob-

IN FLIGHT:
THE STORY OF LOS ALAMOS ECLIPSE MISSIONS

Art Cox: "The pilot chose an emergency option called full military-rated thrust to zoom to more than 40,000 feet. . . when we landed, he wouldn't let me see the fuel gauge". . . Pago Pago '65

ertson and Melvin Mattison of Sandia, the Rube made five important measurements through its 80-inch focal length lens. Detailed photometry of the coronal continuum and six emission lines was obtained, as well as interferometric observations of the shape of two iron emission lines and a photospheric absorption line. In addition, the Rube yielded information on the interferometer fringe pattern of the green iron emission line to 1.7 solar radii. While the data made the crew jubilant, it was only a beginning, as the cadre of astronomy buffs planned for more airborne missions that in February, 1980, would total seven.

Pago Pago 1965 is remembered by those who flew for more than good data and a successful mission; there were lighter moments that have entered the folklore of the tightly knit group of solar eclipse veterans. For instance, Bill Regan, who retired from Laboratory service last year and has flown on six of the seven missions, recalls that on the Pago Pago expedition he was forced to shoot his pictures, as the aircraft nosed into totality, from the only space available—the forward latrine.

Our Bill Ogle, the AEC representative for the mission, is remembered for the aplomb he displayed when he rose to respond to the greeting of a Samoan chief at a party thrown for the crew. Ogle lost his lava lava, the native skirt donned for the occasion, and was thereafter known to the natives as a "smart palangi" for having had the forethought to have donned a pair of underpants.

A few hairy moments during the eclipse flight are also recalled, for the scientists were forced to "go to war." At the plane's planned altitude, just before totality, they found they had to contend with high cirrus clouds that would wipe out any hope of gathering high-quality data. The flight crew chose an emergency option called "full military-rated thrust" to zoom to more than 40,000 feet—an option exercised generally only in wartime and then for just a few minutes because of the increased use of fuel and the possibility of burning up an engine. The ploy worked, but Art Cox remembers that pilot Jim Wells refused to let him see the fuel gauge when the party landed at Samoa.

Veterans now, the Laboratory team began planning for the 1966 eclipse off the coast of Brazil, when no less than five aircraft were to be involved in a coordinated effort to wrest the sun's secrets while its brilliance was diluted by the moon's shadow.

The Los Alamos NC-135 flew first, followed by a Sandia plane, the Lawrence Radiation Lab's aircraft, NASA's Convair 990, and the Air Force Cambridge Research Laboratory's craft.

The original plan called for the three leading planes to maintain a ground speed of 450 knots along the path of totality, but a crosswise 150-knot jet stream necessitated a change in speed and made it extremely difficult to navigate accurately. However, Bob Brownlee reported that Los Alamos managed 160 seconds of total eclipse observation with instruments installed for backward-look angles. Three optical systems were used on the mission: a coronal camera with a 4-inch aperture and 36-inch focal length pointing directly at the sun to avoid the polarization effect mirrors would have introduced in feeding a fixed telescope; a second instrument with eight parallel telescopes to

record the corona in the light of five spectral emission lines and three nearby continuum regions; and the Rube, tracking on three axes to provide a steady rotation-corrected image.

An innovative feature on this mission was the hydraulic tracking system used for several of the telescopes; it was quicker and more powerful than the electric motors used on the first eclipse flight, and the accuracy of 1 minute of arc was improved to 10 seconds of arc.

The presence of five aircraft and simultaneous rocket experiments on this mission gave pause to the veterans who flew off the coast of Brazil; they remember being tempted to watch the sky for payloads heading in their direction!

At the official welcoming ceremony, bemedalled and bemused South American officials greeted the rumpled eclipsers as they deplaned, headed by Bill Ogle, resplendent in a pair of bib overalls.

Electing to use some vacation on the way home, two intrepid souls ventured onto a boat crossing the famed Lake Titicaca that bisects the Andes between Peru and Bolivia. An attempt to pay for a boat ticket with Bolivian currency failed; Peruvian coins were needed; the Peruvians took travelers' checks, however. At least, they took them long enough to admire them politely, having no idea they could be exchanged for money.

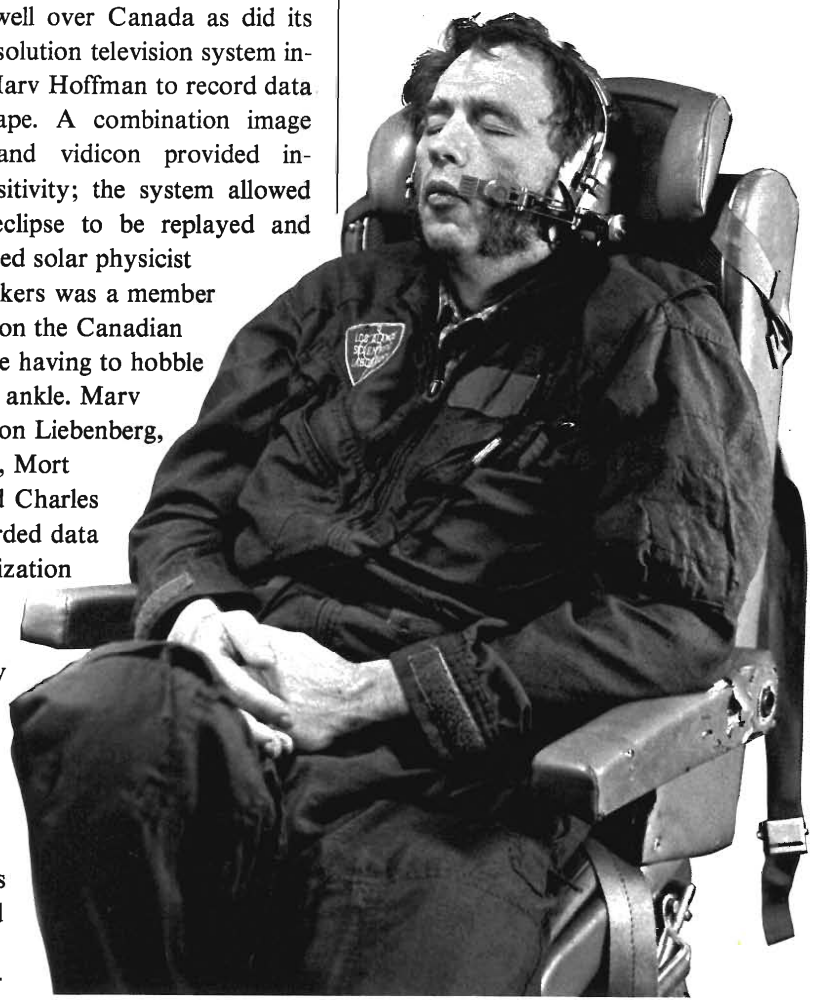
Before the 1970 eclipse, heavy rain in San Antonio, Texas, jeopardized experiments before takeoff, but subsided in time for windows to be cleaned; the Rube recorded data out to two solar radii; streamers to 10 solar radii were photographed and one streamer, dominant in the northwest quadrant, was photographed out to 13 solar radii.

At the traditional open house for the San Antonio press hosted aboard the aircraft before the mission, reporters edged gingerly around the mounds of equipment. One timid neophyte of the printed word stopped before the Rube and asked for an interview with "Dr. Goldberg."

Progress in sun missions continued with the interception of the 1972 eclipse by flying astronomers at 39,000 feet northwest of Canada's Hudson Bay. The Rube, which had delivered coronal light to interferometers on three missions, performed well over Canada as did its new high-resolution television system installed by Marv Hoffman to record data on video tape. A combination image intensifier and vidicon provided increased sensitivity; the system allowed the entire eclipse to be replayed and studied. Noted solar physicist Jacques Beckers was a member of the team on the Canadian flight, despite having to hobble on a broken ankle. Marv Hoffman, Don Liebenberg, Joe Calligan, Mort Sanders, and Charles Millich recorded data on the polarization of the green emission line of highly ionized iron at 5303 Å.

Data collected by the Rube on four missions now spanned much of the 11-year solar cycle from minimum activity (1965) to near maximum (1970), and a comparison of temperatures, as the data were reduced, indicated that average coronal temperature varies little over the solar cycle and is more dependent on activity on the surface of the sun.

The grueling aftermath of the 1973 eclipse is evident in sleeping Art Cox. (Bill Regan Photo)



IN FLIGHT:
THE STORY OF LOS ALAMOS ECLIPSE MISSIONS

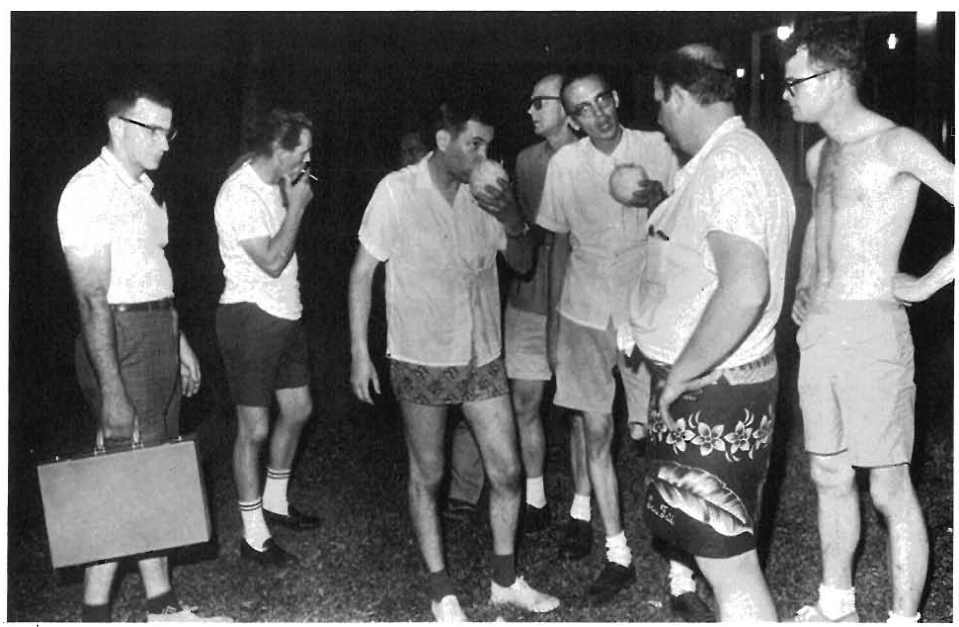
Don Liebenberg: The moon appeared to work as a slow camera shutter, slicing off only a tiny piece of the atmosphere each second. . . from the SST we managed an incredible 74 minutes of totality. . . Africa '73

The Laboratory eclipse experts scrambled next for the 1973 mission, which came to be called "the frosting on the cake." Liebenberg boarded the SST prototype—the French-British Concorde—to monitor the 20th century's 50th eclipse, which occurred over Africa. Racing at nearly 1300 miles per hour (Mach 2), Liebenberg, Kitt Peak's Don Hall, and others from Britain and France managed man's longest look at a total eclipse—an incredible 74 minutes of totality.

Liebenberg recalls looking with awe at the shadow of the moon racing across the Sahara Desert. The flying altitude of the SST was so high that by Air Force definition it was in space, and the curvature of the earth was clear against a purple sky, as the plane raced to intercept the elliptic cone of the eclipse shadow. The precise flight plans were drawn using a computer code developed by Art Cox, a code so accurate that the plane arrived at its destination within one second of the predicted time! As the mission proceeded, those on the 135 watched in disbelief as the SST streaked by and above at 55,000 feet.

Liebenberg's emission line experiments were coordinated with work being done on the slower NC-135, and were designed to probe the transition region between the inner and outer parts of the sun's chromosphere, about which little was known. Hoffman, in the slower subsonic plane, had almost 13 minutes of observation time, little by comparison with the SST's 74 minutes, but still the longest time for airborne eclipse observa-

It was hot and humid at the Tafuna Airport just before the 1965 mission. Charlie Hyder went native in a Samoan lava lava; the rest opted for more conservative attire. From left are Carl Young, Harold Staake, Don Liebenberg, Paul Rudnick, Ken Williamson, and Jerg Jergensen. (Bill Regan Photo)



The radially graded filter experiment produced magnificent pictures in 1973 for the three-man team of (from left) Sid Stone, Darrell Call, and Bill Regan, and a stunning photo of a coronal bubble in 1980. (Bill Regan Photo)

Design genius Bobby Strait (left), Joe Montoya, and Charles Keller study the coronal camera used on the expedition out of Madrid, Spain, in 1973. (Bill Regan Photo)

tion from a subsonic aircraft then and now, and he used the time to measure the polarization of the red, green, and yellow emission lines in the corona.

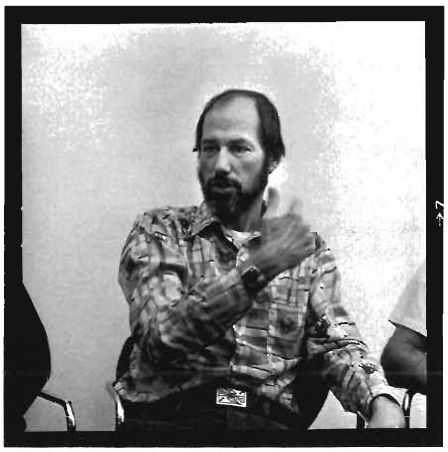
With the added time gained by the Concorde's keeping pace with the moon's shadow, Liebenberg triumphantly noted the onset and departure of totality as the speed of the plane made the moon appear to work as a slow camera shutter, slicing off only a tiny piece of the sun's atmosphere each second. For the first time, pulsations in light intensity, presumably due to traveling waves, were observed in the corona.

Aboard the 135, Sid Stone, a veteran of 1965, joined Bill Regan and Darrell Call in mounting a radially graded filter experiment that succeeded in recording on one photograph intensities ranging over a factor of 10,000, from the glaring base of the inner corona to the faint emanations from the outer corona at 10 solar radii.

The coordinated experiments of 1973 yielded such a wealth of information that the shipboard 1977 eclipse event was a profound disappointment. Los Alamos had tremendous cooperation from the crew of the vessel on which they set up shop off the coast of Mexico, but Murphy's Law struck with a vengeance: high clouds obscured much of the eclipse and failure of a power cord, dislodged by a stranger who opened a door near the experiment, put an end to the effort.

A quick and dirty mission was mounted in 1979. Los Alamos eclipsers heard that the USAF 4950th Test Wing was considering a request from a major television network to cover the mission in a 135 (now used primarily by the Air Force Weapons Laboratory) so that the eclipse could be filmed live. Superb ob-

IN FLIGHT:
THE STORY OF LOS ALAMOS ECLIPSE MISSIONS



Charles Keller: We mounted a quick and dirty mission... for the first time in 16 years of eclipse observation continuous intensity data out to 20 solar radii were collected... North Dakota '79

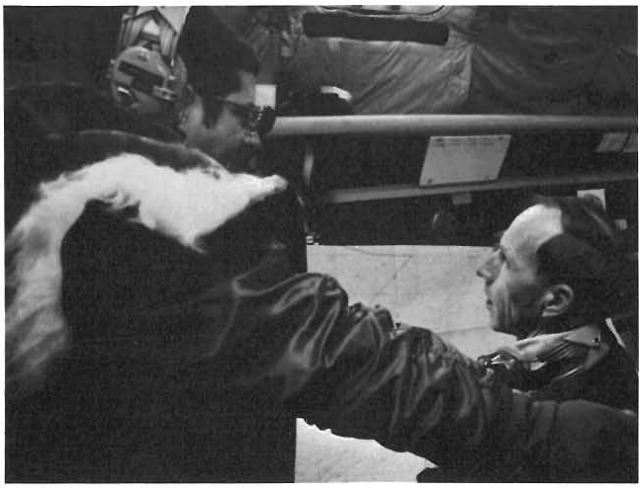
servation possibilities at 37,000 feet above North Dakota and a relatively short flight, plus renewed interest by both the Air Force and the Laboratory resulted in a fruitful mission, although the live TV coverage had to be scrubbed and delayed footage substituted. The flight was also an excellent training opportunity for the 1980 eclipse over Africa. Senior scientist Charles Keller produced some remarkable image-enhanced photographs out to 12 solar radii. For the first time in 16 years of eclipse observation, continuous intensity data out to 20 solar radii were collected. For such a crash effort—only five weeks' preparation—it was a good mission. It also aroused more public interest than any eclipse in recent years, perhaps because much of North America was in on the act.

By contrast, hundreds more man hours were devoted to preparing for the 1980 eclipse, which climaxed on February 16. The Los Alamos contingent flew to Spain on the first leg, and the time was used by Brook Sandford to test an onboard microcomputer designed to interface with his ingenious new experiment to make direct measurements of electron temperatures in the solar corona—the first experiment of its kind. The computer's magnetic disk appeared to function well, but during the flight to Nairobi, the jumping off point for the mission, it began recording intermittently. A back-up unit was installed while the plane was on the ground at the Jomo Kenyatta Airport, and it too functioned well, until a Kenyan ground crew member inadvertently pulled the plug to the power source on the aircraft and destroyed the disk. Sandford and the rest of the crew worked feverishly the night before the eclipse to get the experiments in order. The original disk was put back into the system; it worked, but failed as totality approached, a victim of the stress of turbulent air travel. Sandford glumly watched what appeared to be good data flickering across his screen, and was unable to record the information for later analysis.

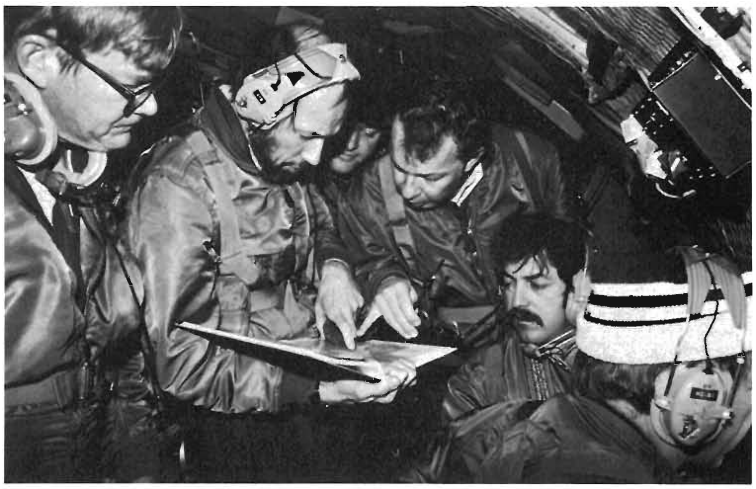
Difficulty with the mirror tracking system in the airplane caused Bill Regan and Carl Lilliequist's radially graded filter experiment not be be centered precisely on the sun. The experiment failed to record the entire corona, but it did produce a marvellous photograph of a hydrodynamic eruption from the sun's surface that formed a huge bubble in the corona.

There were other successes, too. Liebenberg and an interdivisional team, which included Ed Brown, Pete Murray, Mort Sanders, and Bob Kennedy, successfully captured very-high-resolution data on emission lines of highly ionized iron and calcium. Bob Brownlee and Paul Mutschlecner and Kitt Peak Observatory's Don Hall, veteran of the SST 1973 mission, recorded excellent data on the thermal properties of dust in the F corona. Bobby Strait, Charles Keller, and Joe Montoya fielded the photographic polarimetry experiment, first flown in 1970 and designed to determine electron densities in the corona. They captured the first detailed and complete record of a hydrodynamic solar eruption from the base of the corona out to 7 solar radii—a phenomenon seen previously by satellite, but only down to 1.5 solar radii.

Failure and success are common in eclipse study, and there are other commonalties. There is the indescribable feel-



Above, Practice flight arrangements are discussed before the 1980 mission by USAF Maj. Mike Bartlett and the Laboratory's Bill Roach. (Johnnie Martinez Photo)



Above, plotting a course for home after the 1980 eclipse mission are, from left, Bob Jeffries, Charles Keller, Brook Sandford, Paul Mutschlecner, Joe Montoya, and Carl Lilliequist. (Johnnie Martinez Photo)

Four members of the 1965 American Samoa flight (left to right) Ken Williamson, Bill Ogle, Don Liebenberg, Walt Wolff, and the Rube Goldberg. (Bill Regan Photo)

IN FLIGHT:
THE STORY OF LOS ALAMOS ECLIPSE MISSIONS

Bill Regan: The light was so eerie. . . like a mercury vapor lamp on a dark night. . . Lilliequist was on his back, still struggling to center the disc as totality approached. . . Africa '80

ing of being so close to the awesome display, untroubled by weather or other interference, as the plane begins the race-track orbits designed to place it in the path of totality. Typically, three to seven orbits are flown after first contact, when the moon's shadow touches the disc of the sun and the plane turns eastward to the path of the shadow's center line.

In 1980, at least, the frantic preparations continued up to and into totality, as Carl Lilliequist struggled beneath his equipment trying to center a 2.54-mm disc on a second 2.54-mm image, and was only partly successful. Sandford mourned the loss of his data, and in every section of the noisy, chilly, padded interior of the plane, teams conversed tersely on closed-circuit intercom systems linked only to each other and to a frequency used by the flight crew to relay directions in case of emergency.

As the shadow deepened, the chill spread, and the light became ominous, a blue-tinged glow like a mercury vapor lamp on a very dark night. The General Dynamic Company's crew struggled with the obstreperous tracking system up to and beyond totality. Bob Jeffries, watching a video screen monitor for the master tracker, cried "Lock it in, that's it!" The moment had arrived. Months of gut-wrenching effort, the tension, the successes, and the failures were all forgotten in the ensuing minutes, as each team worked in the eerie twilight. For just a few moments, as the light returned, there was silence, broken only by the overwhelming noise of the less than luxuriously appointed aircraft. Then the analysis began.

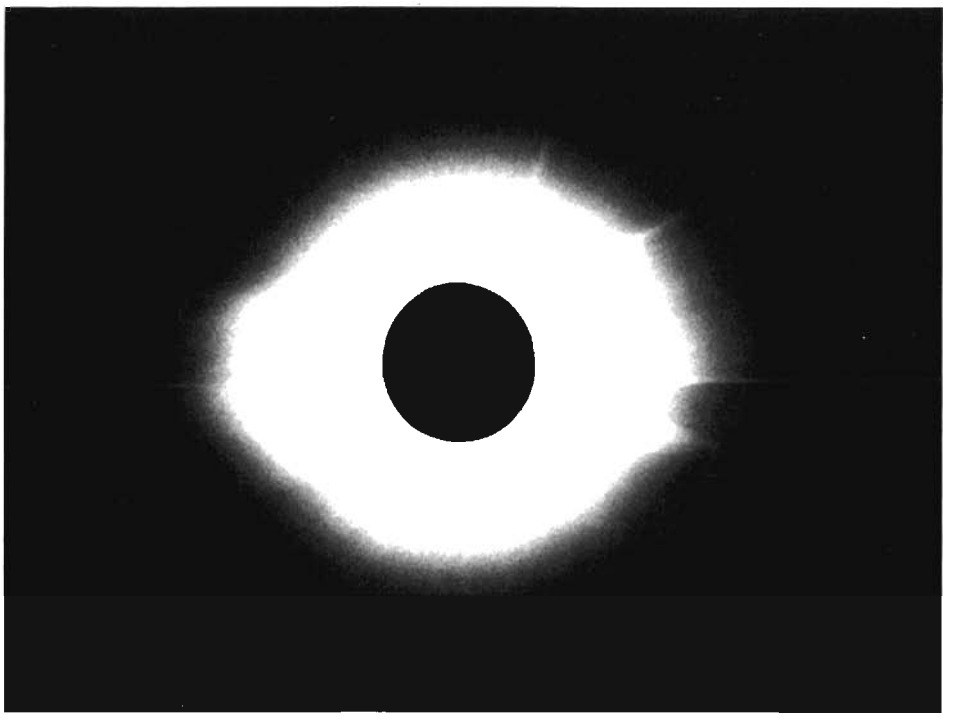
So exciting is the field of astrophysics that the media coverage tends to make the missions sound like fun and games, and the considerable ingenuity and months of hard work are frequently overlooked. Nevertheless, the increasing sophistication of both equipment design and diagnostic techniques has improved the accuracy of coronal data and added to the slim store of knowledge of the star on which our existence depends.

For instance, Art Cox, working with Don Eilers, put together the computer code for predicting aircraft interception of solar eclipses, and the model has been used by all other aerial observers. Ralph Partridge designed and has continually improved photo tracking systems that permit experiments to lock on to the eclipsed sun and hold it centered with a maximum error of six arc seconds—this is the equivalent of holding perfect aim with a rifle for five minutes on a penny-sized target a mile away.

Special lens systems have been produced by Berlyn Brixner using a lens design computer code. Stabilized platforms and servo systems—major contributions—have come from Joe Perry, Ed Brown, and Bobby Strait. Strait, called the magician for his expertise in designing and adapting equipment for aerial eclipse study, is deemed essential to the success Los Alamos has enjoyed. Strait's idea for individual tracking systems that could make future missions even more successful is something that might apply to the eclipse over Indonesia in 1983.

Cox, Brownlee, Keller, Roach, Sandford, Regan, Carl Lyon who smoothed the logistic progress of several trips and handled communications—the list goes on. The veteran Liebenberg, who has covered 11 mission since 1954, including at least one on his own time and at his own expense, has logged more totality observance than anyone in history—117 minutes and 11 seconds, and as Keller *et al.* are fond of pointing out, "sometimes Liebenberg even remembers to put film in the camera, remove the lens cap, and refrain from jiggling the equipment at the wrong moment. . ."

That pretty much sums up the spirit of the flying eclipse corps at Los Alamos: slightly irreverent and doggedly determined. The word most often used by those who know is—dedicated ■



Scientific Collaborators on Los Alamos Airborne Eclipse Missions

1965 Charles L. Hyder, University of California
1970 Merton M. Robertson, Sandia Laboratories
1972 Jacques M. Beckers, Sacramento Peak Observatory
Apostolos Frangos, Scientific Group for Space Research, Greece
1973 Jacques M. Beckers, Sacramento Peak Observatory
Warren N. Arnquist, University of California, Los Angeles
Alberto Righini, Arcetri Astrophysical Observatory, Italy
Manuel Lopez Arroyo, Madrid Observatory, Spain
Ervin C. Woodward, Lawrence Livermore Laboratory
Robert E. Donaldson, Lawrence Livermore Laboratory
Jose L. M. Cortez, Lawrence Livermore Laboratory
1979 R. Kent Honeycutt, Indiana University
1980 Donald L. Hall, Kitt Peak National Observatory
R. Kent Honeycutt, Indiana University

HIGH EXPLOSIVES

The interaction of chemistry and mechanics

by William C. Davis

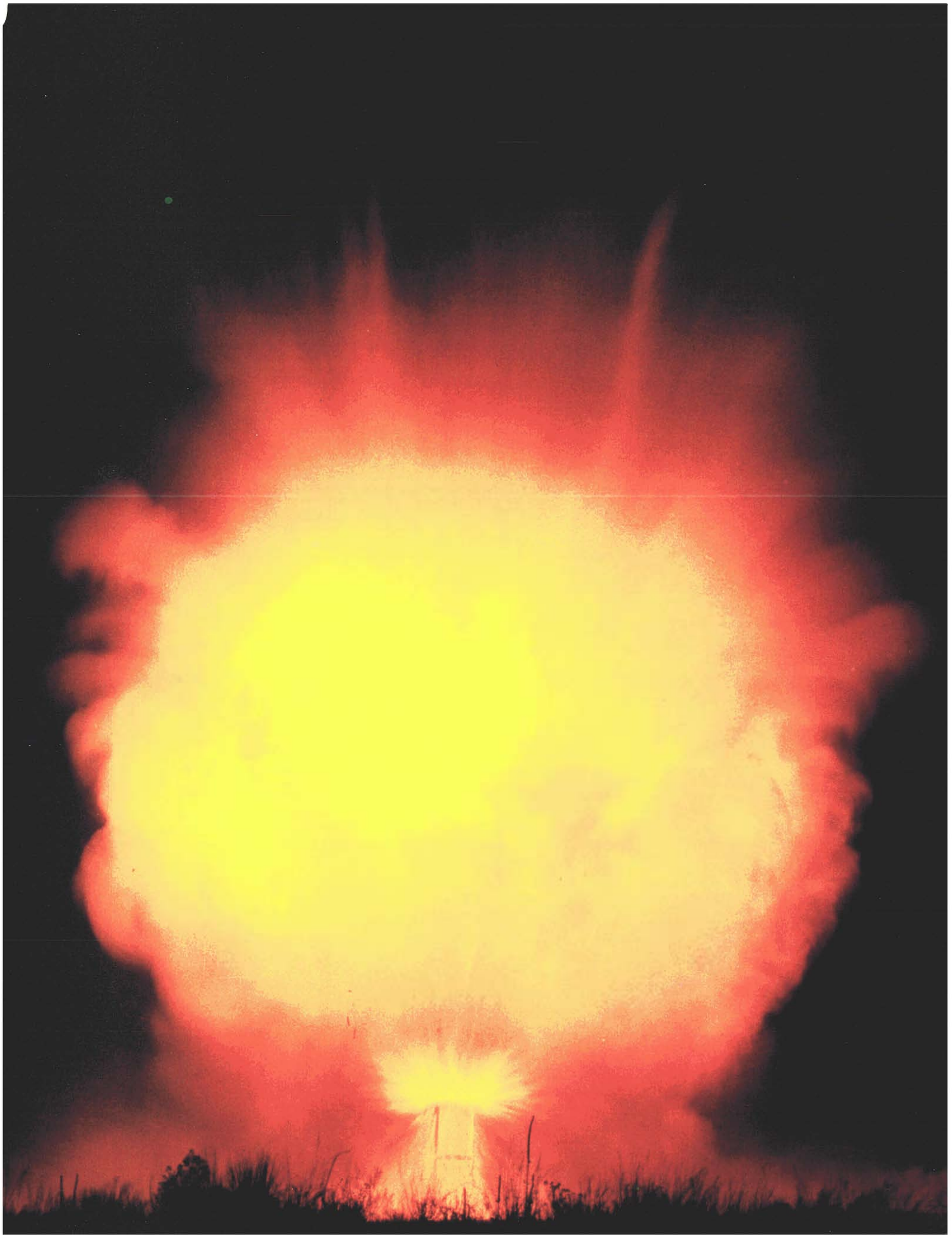
Although explosives have been known for over a thousand years, the science of explosives is still very young. We are only beginning to understand the nonlinear interaction between chemistry and fluid mechanics that produces the rapid energy release known as detonation.

The science of high explosives is basically a coupling of chemistry and fluid mechanics. While each of these fields is in itself quite well-developed and sophisticated, high-explosives science is surprisingly primitive. The fluid-mechanical phenomenon of detonation is reasonably well understood, but the detailed chemical reactions and thermomechanics that cause a detonation are still largely a mystery. For many explosives, even the final chemical composition after detonation is not known accurately and the reaction mechanisms are only guessed at. Similarly, while it is clear that some of the most energetic explosives would not detonate at all were it not for their nonuniform mechanical response to shock waves, the micromechanics of explosive materials is not nearly so well understood as that of metals.

There are two basic reasons why the science of explosives is relatively undeveloped, and an understanding of them indicates why the next decade is likely to produce a dramatic increase in our understanding. First, as is obvious, measurements in the interior of a detonating explosive are extremely difficult. Whereas experimental methods have existed for many decades that can characterize chemical reactions under normal conditions, they are of little use under the extremes of temperature and pressure generated in explosives. Similarly, standard techniques to study the mechanics of flow in metals are of little use when the relevant stress and strain rates produce a violent reaction in the studied material. As a result, few academic institutions have deemed it fruitful to establish a research program in explosives.

The second reason is that applications of explosives technology in the past have not placed a high premium on understanding the details of the detonation phenomenon. Most explosive applications are in the fields of excavation, mining, or conventional munitions. All of these have well-established, albeit crude, "rules-of-thumb" as regards quantities and configurations of explosive required to accomplish the task. Hence there has been little pressure to establish an extensive industrial research base.

Recently, however, there has been a significant increase in both the capability and motivation for expanded research in explosives science. Modern instruments, particularly those employing lasers as probes, have made it possible to selectively investigate chemical phenomena on time scales of less than 100 picoseconds. Such measurements are



stimulating a rapid growth in experimental and theoretical techniques that should soon be applied to explosives. The need for understanding the details of the energy release in explosives is also increasing rapidly. Explosive systems are being demanded that function with increased precision and efficiency and at the same time maximize safety. For example, *in situ* retorting of oil shale or chemical mining of scarce minerals will require blasting techniques that can produce a preselected distribution of small cracks or fragments rather than just a displacement of the ore to facilitate mechanical mining. As another example, modern munitions are increasingly reliant on the ability to focus the energy from explosives to defeat well-protected targets such as tanks and armored vehicles. Designs of such systems are so sophisticated that a giant computer is needed to optimize them, and the behavior of the explosive must be quite accurately predicted. This combination of demand for a refined explosive technology and the availability of new research tools should produce a dramatic improvement in the state of explosives science. In this article we review our understanding of explosives as it has evolved by bursts and starts from the turn of the century to the present. We begin with empirical observations and trace the development from simple to more complex fluid-dynamical models of energy release and propagation.

Because the time scales for energy release are so fast, the simplest model that ignores all details of the chemistry has been remarkably successful in predicting the performance of many explosives presently in use. But the more complex models give us insight into the nonlinear interaction between chemistry and fluid dynamics that is at the heart of the detonation process.

Detonation Physics

Our understanding of explosives begins with empirical observations. Figure 1 shows a block of explosive as it detonates. The detonation wave spreads out from the point of initiation almost like a Huygens construction. The wave velocity is supersonic and almost constant for a particular explosive, but it varies from one explosive to another, depending primarily on the composition and density of the explosive. For most explosives, the detonation velocity is affected little by the time it has run, the size and shape of the block, or the curvature of the detonation wave front. Because the detonation wave velocity is faster than the velocity of sound in the explosive, the material in front of the wave is absolutely unaffected until the detonation wave passes through it. In particular, a second detonation wave in the block propagates independently of the first wave until the two intersect.

The wave front is the moving surface that separates explosive material in motion from stationary material. In solid or liquid explosives, the pressures just behind the front are very high, a few hundred thousand atmospheres (a few tens of gigapascals), and the temperatures are from 2000 to 4000 K.

The high temperatures and pressures are produced by the very rapid release of chemical energy in the explosives. Typically the chemical reaction is 90% complete in 10^{-6} to 10^{-9} second. The energy goes into the motion of the explosive products, creating the high pressures and temperatures necessary to drive the reactions. In other words, the inertia of the explosive itself provides the confinement necessary to maintain the conditions for the fast chemical reaction rates and the self-sustaining propagation of the detonation wave. The distinguishing feature of detonation is the self-inertial confinement of the chemical reaction. Thus, there is an intimate relationship between chemistry and mechanics, and neither can be treated as an independent process in a realistic detonation model.

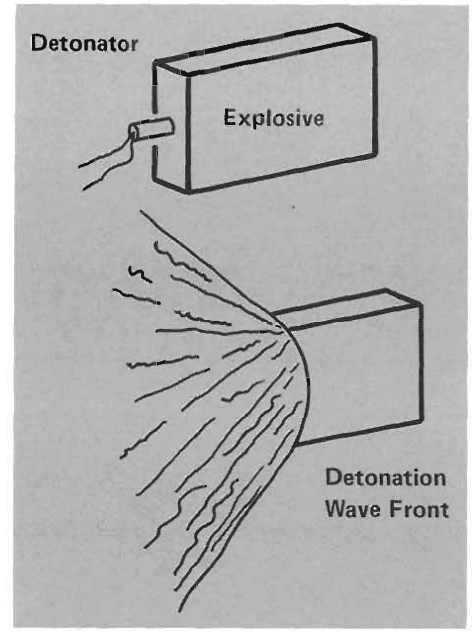


Fig. 1. Detonation of a block of explosive. The detonation wave spreads from the point of initiation as a nearly spherical wave, almost like a Huygens construction. The wave is supersonic (faster than sound in the unreacted explosive), so there is no signal of any kind ahead of the detonation. The entire chemical reaction takes place in a thin layer just behind the wave front.

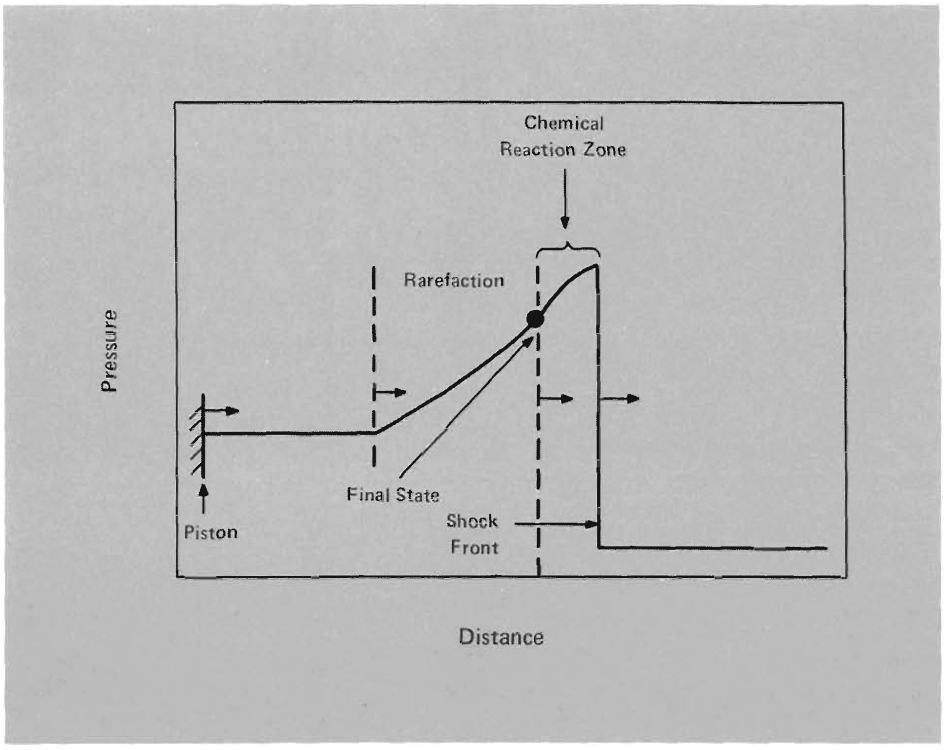


Fig. 2. Plot of pressure versus distance for a detonation wave. The shock wave, at the right, is the leading element of the detonation wave. The explosive behind it, heated by the sudden compression, begins its chemical reaction there. Pressure falls as the reaction proceeds, and reaction is finished at the point marked final state. Behind that point are a rarefaction and a constant-state region; they reduce the pressure and particle velocity to match the motion of the external confinement, shown here as a piston. The reaction zone, between the shock wave and the final state, is a subsonic flow region; energy liberated in it can flow forward to drive the shock wave. The region behind the final state is a supersonic flow region; neither energy released there nor any perturbing waves can move forward to affect the reaction zone or the shock.

The explosive and the inert material it drives are usually solids, but the detonation pressures are so high that material strength may be neglected and the propagation of energy may be understood through the equations of reactive fluid dynamics. Furthermore, energy transport by heat conduction, viscosity, and radiation is negligibly small compared with the transport by motion. The theoretical basis for treating one-dimensional detonation (the ZND theory) is a fluid-dynamical model that was arrived at independently by Zeldovich, von Neumann, and Doering. Figure 2 shows a plane, steady, unsupported detonation wave predicted by the ZND theory. The detonation, initiated by a pressure pulse from the piston at the left, is called unsupported because the piston velocity is less than the fluid or particle velocity of the explosive products. The detonation front is a shock wave, supersonic relative to the material ahead of it, so no signal precedes it. Compression heats the explosive, and rapid chemical reaction follows. Finally, reaction is complete, and the product gases expand as an inert flow. The inert flow of the explosive products is affected by the surrounding inert materials. In other words, the inert flow must match the boundary conditions provided at the left of Fig. 2 by the piston and at the right by the final state of the explosive products at the end of the reaction zone.

The speed of the chemical reaction rates and the seemingly independent propagation of the detonation front lead naturally to a division of the problem into two parts: (1) the study of the chemical reaction zone where the detonation process goes on and (2) the study of the acceleration of inert components, such as the metal of a hand grenade or the rock around a borehole, by the expansion of the explosive gases after the reaction is finished. Although the two parts are interrelated, until recently they have been treated as separate problems.

In most practical cases, the chemical reaction zone is so thin compared to the size of the explosive charge that its length is neglected completely in explosive performance calculations. We assume that the reaction takes place instantaneously at the detonation front and calculate the expansion of final explosive products as they push whatever material may enclose them. If this idealized calculation is compared with measurements, the effect of finite reaction zone length (or finite time of chemical reaction) appears as a small rise in pressure or velocity at the detonation front.

In many cases, this simple way of treating detonation phenomena is sufficient for determining the equation of state of the explosive products and for calculating the inert flow and explosive performance. However, modern applications of explosives have stimulated attempts to treat the entire problem as a whole, to learn in more detail the chemical reaction rates in the reaction zone and how they are affected by changes in the boundary conditions, by the addition of new materials, and by the effects of inhomogeneities and transverse waves. These details became important in applications of nonideal explosives, such as TATB and other insensitive high explosives, that have relatively long reaction times.

Our methods may baffle the newcomer to the field unless he or she recognizes the difficulty of studying det-

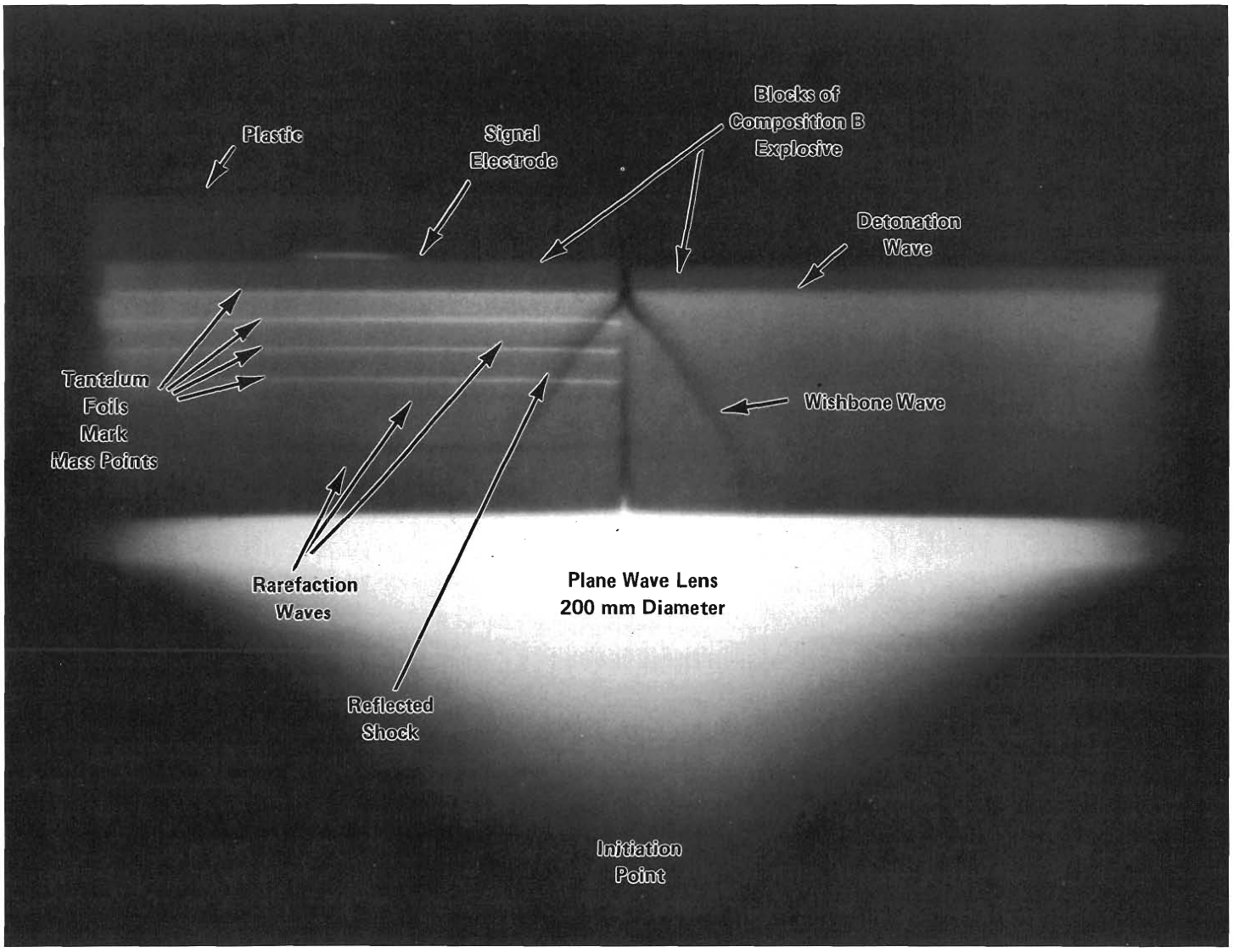


Fig. 3. Radiograph of detonating Composition B explosive, taken with the PHERMEX flash x-ray machine. The detonation wave, which has advanced from the initiation point up through the conical plane wave lens (see Fig. 7) and almost to the top of the Composition B blocks, appears as a light streak across the radiograph. Light areas on x-ray pictures indicate high density, darker areas indicate low density. The high density remains of the plane wave lens also appear as a very light area.

Above the detonation wave is a thin layer of unreacted explosive. Below the detonation wave we can see various waves in the explosive product gases: a wishbone wave from the gap between the two blocks, rarefaction waves from the air gaps around the tantalum foils, and the reflected shock wave hardly visible in the remains of the plane wave lens. The foils mark distinct mass points in the explosive. The spacing between them decreases as the explosive is compressed by the passage of the detonation wave.

onation phenomena experimentally and the necessity of inferring an explosive's material properties and chemical reaction rates from theoretical models and indirect measurements. For example, there is no way to take a sample of material from the reaction zone to see how the chemistry is progressing. There is no way (yet) to study the chemicals and their reactions at the conditions of pressure and temperature in the detonation reaction zone in any sort of laboratory apparatus. Pressure gauges, velocity gauges, and thermometers for the study of conditions in explosions are being developed, but are not yet satisfactory. In addition to experimental difficulties like making electrical connections that are not destroyed by the violent motions of the explosive products, the massive apparatus perturbs the flow so much that the system is changed and the measurement is meaningless. Perhaps developments, particularly in laser spectroscopy, will eliminate these difficulties.

With current instrumentation we can only follow the motion of an inert material driven by explosive or follow the positions of the shock front and other waves. Pulsed x-ray photographs can show positions of matter or waves, and cameras and electrical contacts can record what happens at accessible surfaces. (See Fig. 3.) This very limited information makes the interpretation of measurements depend heavily on theory and models.

A Simple Theory

An explosive's performance or usable energy is determined by the expansion of product gases following completion of the chemical reaction in the explosive. Thus, to calculate performance we must know the state (pressure and particle velocity) of the materials at the end of the reaction zone and their equation of state, that is, how the pressure varies with the particle velocity of the product gases during adiabatic or free expansion. With no direct measurements of these material properties nor fundamental theory to help us, how do we proceed?

The usual practice has been to apply a simple fluid dynamical model of detonations known as the CJ theory. This generalization of the theory of shock waves provides a framework for inferring the relevant material properties from standard detonation experiments. It relates the detonation wave velocity to the properties of the gases behind the detonation wave front.

The CJ theory assumes that all chemical energy is released at the detonation front so the reaction zone in Fig. 2 has no thickness. The detonation wave is thus approximated by a self-sustained supersonic wave traveling through the explosive at constant velocity. We are interested in determining four quantities: the velocity of propagation D, and the pressure p, density ρ, and particle velocity u behind the wave front.

For shock waves in inert materials, the three equations of conservation of energy, momentum, and mass across the shock front, the so-called jump conditions, are sufficient to determine the shock velocity U in terms of the variables p, ρ, and u.

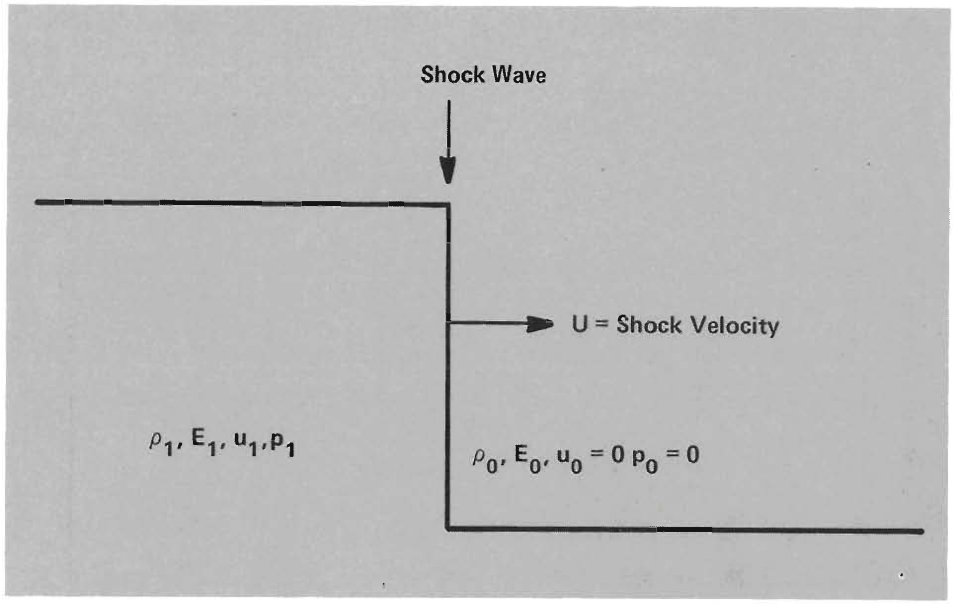


Fig. 4. Wave propagating to the right into stationary material.

In the CJ theory for detonation waves, the jump conditions apply, but because energy is released at the front, making the wave self-propagating, an additional condition is needed to determine the detonation wave velocity D. The condition, postulated by Chapman and Jouguet around the turn of the century, is known as the CJ condition. Before discussing it, we review the jump conditions for simple shock waves and show how chemical energy release in explosives complicates the analysis of the conditions behind the wave front.

A plane shock wave propagating in a medium initially at rest is shown in Fig. 4. We use U for wave velocity, u for particle velocity, p for pressure, ρ for density, and E for specific internal energy. Subscripts 0 and 1 indicate the regions before and after passages of the shock, respectively. Velocities are positive for motion to the right.

We treat a tube of area A for a period t. During the period t, the wave front, moving at velocity U, passes over a mass of material equal to $\rho_0 AUt$. During the same period, the fluid element located at the wave front at $t = 0$ moves a distance ut, so material passed over by the wave front is now within a volume $A(Ut - ut)$, and its mass is $\rho_1 A(U - u)t$. Its quantity has not changed, so we equate the two expressions for the mass. Cancelling out At, we obtain

$$\rho_0 U = \rho_1 (U - u_1) \ . \qquad (1)$$

The passed-over material is accelerated to velocity u_1, so its momentum changes from zero to $\rho_0 AUtu_1$. The force acting on the material (if p_0 is assumed negligible) is $p_1 A$, and it acts for a time t, so the impulse is $p_1 At$. Equating the impulse and the change of momentum (and again cancelling At), we obtain

$$p_1 = \rho_0 U u_1 \ . \qquad (2)$$

The internal energy changes by $\rho_0 AUt(E_1 - E_0)$, the kinetic energy changes by $1/2\ \rho_0 AUtu_1^2$, and the work done on the material is the force times the distance, $p_1 Au_1 t$. The change in energy and the work done are equal, so we equate them. Cancelling At and dividing by $\rho_0 U = p_1/u_1$ from Eq. (2), we obtain the conservation of energy equation

$$E_1 - E_0 = 1/2\ u_1^2 \ . \qquad (3)$$

Equations (1), (2), and (3) are called the Rankine-Hugoniot shock relations, or the jump conditions. The Rankine-Hugoniot relations along with an equation of state for the material (an expression for E in terms of p and ρ) define a smooth curve in the p-ρ or p-u plane called the Hugoniot curve. The curve describes all the states of the material that can be reached by the passage of a single shock wave. The Hugoniot curve is steeper than isotherms and isentropes for the material.

The jump conditions, Eqs. (1) - (3), derived under the assumption that the material has no strength, are often applied to metals because the material strength is small relative to the shock forces and the shear forces quickly relax to zero. This approximation is good at high pressures, but the deviations may be large at low pressures.

Although we don't have a theoretical equation of state for the shocked material, we know empirically that the relationship between the pressure p and the particle velocity u on the Hugoniot curve is described adequately by a few terms of the series expansion

$$p_1 = \rho_0 (c u_1 + s u_1^2 + \ldots) \ , \qquad (4)$$

where c, s, and the coefficients of higher powers of u_1 are constants. These constants are determined experimentally. In this simple approach that treats the shocked material as nonviscous fluid, the Hugoniot curve given in Eq. (4) is a complete statement of the important material properties. Comparison with Eq. (2) shows that the velocity of the shock wave in the material is given by the expansion

$$U = c + s u_1 + \ldots \ . \qquad (5)$$

Thus, the constant c is the sound speed since a very weak wave, with u_1 negligible, propagates with velocity c.

For a fixed shock-wave velocity U, Eq. (2) describes a line in the p-u plane along which momentum is conserved, called the Rayleigh line. The slope of the line is given by U times the initial density ρ_0.

Figure 5 shows the Hugoniot curve for a particular material and the Rayleigh line for a particular shock velocity U and initial density ρ_0. The intersection of the two curves gives the state of the material behind the wave front with shock velocity U.

Now we turn from shock waves in inert material to detonation waves in explosives. We are going to ignore the thickness of the reaction zone and approximate the detonation wave by the discontinuity shown in Fig. 4. Now the material to the right is unreacted explosive, and the material to the left is completely reacted explosive products. The jump conditions apply just as they do for shock waves, but the Hugoniot relation describing the final state of explosive products behind the detonation wave front has an additional term reflecting the fact that energy is released as the chemical bonds are rearranged. The Hugoniot relation becomes

$$p_1 = \rho_0 (nQ + c u_1 + s u_1^2 + \ldots) \ . \qquad (6)$$

Here Q is the specific chemical energy of the explosive, and n is another material constant to be determined experimentally. The Hugoniot curve for an explosive is shown in Fig. 6, along with the Rayleigh line that is just tangent to it. We compare Figs. 5 and 6 to illustrate the differences between inert materials and explosives. An inert material has a minimum shock velocity U = c, representing a sound wave with zero pressure, and a unique shock wave pressure and particle velocity for any faster wave. Notice that the Hugoniot curve for the explosive does not pass through p = 0. The explosive has a minimum detonation wave velocity, but the pressure at that velocity is not zero but large, and for any higher shock velocity the Rayleigh line and the Hugoniot curve intersect at two points rather than one so that the final state for the shocked explosive is not uniquely determined by the jump conditions and the equation of state.

In the late 1890s, Chapman, in England, and Jouguet, in France, eliminated this ambiguity. They studied the propagation of waves in the flows that might follow a detonation front and made the plausibility argument that an unsupported detonation proceeds at the minimum detonation velocity, which is the

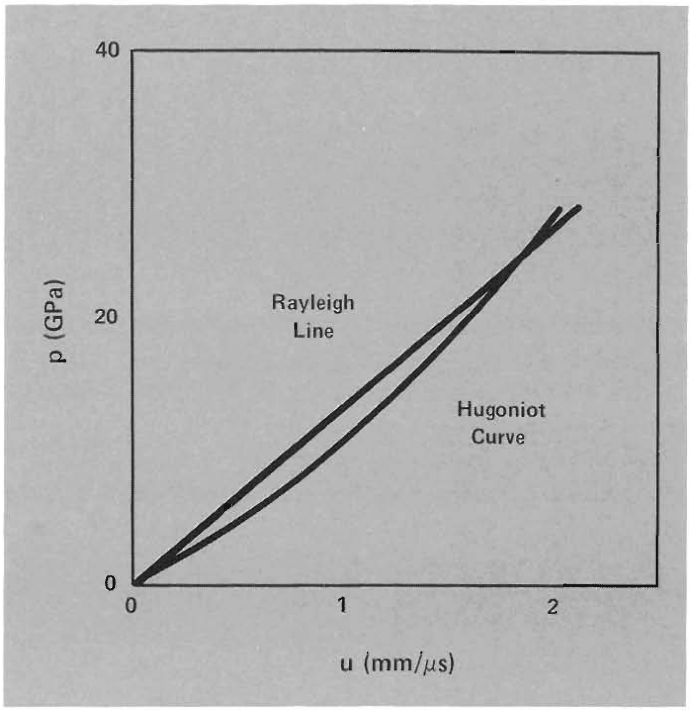


Fig. 5. The Rayleigh line and Hugoniot curve shown in the p-u plane. The material is an inert, NaCl; $\rho_0 = 2.165$ g/cm^3, $c = 3.528$ mm/μs, and $s = 1.343$. The Rayleigh line is drawn for $U = 6.05$ mm/μs.

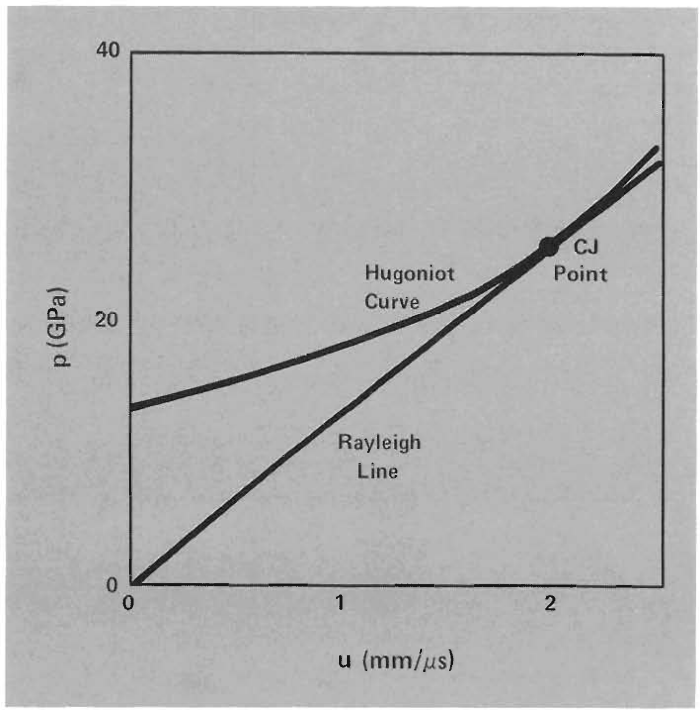


Fig. 6. The Rayleigh line and Hugoniot curve for an explosive with a detonation velocity D of 8 mm/μs, $\rho_0 = 1.6$ g/cm^3, and $p_1 = 25.6$ GPa. The Rayleigh line, drawn for $D = 8$ mm/μs, is the slowest one (least slope) that intersects the Hugoniot curve. The single intersection, the tangent point, is the CJ point.

unique velocity given by the Rayleigh line tangent to the Hugoniot curve. This is called the CJ condition.

The selection of the unique minimum velocity is in agreement with the observation that detonations have a well-defined velocity, determined by the composition and density and little affected by any external conditions. The plausibility is increased because simple thermodynamic arguments show that the CJ point is a sonic point. That is, the detonation velocity D at that point is given by

$$D = c' + u' , \qquad (7)$$

where c′ is the local sound velocity and u′ is the local particle velocity. Thus, any signal, such as energy liberated at that point, propagates forward at the wave velocity, just keeping up with the front and not overrunning it. Farther back in the rarefaction region the flow is supersonic, and any signal falls farther and farther behind the front, so the pressure decrease in that region does not interfere with the propagation of the detonation wave.

Thus, the simple theory gives a prescription for the final state of reacted explosive behind the detonation front—namely, the CJ point on the Hugoniot curve.

Determining the CJ Point and Hugoniot Curve

The CJ point is the starting point for calculations of inert flow behind the detonation front. This point is determined by measuring the detonation velocity and the Hugoniot curve for a

given explosive.

During the 1950s, W. E. Deal of Los Alamos carried out an extensive experimental program to determine the Hugoniot curves [the constants in Eq. (6)] for several explosives. Because computing power was very limited, Deal had to design experiments that could be analyzed by using pencil and paper. He simplified the required analysis by using plane detonation waves in the experiment.

Figure 7 shows the plane wave lens for Deal's experiment and Fig. 8 shows the diagnostic part. The plane detonation wave in the explosive reaches the inert material simultaneously over the interface and drives a plane shock wave through the inert material. The free surface of the inert material moves upward, driving the shim against the Plexiglas and compressing and heating the argon in the gap to produce a brief flash of light. The free surface moves only a very short distance to close the two outer gaps, but it moves an extra distance d to close the center gap. The film of an experiment recorded with a smear (or streak) camera is shown in Fig. 9. Measurement of the time offset between the flash from the outer gaps and the flash from the center gap gives the time it took the free surface to move the distance d. Division of the distance by the time gives the free-surface velocity.

How does measurement of the free-surface velocity of the inert material determine the CJ point and Hugoniot curve in the explosive? The analysis involves determining the pressure and particle velocity of the inert material at its interface with the explosive. Deal did this through a series of experiments with thinner and thinner plates of inert mate-

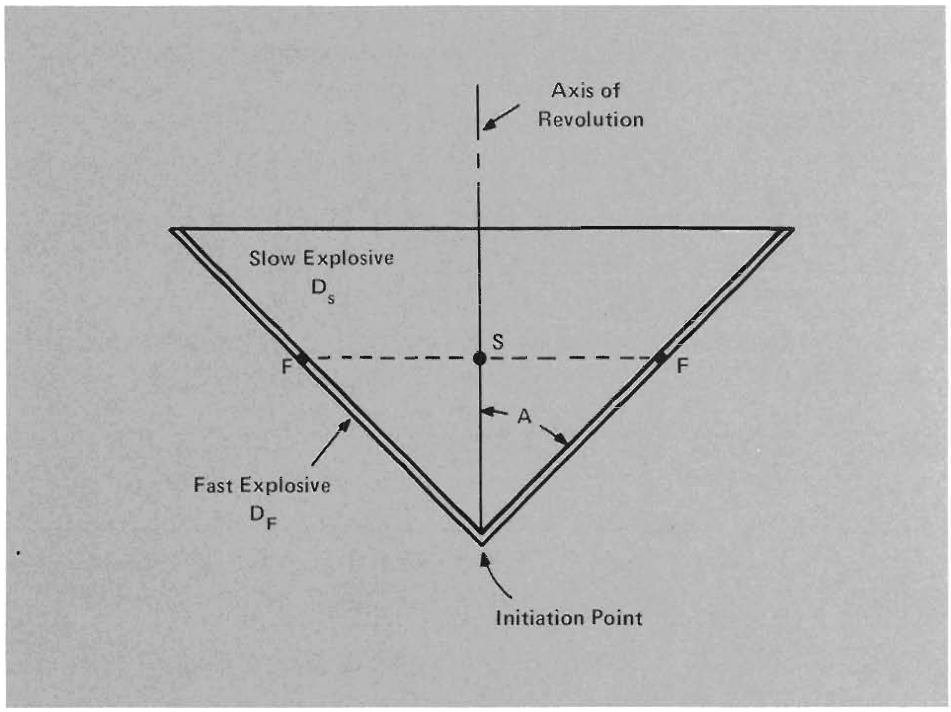


Fig. 7. Plane wave lens generates a plane wave from the point of initiation. It is made of a cone of slow explosive, with detonation velocity D_S. The conical surface is covered with a layer of fast explosive, with detonation velocity D_F. Initiation is at the point apex of the cone. After a time, the detonation wave in the fast explosive arrives at point F, and the detonation wave in the slow explosive at point S. If the cone angle A is chosen so that cos $A = D_S/D_F$, points S and F will lie in a plane. Because the fast explosive initiates the slow explosive as it proceeds, the detonation wave in the slow explosive is plane at every level. The lens can be used to initiate a plane wave in the test explosive for the experiment.

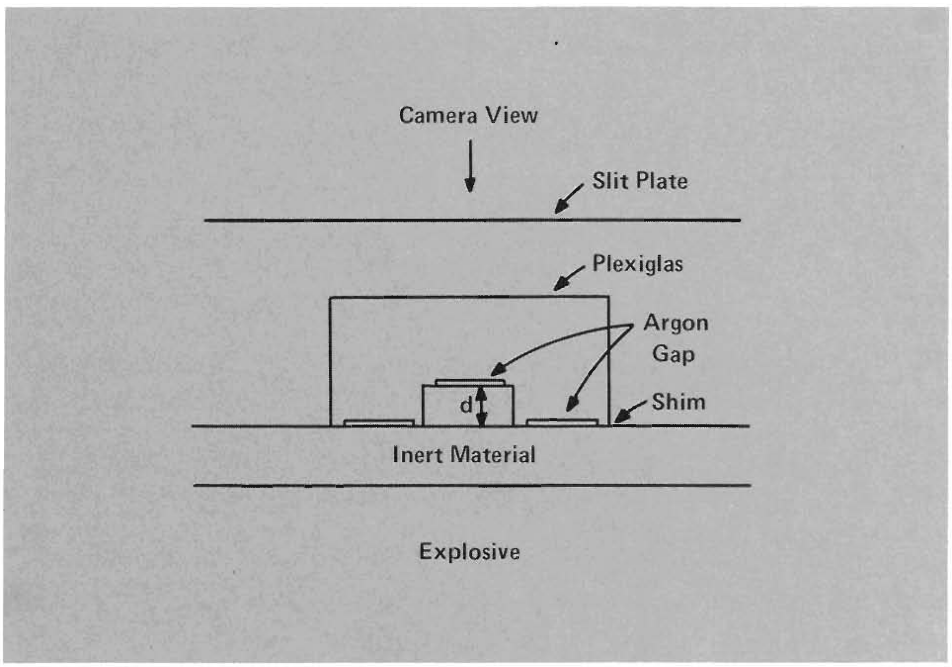


Fig. 8. The Plexiglas block assembly used by Deal for measurement of free-surface velocity of an explosive-driven plate. The argon gaps produce a flash of light when they are closed by the moving inert material. The time of flight of the inert material across the gap d is measured by the camera.

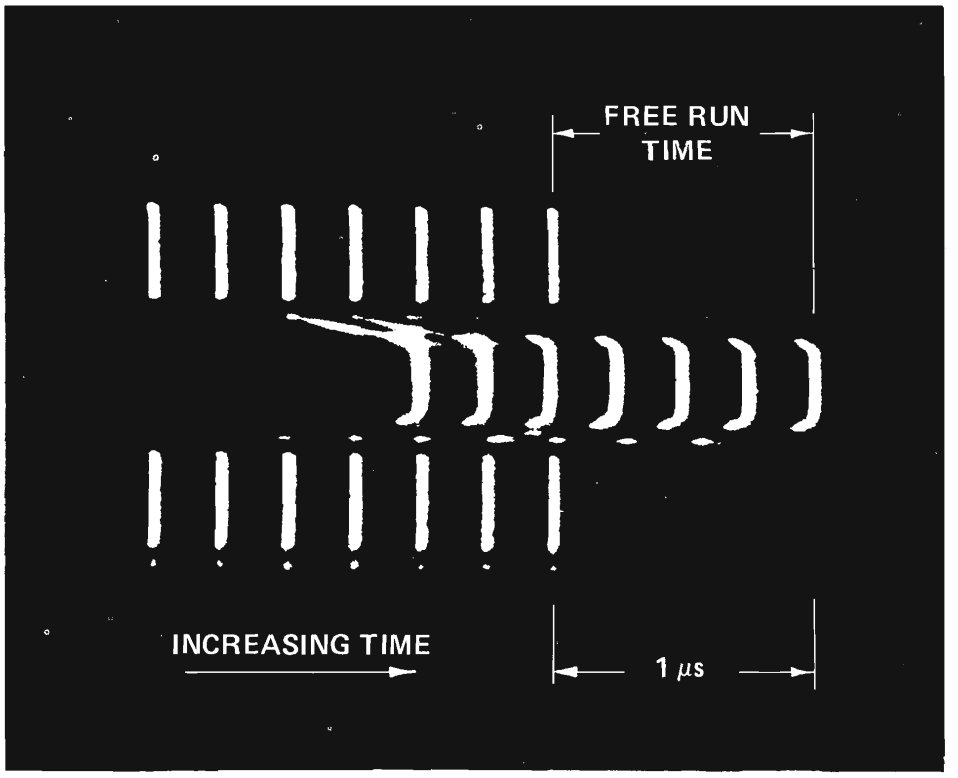


Fig. 9. Smear camera record showing seven records of free-run time of a metal plate accelerated by explosive. Free-run time is measured from each pair of side traces down to the corresponding central plate arrival time.

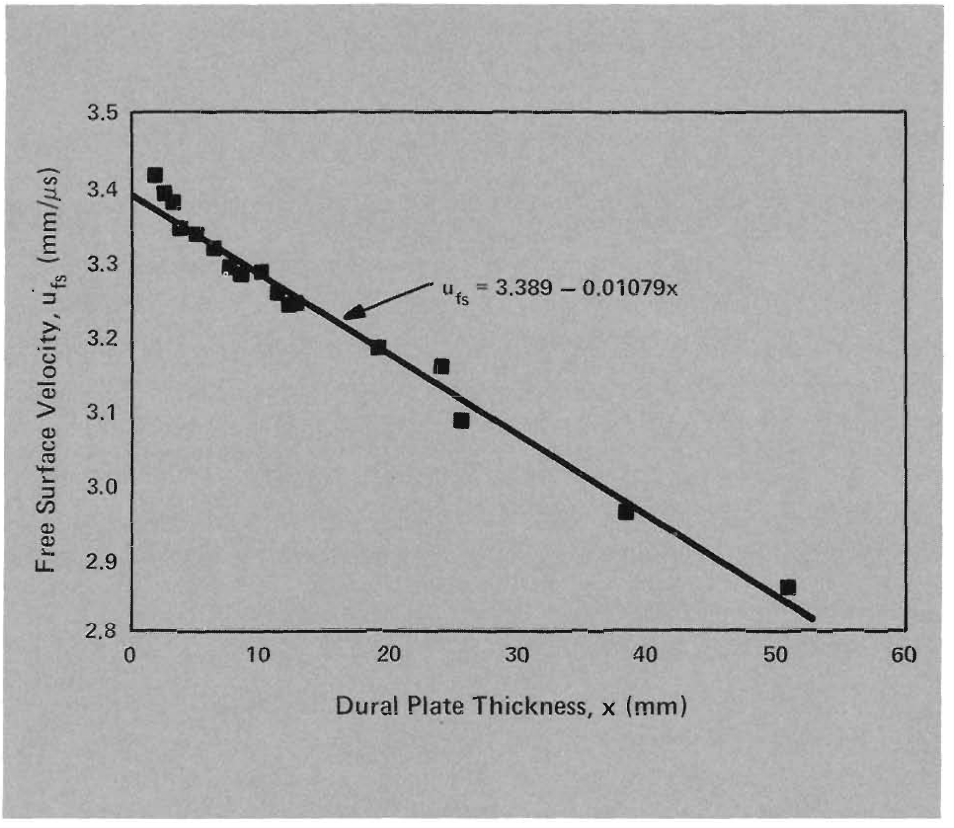


Fig. 10. Experimental values of free-surface velocity imparted to 24ST aluminum plates by Composition B explosive as a function of plate thickness. The line is the linear least squares fit to the data. The intercept of the fit with the ordinate corresponds to the free-surface velocity of a zero-thickness plate.

rial and extrapolated the results to determine the free-surface velocity of a zero-thickness plate (Fig. 10). From this velocity, he calculated the particle velocity in the inert material at the interface at the instant the shock wave was transmitted from the explosive to the inert material. The particle velocity in the inert material is very slightly less than half the free-surface velocity. The exact ratio is determined from known properties of the inert material on the Hugoniot and related curves in the p-u diagram. The particle velocity and the Hugoniot curve determine the pressure in the inert material at the instant of transmission.

The pressure and particle velocity determined for the inert material at the interface must be the same as those in the explosive products. Thus, these values describe one point on the p-u diagram for the explosive products. But is this the CJ point for the explosive?

The answer is yes only if the properties of the inert material and the explosive match exactly. In practice, they do not. If the inert material is more dense than the explosive, it reflects a shock wave back into the explosive; the shock wave moves the explosive away from the CJ state. If the inert material is less dense, a rarefaction wave goes back into the explosive, again moving the explosive away from the CJ state. Thus, the particle velocity and pressure determined for the inert material by experiment describe one possible state for the explosive products. Other possible states are determined by repeating the plate experiments with other inert materials more and less dense than the explosive. Figure 11 shows the experimental results. The smooth curve through these points must go through the explosive's CJ point.

Above the CJ point, the curve corresponds to the reflected shock Hugoniot curve for the explosive. Below the CJ point, it is an expansion isentrope. The CJ point, the intersection of the Rayleigh line with the curve in Fig. 11, is determined by measuring the detonation velocity in the explosive.

The measured curve in Fig. 11 differs slightly from the Hugoniot relation defined by Eq. (6). However, with some assumptions about the behavior of the explosive products, we can use the measured curve to determine the coefficients in that equation.

This method to determine the equation of state for an explosive obviously requires many experiments. Moreover, the experiments are very expensive. As computers became powerful enough to calculate the motion of metal driven by explosive products, workers in the field tried to devise less expensive ways to obtain the information. A group at the Lawrence Livermore National Laboratory (Kury, Hornig, Lee, McDonnel, Ornellas, Finger, Strange, and Wilkins) developed a method known as the cylinder test. They filled a cylindrical copper tube with explosive, detonated it at one end, and carefully measured the expansion of the copper wall. Figure 12 shows a flash silhouette photograph of such an experiment. Only the short section of the tube at the top remains unexpanded. The Livermore group compared their experimental data with a calculation using an assumed equation of state for the explosive and adjusted the equation-of-state parameters to obtain good agreement. This approach yields all the data below the CJ point in Fig. 11 with one relatively inexpensive experiment. The analysis must be done

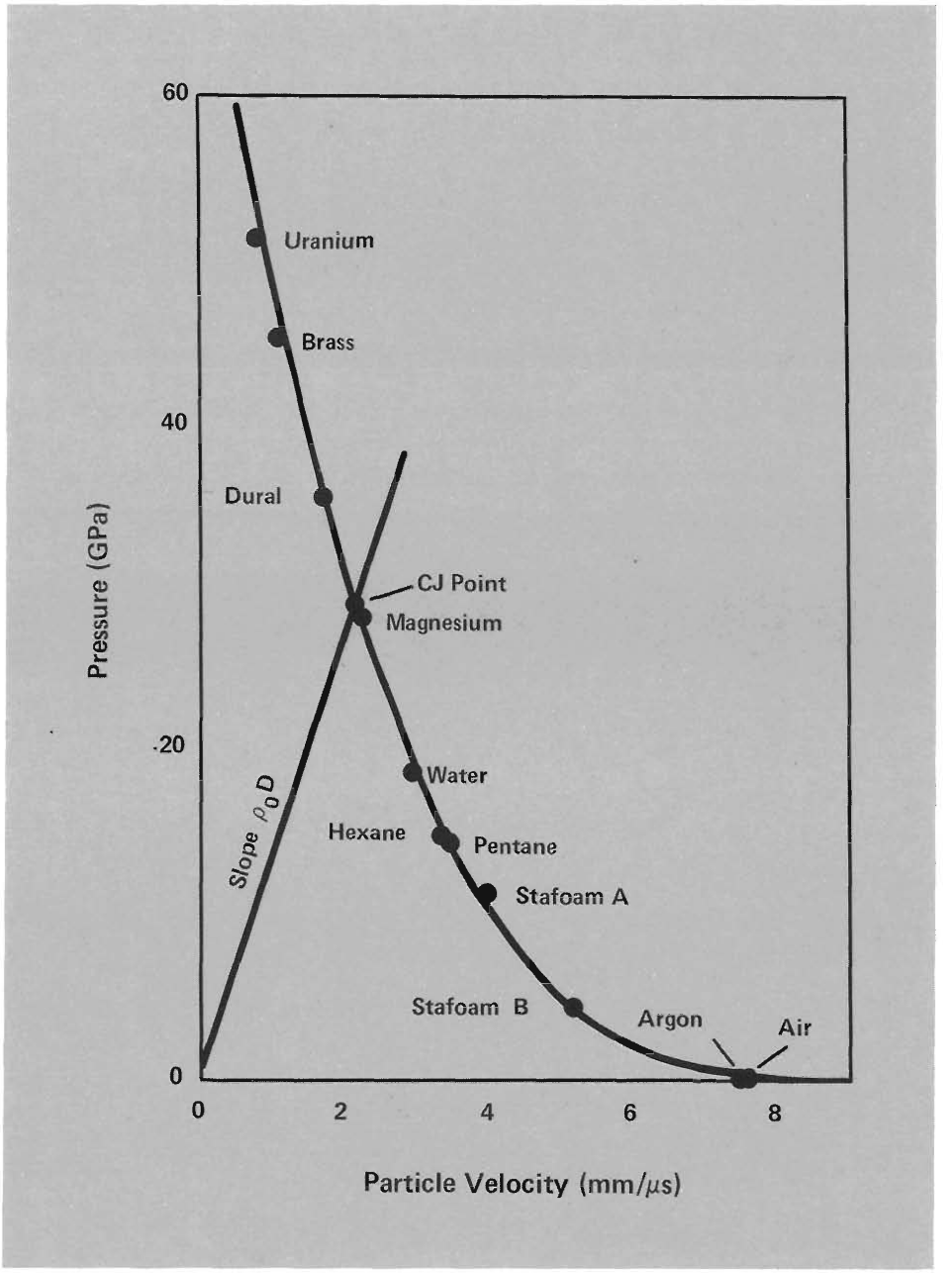


Fig. 11. Plot of pressure and particle velocity obtained from zero-thickness free-surface velocity measurements for Composition B explosive. The CJ point must lie on the Rayleigh line for the measured detonation velocity; it is determined from the line's intersection with a smooth curve. The points above the CJ point determine the reflected-shock Hugoniot curve, and those below determine the expansion isentrope for the explosive products.

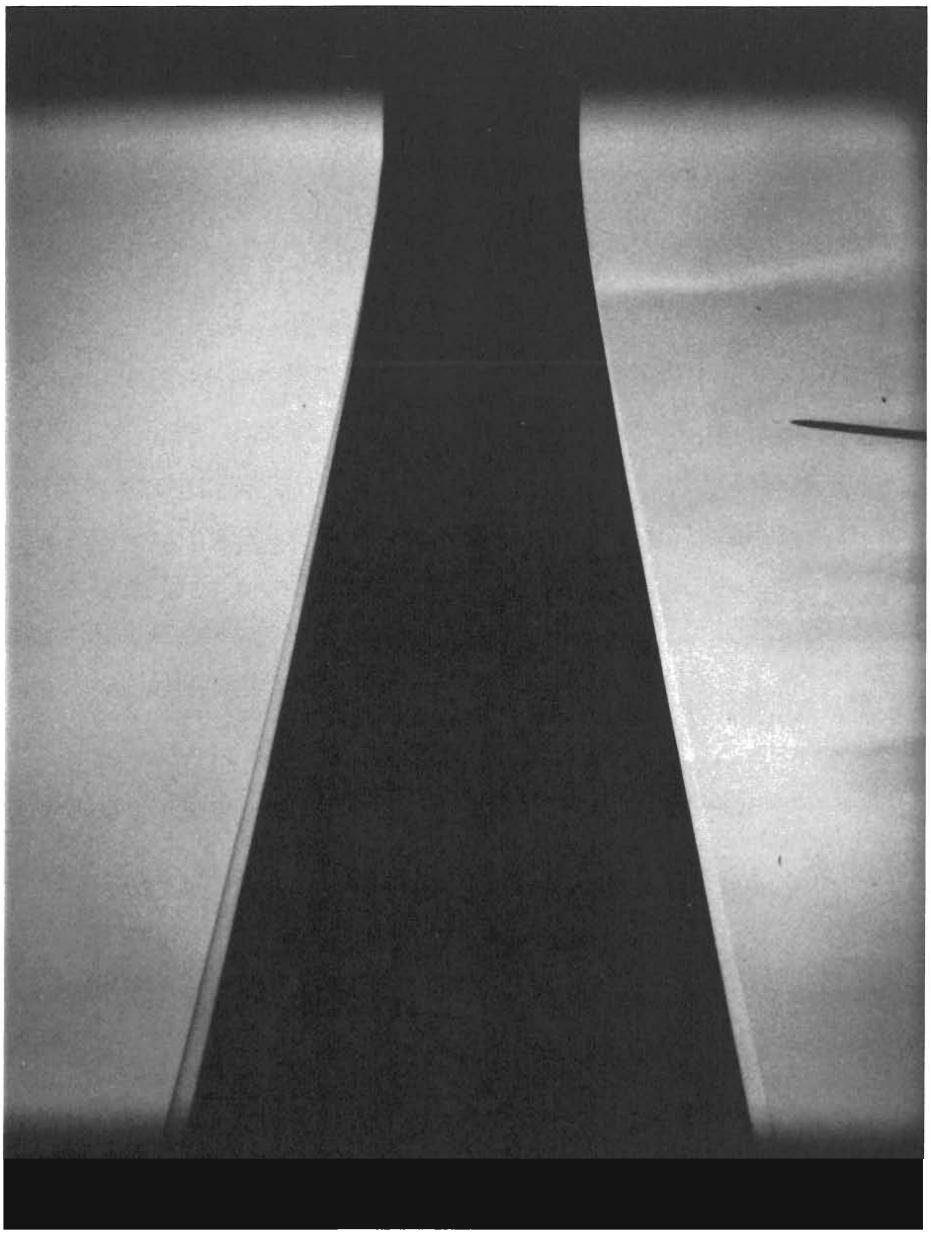


Fig. 12. Silhouette flash photograph of a cylinder test. The explosive charge, 300 mm long and 25 mm in diameter, was encased in a copper tube with 2.5-mm wall thickness. Explosive in the tube has been detonated at the bottom and is accelerating the copper outward. A very short section of undisturbed tube can be seen at the top. Measurements of the shape of the copper wall can be used to compute the expansion isentrope of the explosive products. Shock waves in the surrounding gas can be seen around the tube.

carefully, but the two-dimensional calculation, while neither cheap nor easy, is less expensive than the plate experiments.

Performance of Explosives

As one can easily imagine, there has been a tendency to avoid long calculations by using the experimental measurements of cylinder-wall velocity directly to rank explosives in order of their performance for accelerating metal. This is a mistake. Consider a series of explosives with fixed energy per unit volume but with various densities. In experiments, the energy released by chemical reaction is partitioned between the explosive products and the copper wall. Obviously, the lower the density of the explosive, the more energy is transferred to the copper. The lowest density explosive will rank highest. This ranking is correct if the explosive is used to expand the tube. But suppose the explosive must collapse the tube. In this case, the explosive is applied as a layer on the outside of the tube. Now the confinement effect of the tube is gone, and the inertia of the explosive product gases provides the confinement for energy transfer to the copper. In this imploding configuration, there is an optimum density for the explosive. If the explosive is too light, its products, containing almost all the kinetic energy, will fly off at very high velocity. If it is too heavy, both the metal and product gases will move slowly, and less than the optimum fraction of the energy will be transferred to the copper. Therefore, calculations are necessary to determine the optimum-density explosive to achieve the best

performance in any particular configuration of explosive and inert material.

Today, the performance of explosive systems for most practical applications is calculated by using the CJ point and equations of state determined from free-surface velocity and cylinder tests. Most engineering calculations neglect the reaction zone length and calculate only the inert flow. The calculations are very accurate when they are done correctly, but they have subtle problems. C. L. Mader, of Los Alamos, discusses many aspects of detonation calculations in his book, *Numerical Modeling of Detonations*, a volume in the Los Alamos Series in Basic and Applied Sciences.

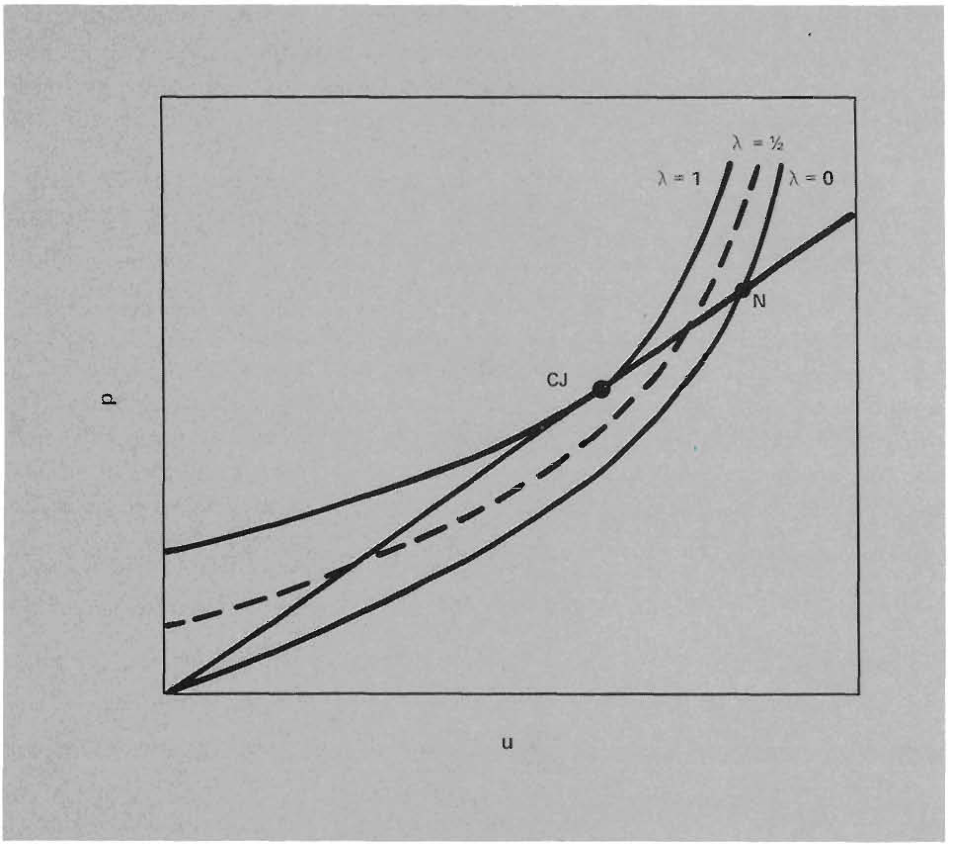


Fig. 13. Three Hugoniot curves and the Rayleigh line for a ZND reaction zone of a CJ detonation. The progress variable λ is zero for unreacted explosive, and 1 for completely reacted products. The pressure at point N is the pressure immediately behind the shock in Fig. 2; it falls as reaction proceeds (as λ goes to 1) until the CJ point is reached at complete reaction.

The Reaction Zone

The assumption that the reaction zone is small compared with distances of interest can be violated in two ways: the charges can be abnormally small, or the explosives can have abnormally long reaction zones. Explosive logic (computers for extreme environments that use explosive to make *and* and *or* gates) is an application using extremely small charges, and one in which the length of the reaction zone must be taken into account. Accurate modeling of the reaction zone is important in applications using insensitive high explosives that have long reaction times. These applications include design of nuclear weapons and studies of *in situ* retorting methods. The discovery that TATB and nitroguanidine (NQ) are extraordinarily safe explosives, with accidental initiation between 100 and 10,000 times less likely than for common military explosives, was of great interest to weapon designers. Although accidental initiation of explosive cannot produce a nuclear explosion, it can result in the scattering of toxic and radioactive nuclear material. Consequently, insensitive high explosives are used in nuclear weapon design. These explosives are safe in part because their chemical reactions proceed more slowly than those of other explosives. Slow chemical reactions mean long reaction zones.

Explosives used for blasting also have long reaction zones. For economy and safety, some are made from coarse granular ammonium nitrate (a source of oxygen) coated with hydrocarbon fuel. The physical separation of fuel and oxidizer means that the components must diffuse into each other. Because the diffusion is slow, the reaction zone is very long. Until recently, blasting was done with the guidance of simple tests, but with no detailed computer calculations. However, new *in situ* retorting methods for oil shales and coal require precise fragmentation of the rock. Los Alamos, Livermore, and others are doing experiments and calculations to try to solve some of the problems. The relatively small charges of these materials used at laboratory firing sites don't stay together long enough for the reactions to go to completion. Therefore, to scale measurements of small charges up to the large sizes required for blasting, we must understand the chemical reaction rates and their dependence on boundary conditions.

The ZND Theory

Present Laboratory efforts to model the reaction zone are built on theories developed during the war years. About 1940, the theory of detonation was extended independently, but in almost ex-

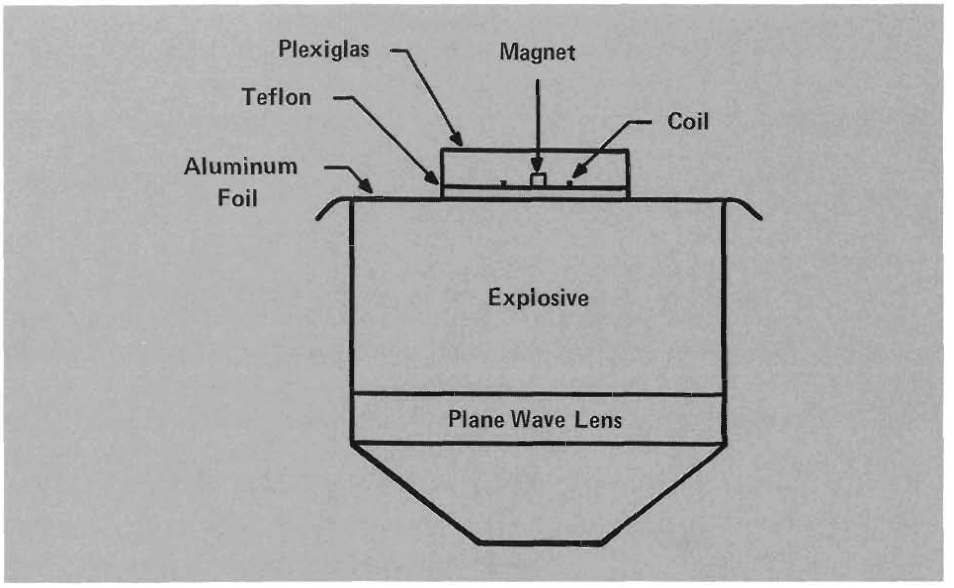


Fig. 14. An experimental assembly for measuring particle velocity with the magnetic probe technique. The motion of the aluminum foil (75 μm thick) between the explosive and the Teflon in the nonuniform magnetic field induces about 150 mV in the single-turn pickup coil. The voltage is recorded during the interval, about 1 μs, between the time when the foil starts to move and the time when the shock wave reaches the coil and destroys it. The voltage record can be unfolded to give foil velocity versus time.

actly the same way, by Zeldovich in Russia, von Neumann in the United States, and Doering in Germany. The ZND theory, describing a steady reaction zone with a finite chemical reaction rate, took some of the mystery out of the CJ theory. In the ZND theory, the front of the detonation is a shock wave where the pressure and temperature rise, but where little or no reaction occurs because the time is so short. Behind the shock wave, the explosive reacts at high pressure and temperature until all of it is changed into product gases. To make the equations tractable, the ZND theory also assumes that the reaction zone is steady; that is, it maintains exactly the same form as it moves through the explosive. The shock wave in Fig. 3 is unchanged, but instead of a single state (pressure and particle velocity) behind it, there is a continuous range of states described by a new variable λ, $0 \leq \lambda \leq 1$, that marks the progress of the chemical reaction. The assumption that the reaction zone is steady, however, reduces the problem to the kind already considered. Equations (1), (2), (3), and (6) apply between the initial state and any selected state in the reaction zone. Equation (6), of course, must be changed to represent the heat released up to that point. If λ is the fraction of material already reacted, it is also the fraction of heat already released. The replacement for Eq. (6) is

$$p_\lambda = \rho_0 (n\lambda Q + cu_\lambda + su_\lambda^2 + ...) , \quad (8)$$

where the subscript λ indicates the point in the reaction zone where λ has that value. The coefficients n, c, s, ..., are also functions of λ. Figure 13 is a plot of curves represented by Eq. (8). The Hugoniot curve marked $\lambda = 0$ for the unreacted explosive at the detonation front is just like the Hugoniot curve for an inert material shown in Fig. 5. As chemical reaction proceeds toward $\lambda = 1$, the state in the reaction zone is represented by points along the Rayleigh line between N and CJ.

The pressure-distance curve for the ZND detonation was shown in Fig. 2. The shock wave at the detonation front in Fig. 2 raises the pressure to point N in Fig. 13. As the reaction proceeds, the pressure falls until it reaches point CJ, where reaction is complete. (This point is marked "final state" in Fig. 2.)

The new feature of the ZND theory is the reaction zone and, in particular, the high pressure at the detonation front. The fall in pressure caused by chemical reaction is contrary to almost everyone's intuition. When von Neumann's report detailing his new theory was first circulated in 1942, it was greeted with disbelief because the "von Neumann spike" seemed patently absurd.

Measurable Effects of the Reaction Zone

Figure 13 shows that a particle-velocity spike accompanies the pressure spike at the detonation front. We can measure particle velocity indirectly by placing a foil in the explosive and measuring the foil velocity induced by the passage of the detonation wave. The foil velocity is equal to the particle velocity of the explosive products. The

experiment is diagrammed in Fig. 14, and the results are plotted in Fig. 15. The dashed line in Fig. 15 is the rarefaction behind the detonation front calculated with the simplest theory, that is, as if the reaction zone were negligibly thin. The solid line is the particle velocity measured in the experiment. Where the measured velocity slope is greater than the calculated slope, chemical reaction is still appreciable. The von Neumann spike is real.

The time during which the slope of the solid line differs from the slope of the dashed line is a measure of the chemical rate or reaction zone length. In Fig. 15, the time for passage of the reaction zone is about 0.5 μs, corresponding to a reaction zone length of 3 mm. These data determine a chemical reaction rate of the form

$$r = \frac{d\lambda}{dt} = k(1 - \lambda)\,, \qquad (10)$$

where λ is the progress variable and the coefficient k, with units of reciprocal time, measures how fast the reaction goes. The depletion term, in parentheses, makes the rate go to zero when all the explosive has reacted.

Obviously, one learns very little about chemistry from mechanical measurements. Any information about the real chemistry of the explosive reaction would be extremely valuable but, unfortunately, the pressures and temperatures in the reaction zone are far higher than laboratory measurements. Detonation data must be obtained from detonation measurements.

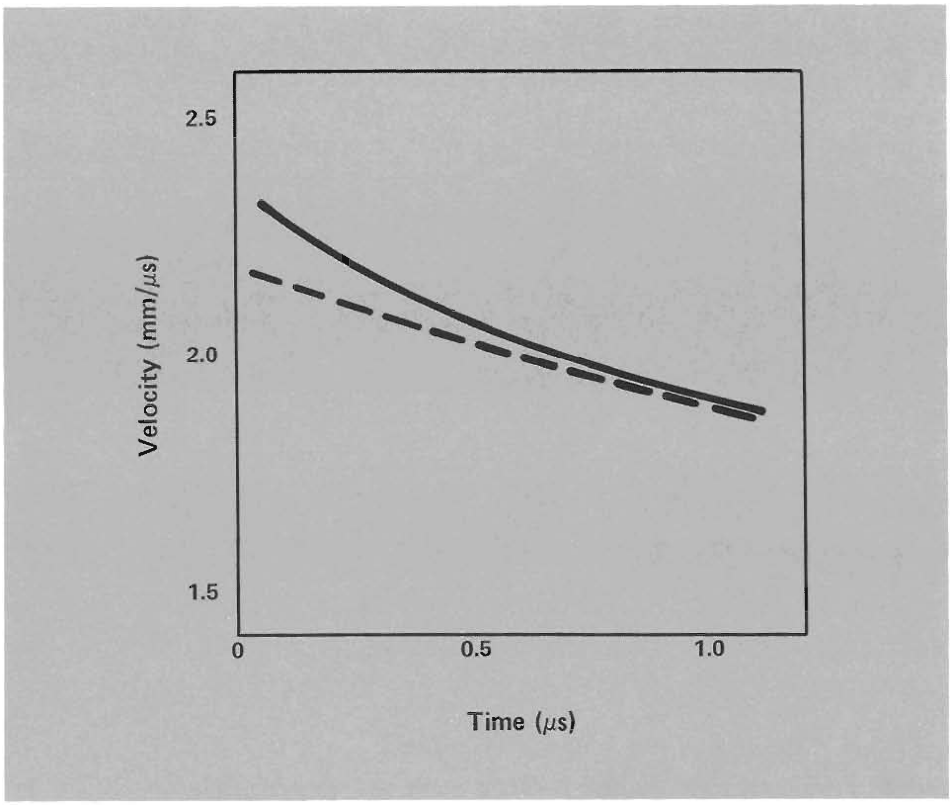


Fig. 15. Particle velocity versus time at the explosive products-Teflon interface. The explosive was PBX-9502, a TATB insensitive composition with a long reaction zone. The solid line is the experimental result, and the dashed line is a calculation using the simplest model, with an infinitesimal reaction zone. Where the experimental curve has a steeper slope than the calculation, reaction is appreciable. The von Neumann spike at the front is obviously real.

Detonations in Cylindrical Sticks

One way to obtain more information about the chemistry in the reaction zone and, in particular, about its dependence on pressure, is to change the boundary conditions so that they affect the chemistry. For example, we can study the detonation process in long cylindrical sticks, where the process has a chance to reach a steady state, but where surface effects, especially the decrease of pressure from the center of the stick to its edges, may change the chemistry.

Over the years, many experimenters have studied detonations in cylindrical sticks and have perfected measurements of the detonation velocity to achieve accuracies of a few parts in ten thousand. The most dramatic result of these experiments is the discovery of the failure radius. That is, for each explosive, there is a radius below which detonations fail to propagate.

Figure 16 shows a cylindrical stick

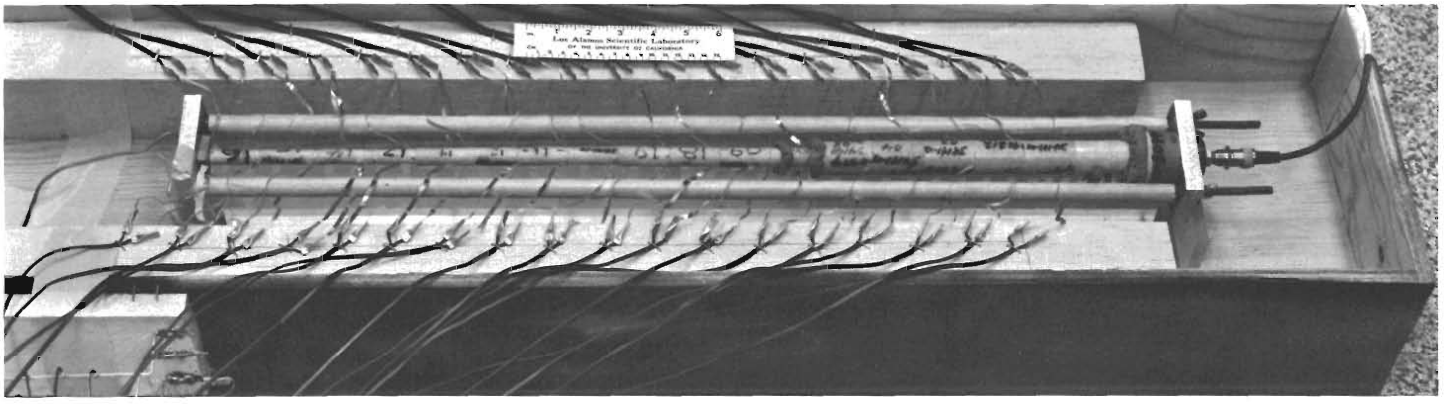


Fig. 16. Detonation velocity experiment ready for firing. Foils placed between segments of the long cylindrical stick of explosive are connected by coaxial cables to oscilloscopes. They measure the arrival time of the detonation wave.

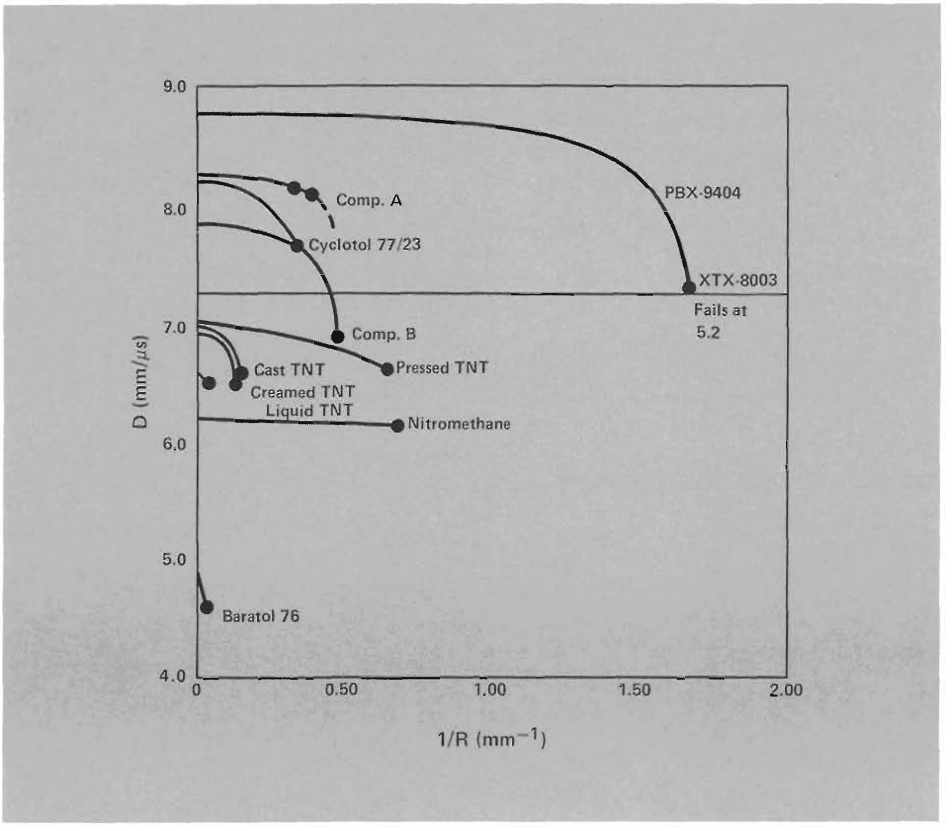


Fig. 17. Plot of measured detonation velocity for solid and liquid explosives versus reciprocal charge radius. The curves end at the measured or estimated failure points. These data show that detonation propagates only for a very limited range of velocity for any particular explosive, the maximum decrement here being only about 15%.

ready for firing to measure its detonation velocity. The very thin foils inserted between the segments of the stick act as electrical switches to indicate the exact detonation wave arrival time. Figure 17 shows plots of measured detonation velocities (D) versus reciprocal charge radii of several explosives. Each curve terminates at a radius, called the failure radius (R_f), below which a detonation will not propagate. Evidently, surface effects take so much energy from the detonation that it cannot proceed. The failure radii for the explosives in the figure vary by a factor of 100, from 0.2 to 20 mm. The detonation velocity decreases as the radius of the explosive charge decreases, but the velocity decrease is surprisingly small. For nitromethane, the detonation velocity near failure is within about 0.5% of the velocity in very large charges of the same explosive. Even the largest velocity decrease is only about 15% in these explosives. Detonation seems to be a very fussy process that can occur only in a very limited range of velocities.

The Two-Dimensional Reaction Zone

In the last 5 years, John Bdzil, of Los Alamos, has extended the theory of the detonation reaction zone to the case of a cylindrical stick of explosive in an inert tube, like the cylinder test described above. In his model, Bdzil assumes the detonation wave has run far enough in the tube that the flow is steady. Figure 18 shows a diagram of a steady detonation in a metal tube. Along the center line the reaction zone is much like the ZND detonation, but away from the line the difference becomes pronounced as the transverse flow transfers some of its energy to the metal tube. The shock wave is curved, and the shock pressure decreases with increasing distance from the center. Because the shock pressure is less, the explosive is heated less and the reaction takes longer. The dashed line shows where reaction is complete. The sonic surface, between the shock and the complete reaction line, is an important dividing surface in this flow. Only the portion of the chemical energy released above this surface contributes to driving the detonation. Even more important, energy used in the transverse expansion in the reaction zone is not available for driving the detonation. Thus, the detonation velocity is less than it would be in a plane wave detonation, or even in a stick with a larger radius.

The theory of this two-dimensional detonation reaction zone is complicated. The shape of the shock determines the initial conditions for the reaction zone, but is itself part of the solution. Bdzil obtained an ordinary differential equation for the shock shape, which can be solved analytically for some simple

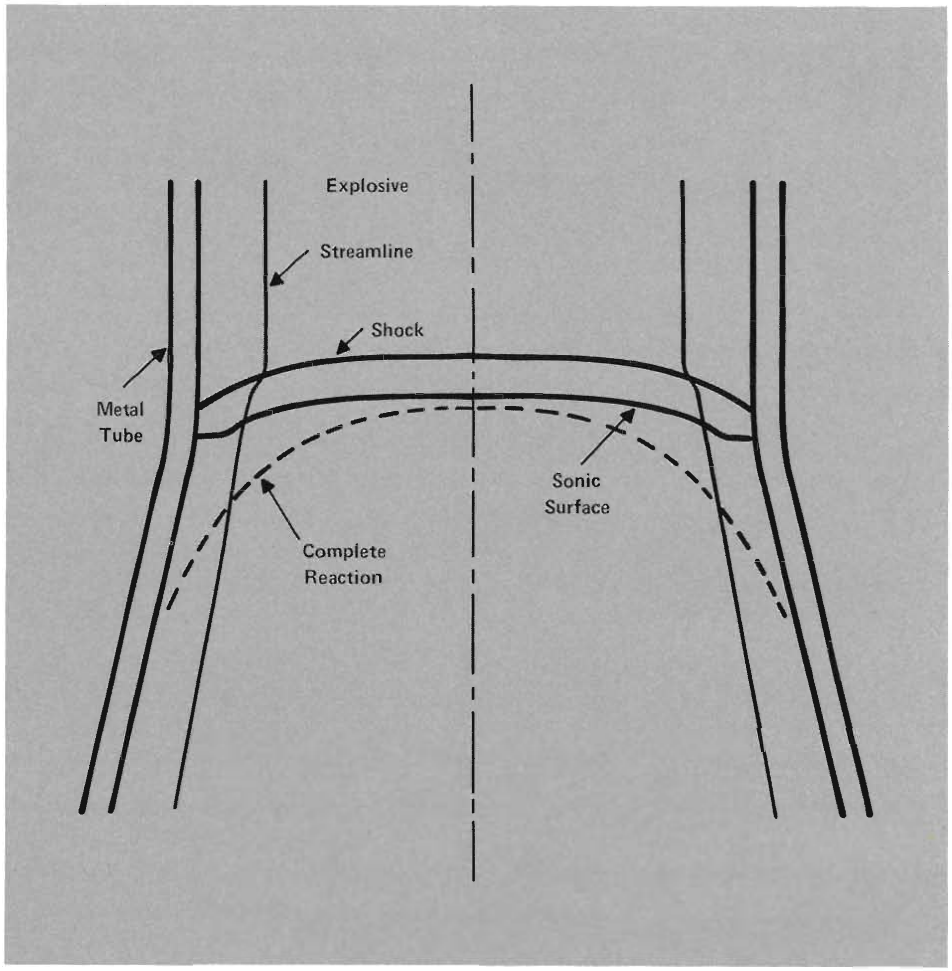


Fig. 18. Cross section of the detonation reaction zone in a metal tube. The curved shock wave is the leading front of the detonation. Explosive heated by the compression reacts behind the shock. Only material that reacts ahead of the sonic surface can contribute to driving the detonation wave.

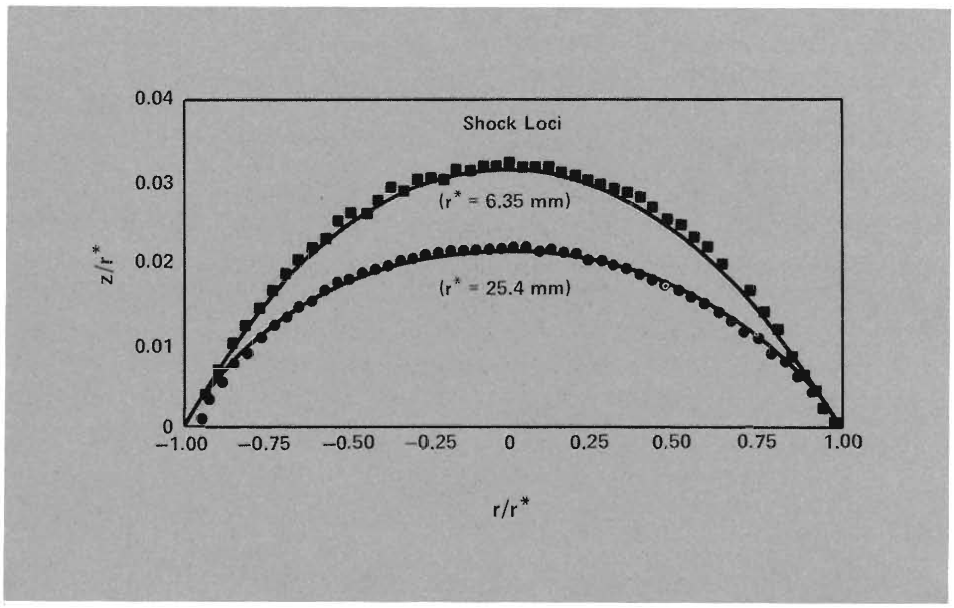


Fig. 19. Comparison of measured detonation shock wave shapes with calculations of the shape for an assumed rate law. Getting agreement between measurement and calculation places restrictions on the form of the rate law.

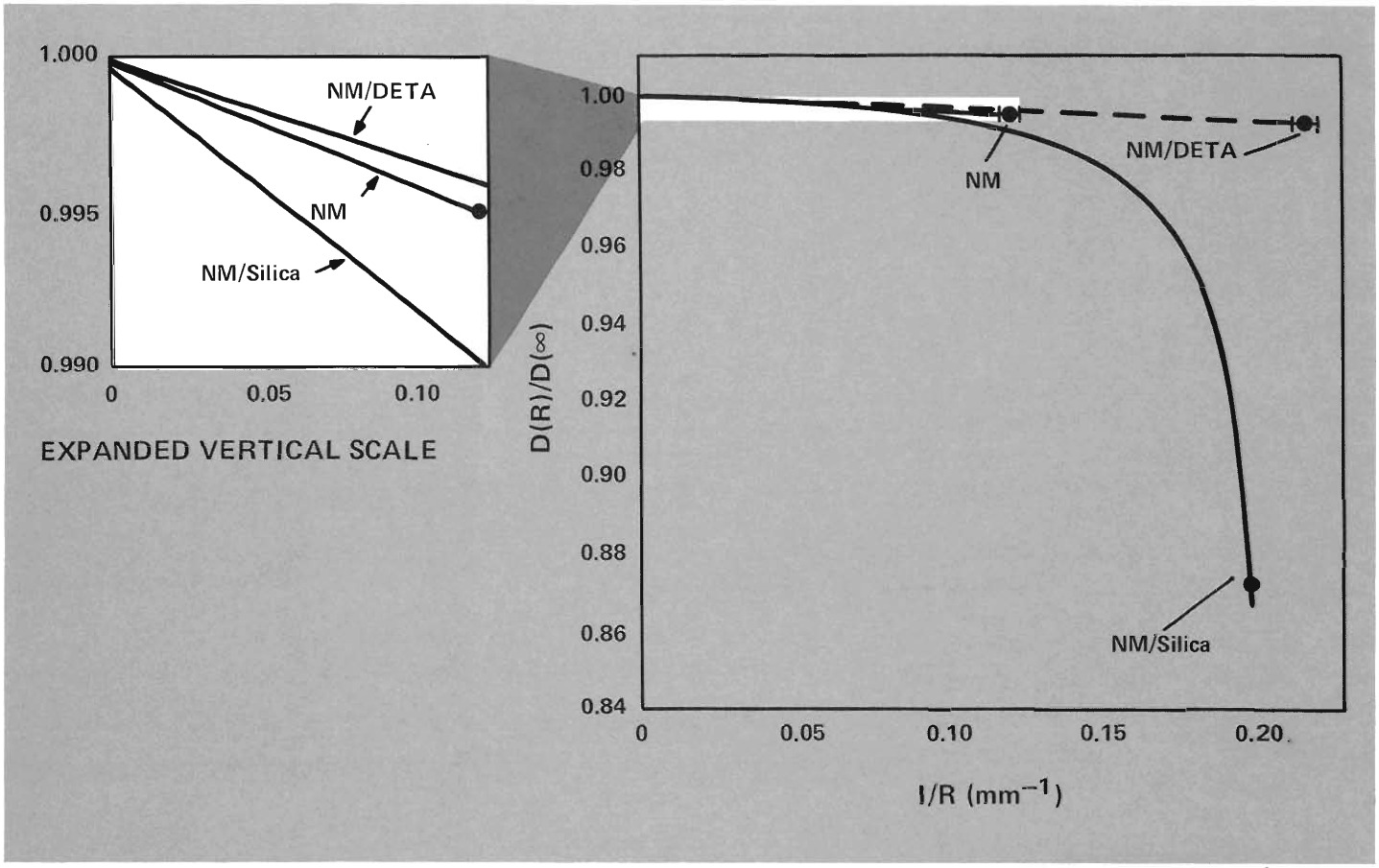


Fig. 20. Plot of measured detonation velocity versus reciprocal radius for nitromethane, nitromethane with the catalyst DETA, and nitromethane with silica particles. The catalyst and the particles both decrease the failure radius by a factor of 2, but the silica particles change the shape of the curve. The graph on the left is a magnified view of the curves where they intersect the ordinate. The slopes of the curves in this region indicate the reaction zone lengths. The catalyst shortens the reaction zone, and the particles lengthen it.

chemical reaction rates and integrated numerically for others.

The theory must match the measured detonation velocity, the measured shape of the detonation front, and the measured velocity decrement at failure. Applying the theory to these data gives some insight into the chemical reaction mechanisms. If the chemical reaction rate is assumed to have a form

$$r = k(p/p^*)^n(1 - \lambda) , \qquad (11)$$

where p^* is the limiting shock pressure for the large charges at the left edge of Fig. 17, the theory can be used to fit the observations. The coefficient k, with units of reciprocal time, measures the scale of the chemical reaction time and, therefore, the reaction zone length. It is determined principally by the slope of the velocity curves near their start at the left in Fig. 17. Larger slopes correspond to larger edge effects and, therefore, to longer reaction zones. The value of the exponent n in Eq. (11) is determined principally by the amount the velocity decreases before the failure point is reached. For n small, the effect on the chemical reaction rate of the pressure decrease at the edges of the stick is small, and the detonation can proceed at lower velocity. For n large, the effect is greater, and the detonation fails with less decrease in velocity.

The measured shock wave shape at the front in a detonating stick is compared with the shape found from the theory, and the reaction rate is adjusted for the best match. Matching the shock shape for several different radii, but using the same chemical reaction rate for all, places functional requirements on the rate law. Figure 19 compares theoretical fits with data for cylindrical nitromethane sticks with two different radii.

New Reaction Zone Experiments

These advances in the theory have encouraged new experiments to explore the chemical reaction zone. The reaction rate and its dependence on pressure can be influenced by adding a catalyst or by suspending solid particles in a liquid explosive. In one series of experiments, nitromethane, a water-white, nonviscous liquid explosive, was chosen as the standard. The experiments compared the detonation velocity and failure radius of the standard with those of the standard plus small amounts of additional materials. Because the same explosive was used in all the experiments, any uncertainties in the equation of state or any other properties cancel out and the effects of the added materials on the reaction rate can be determined. The diameter-effect curve for nitromethane, in Fig. 20, appears to be a straight line, and its

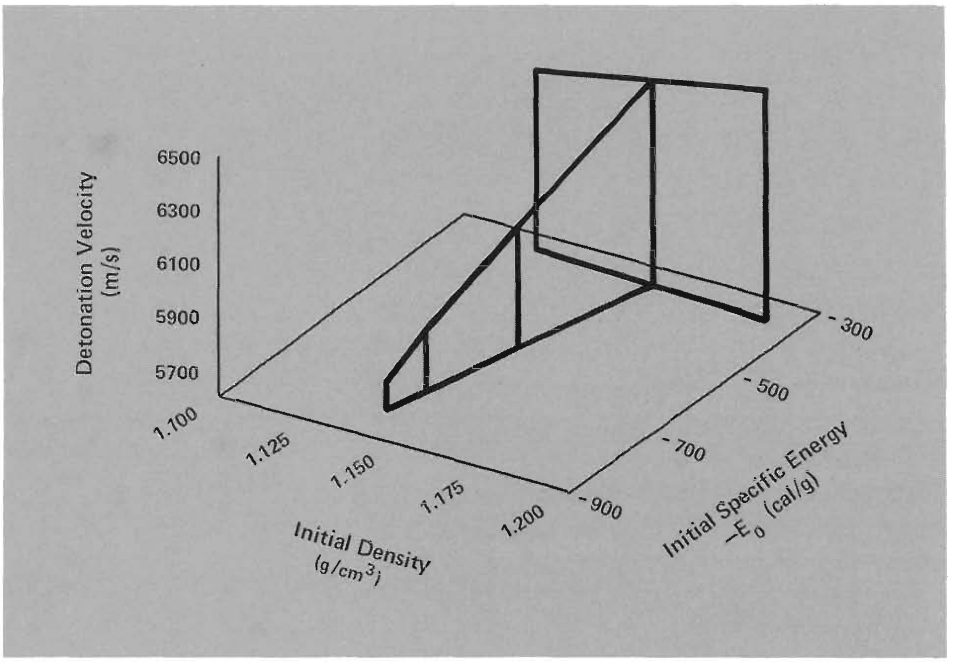


Fig. 21. Three-dimensional plot of detonation velocity as a function of initial density and initial energy, for nitromethane (at the rear) and mixtures of acenina and nitromethane (leading up from front to back). These data provide a test of the theory.

For nitromethane, changing the temperature changes the density quite a lot, and the energy a little. For mixtures of nitromethane and acenina, the energy changes a lot and the density a little. Acenina is an equimolar mixture of nitric acid, water, and acetonitrile. It has the same atomic composition as nitromethane, so the explosive products are the same. These experiments define the detonation velocity surface in the neighborhood of the intersection of the two lines.

velocity decreases only 0.5% at failure. Adding the catalyst DETA in a very small amount (0.03%—about 1 molecule of catalyst for each 5,000 molecules of explosive) decreases the failure radius by a factor of 2, but leaves both the detonation velocity and the velocity decrease unchanged. Adding silica particles and a little gelling agent to hold them in place also decreases the failure radius by a factor of 2, but changes the velocity decrease to about 13% at failure. The slopes of the curves at large radius, shown in the magnified view on the left, indicate [as discussed with Eq. (11)] that adding DETA shortens the reaction zone. On the other hand, adding silica particles lengthens the reaction zone appreciably.

Ray Engelke, of Los Alamos, who is doing these experiments, interprets the effect of the catalyst DETA as the expected one; it speeds up the reaction, thus shortening the reaction zone, decreasing the edge effects, and thereby decreasing the failure radius. The fact that the velocity decrease stays the same indicates that the catalyst doesn't change the dependence of the chemical reaction rate on the pressure in this explosive. The silica particles lengthen the reaction zone but decrease the failure radius. Probably the particles create hot spots; that is, the shock wave develops irregularities as it passes over them, and the high temperature in these spots keeps the reaction going even when the average pressure and temperature fall near the edge of the cylinder. Another possibility is that real failure is not as simple as the model, and that strong transverse waves traveling from the edge into the reaction zone quench the reaction. If so, the silica particles may interfere with the action of the waves by diffusing their sharp fronts as they envelop the individual particles.

Perhaps experiments like these will develop enough understanding of the factors influencing the reaction rates in the detonation reaction zone that we can formulate new explosives with special useful properties. For example, the insensitive high explosives now in use are safe partly because of their slow reaction rates at low pressures. But their slow reaction rates at the high pressures in the detonation reaction zone cause unwelcome behavior. Understanding may lead to explosives with slow rates at low pressures and fast rates at high pressures.

Is the End of Reaction a CJ Point?

In the limit of large charge size and plane flow, we expect that chemical reaction ends at the CJ point, where the flow is exactly sonic. Direct comparison of theory and experiment to determine whether the end point of chemical reaction really is a CJ point is not possible for most solid and liquid explosives because the equations of state needed to do the calculation must be determined experimentally and the experiments incorporate the assumption that the end point is a CJ point.

To avoid this logical dilemma, we have developed a pressure-determination method that is independent of the exact equation of state of the final products. The method involves measuring how much the detonation velocity varies with initial energy and density of the explosive. The measurements are made for two explosives with the same atomic composition, and the data define a detonation velocity surface as described in Fig. 21. To determine the pressure of the explosive products from this surface, we apply the CJ theory. No knowledge of the exact equation of state is needed; we need only the assumption that for two explosives with the same atomic compositions, the equilibrium equation of

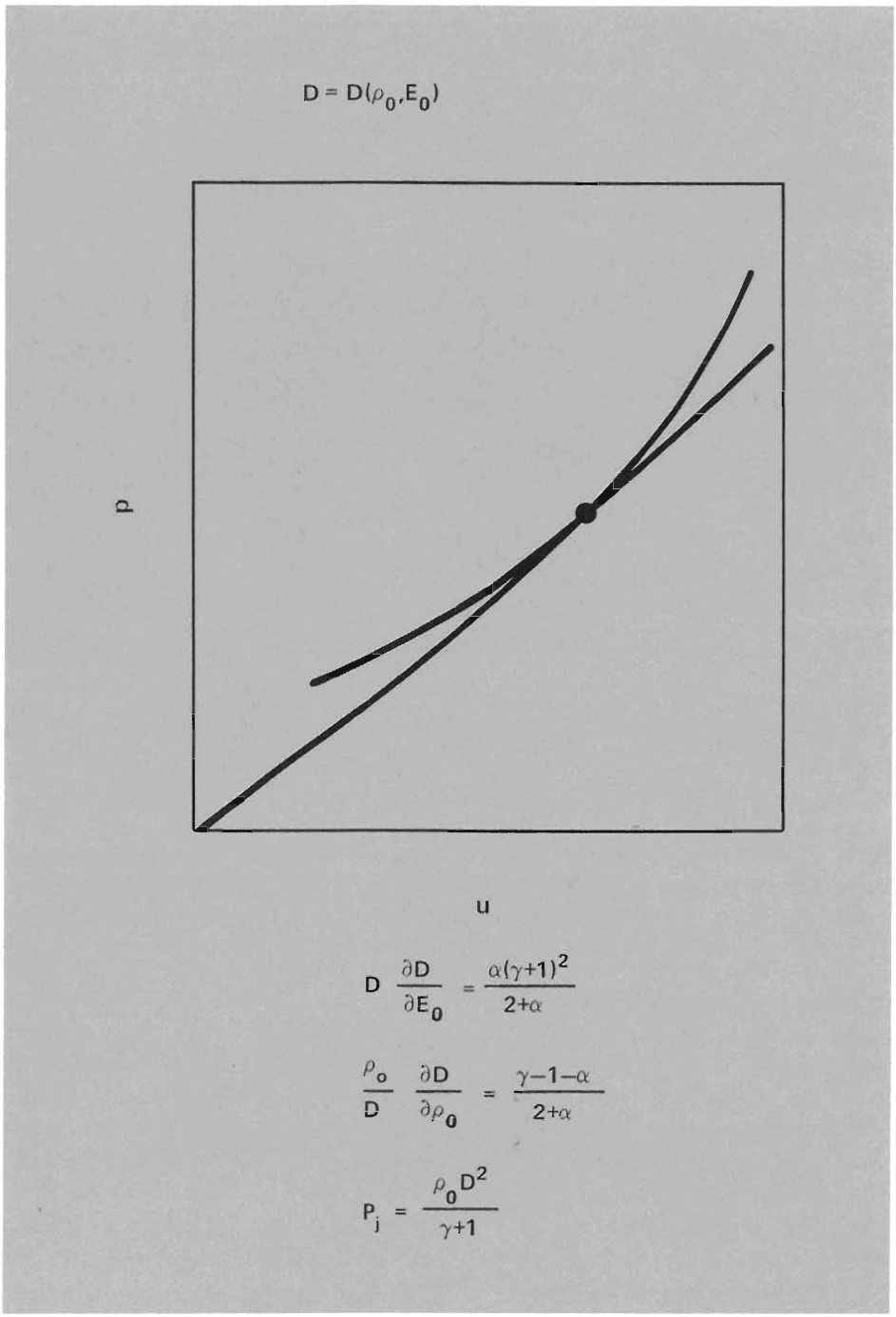


Fig. 22. Outline of the theoretical results needed to use the detonation velocity data (see Fig. 21) for a test of the theory. The pressure obtained from this interpretation is compared with other pressure measurements.

state for the explosive products is the same for both.

The theory to determine the pressure is shown in Fig. 22. The expression for D at the top means that from experiment (Fig. 21) the detonation velocity is known as a function of initial density and energy over some small region. The sketch of the p—u plane shows the theoretical Hugoniot curve and Rayleigh line for a plane, steady, laminar detonation wave calculated with any of a class of simple reaction rate laws. The simple theory says that the straight line (the Rayleigh line) and the curved line (the Hugoniot curve) are tangent at the state point that occurs at the end of the reaction zone. The next two equations, derived from the CJ theory, determine the variation of CJ detonation velocity with initial energy and density in terms of certain thermodynamic derivatives (α and γ) evaluated at the tangent point. The derivatives of D on the left-hand side are determined from measurements shown in Fig. 21. The two equations can be solved to obtain the values of α and γ. Finally, the pressure at the tangent point can be obtained from the last equation. The pressure obtained from these detonation velocity measurements, interpreted using this theory, can be compared with the pressure from the free-surface measurements and cylinder tests described earlier. The best values I have for the pressures of nitromethane are 12.2 ± 0.6 GPa, determined from the detonation velocity measurements, and 14.2 ± 0.4 GPa, determined from the more conventional measurements. Disagreement between the values shows that the end point for the reaction zone in nitromethane (the final state) is not a CJ point.

Extended Theory for the Final State

Why is the final state not a CJ point? Some ideas about the final state can be garnered by looking at Fig. 23, which is like Fig. 13, but has a higher detonation velocity. The shock wave takes the material to state N, and as the reaction proceeds the pressure falls to state S (for strong point). In the case we discussed before, a rarefaction from behind can overtake state S and reduce the pressure until the CJ state is reached. Point W (the weak point) is a possible solution of the conservation equations, because it is an intersection of the Rayleigh line (conservation of momentum and mass) and the Hugoniot curve (conservation of energy). However, point W cannot be reached. The reaction starts at point N and is finished at point S; according to our simple theory there is no path from S to W.

The real world, however, is not so simple. Explosives are made from complicated molecules, and the chemical reactions take place in many steps, not just one. The Hugoniot curves for a partially reacted explosive cannot be described by a single progress variable λ, because there are many possible pathways. The theory of detonation has been extended to consider all the added pathways for reaction. With some rates, the space between S and W is filled with possible Hugoniot curves, and point W becomes an attainable state.

The theory is discussed in detail in *Detonation* by W. Fickett and W. C. Davis, another book in the Los Alamos Series in Basic and Applied Sciences. To give an idea of the theory's flavor, Fig. 24 shows the possible forms for a detonation reaction zone when there are two

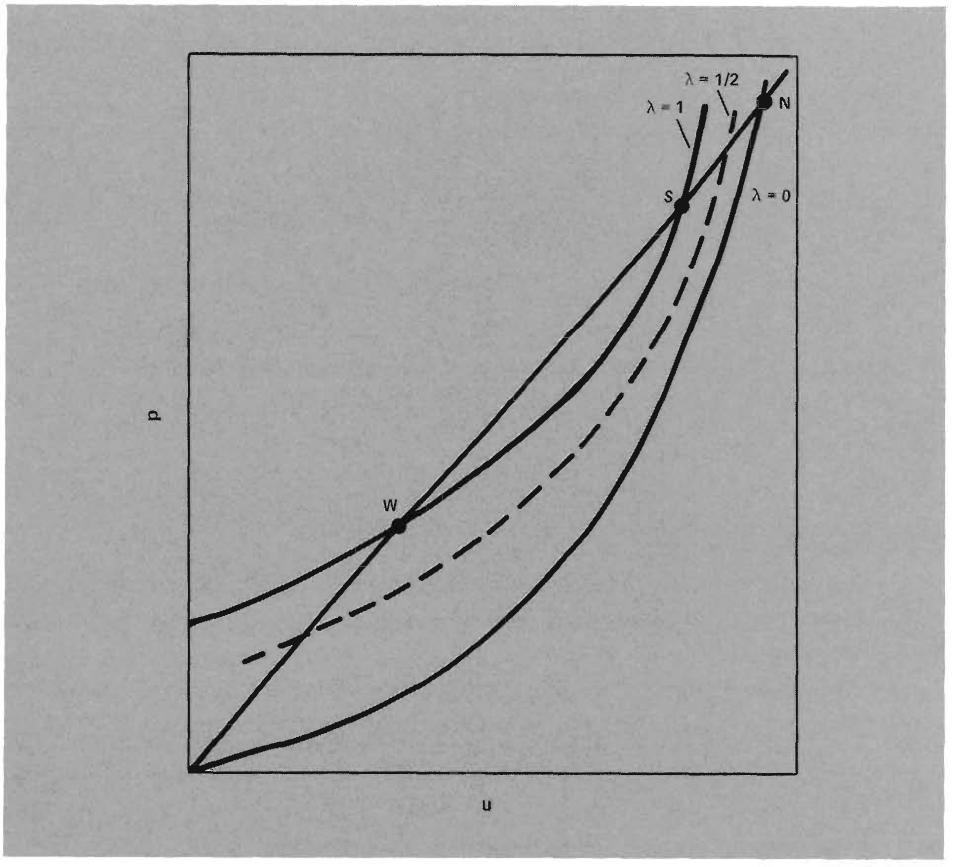


Fig. 23. Three Hugoniot curves and the Rayleigh line for a detonation wave moving faster than the CJ detonation. The initial shock is to point N, and as reaction proceeds the pressure falls to the final state at the strong point S. The weak point W is not attainable in this simple model.

reversible reactions. The abscissa k is the ratio of the reaction rates, and the ordinate u is the particle velocity at the end of the reaction zone. The two forms to the right of k_e correspond to Figs. 13 and 23. All the other forms are new from the extended theory. Detonation physics is richer than was once thought.

These arguments show that the final state of the reaction zone need not be the CJ state, but they fail to show why the detonation in nitromethane is not a CJ detonation. Perhaps the many reactions lead to a weak state like some of those in Fig. 24, but I don't think that is very likely. The ratio of the reaction rates must be extreme, say about 1,000 to 1, to reach a weak state. With many reactions, I think there must be a route to the completion of reaction without such extreme ratios.

Although this part of the extended theory does not seem to furnish an explanation directly, it does show what is needed to reach a weak point instead of the CJ point. What happens in a detonation with two rates, one much faster than the other, is that the fast rate goes too far

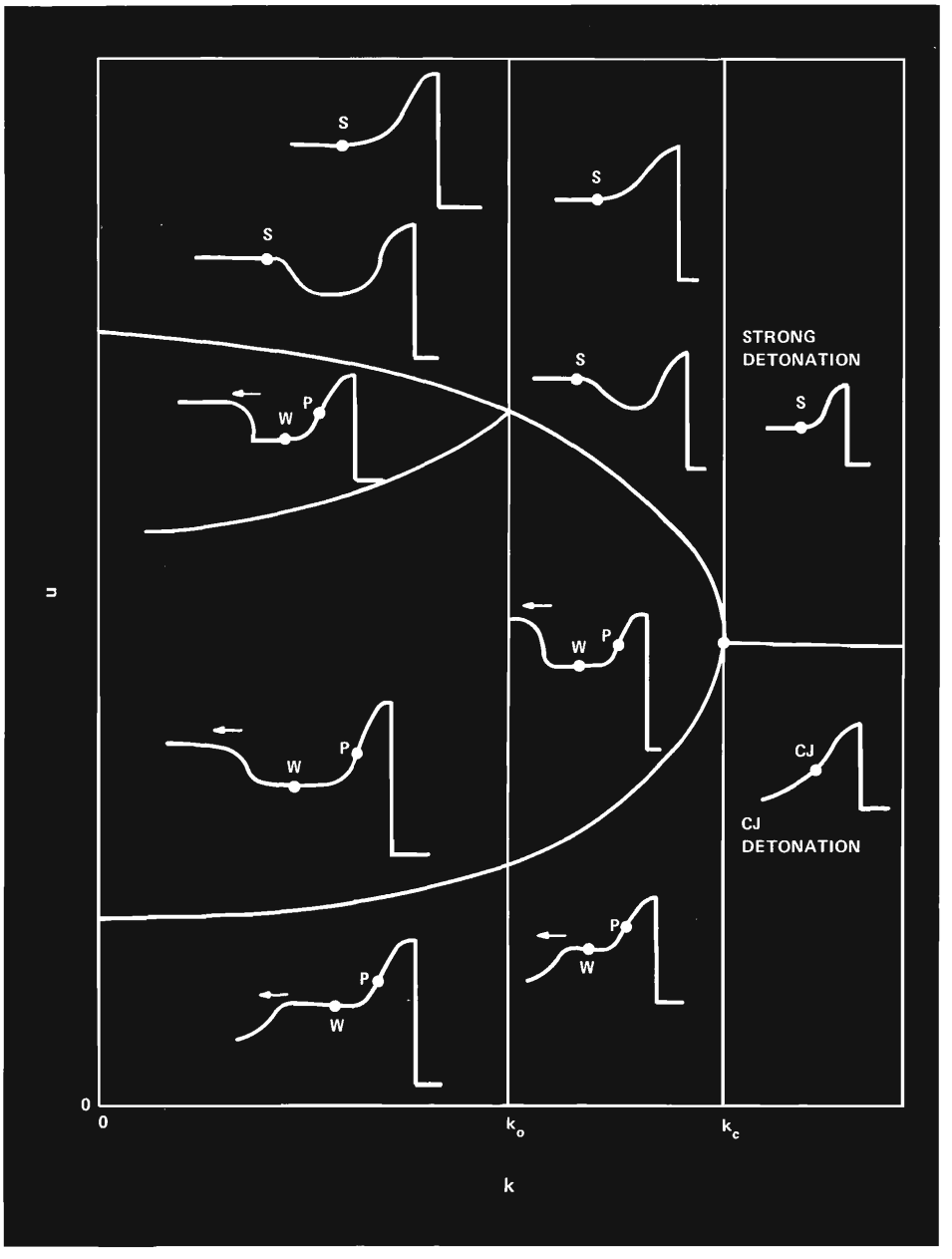


Fig. 24. Diagram of the u versus k plane, showing how it separates into regions where different kinds of strong and weak detonations occur. The particle velocity u through the reaction zone is plotted versus ratio of the two reversible reaction rates assumed for this example. The CJ detonation of Fig. 13 is the pressure-distance plot at lower right, and the strong detonation of Fig. 23 is the plot at upper right. All the others, many of them weak detonations, appear when there are two reactions instead of only one.

and gets beyond its equilibrium composition. Then the slow rate keeps going, using material to get to its equilibrium state, and the composition controlled by the fast rate must shift back from its early state. When this happens, other things being proper, energy is returned from the flow to the internal energy of the chemicals. The return of energy from the flow to internal states is the process needed to make a detonation reach a weak point instead of the CJ state.

Hot Spots and Transverse Waves

The most likely reason that the final state is not a CJ state seems to me to be that energy from the flow is returned to local regions as kinetic energy of fast, small-scale, random flow, like turbulence. The flow contains spatial inhomogeneities, and the pressure and temperature vary from point to point. Thus the rate, dependent on pressure and temperature, varies, and energy is coupled selectively to the small-scale motions in the flow. Although this energy appears as kinetic energy, it is not available to drive the detonation, and it is equivalent to energy in an internal state. The inhomogeneities in explosives and their reaction zones seem to be important for determining the final state of the reaction zone, and they have a large effect on the apparent chemical rate.

Most practical explosives are polycrystalline materials, with small crystals and boundaries between the crystals. Also, they contain defects, such as voids and low-density regions. As the shock wave interacts with the polycrystalline material, it compresses and heats the explosive, but the tem-

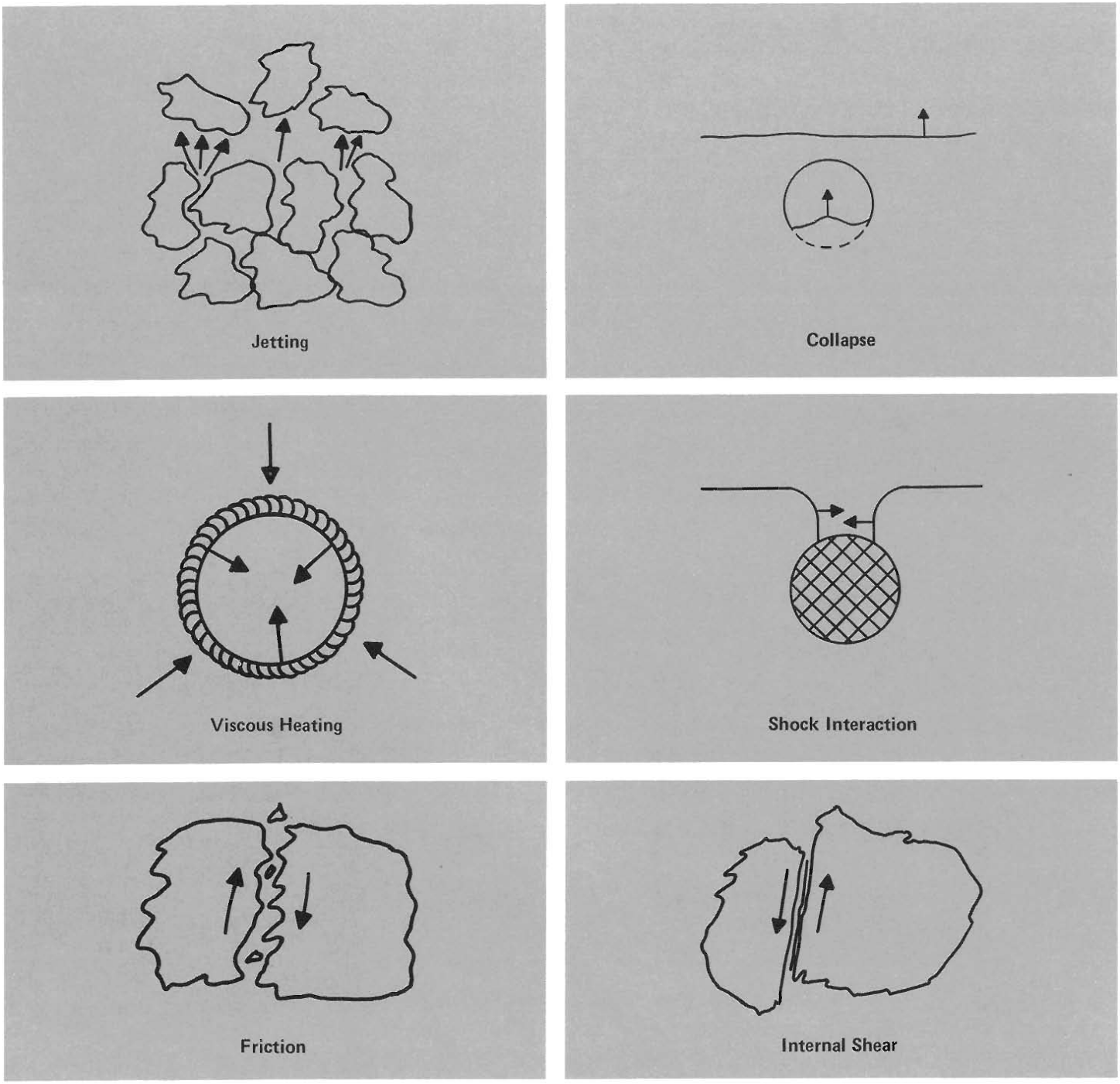


Fig. 25. Six ways to concentrate energy in explosives as hot spots. Hot spots may be formed by material jetting from between explosive grains, by collapse of voids, by viscous heating in material around a void, by interaction of shocks around a high-impedance region, by friction between crystallites, and by internal shear of crystallites.

perature is not uniform throughout the material. Figure 25 shows some of the processes that have been suggested for producing local high temperatures, or hot spots. Hot spots may come from jets of material ejected from little wedge-shaped intersections of crystallites, from impact of material thrown across a void, from viscous heating in material near the surface of a collapsing void, from shock collision around a high-impedance inclusion, from friction between two crystallites, or from internal slippage in a single crystallite. All these processes are important in one special case or another. Figure 26 shows average shock-wave temperature as a function of shock-wave pressure in an explosive called PETN. The two points are the shock points required to initiate detonation in 1 μs in this explosive when it is in two states. The high-pressure point is for a single crystal of PETN, and the low-pressure point is for the same chemical in powder form pressed to almost crystal density. In the single crystal, the temperature is the same everywhere and must be high to start detonation in 1 μm. In the pressed PETN, the average temperature is too low to start any reaction in a short time; if the pressed explosive is put in an oven at this temperature, it decomposes only slowly. At the hot spots the temperature is high, perhaps between 2,000 and 3,000 K; fast reaction takes place at these spots and initiates the explosive. The effect of hot spots is large and exists in most practical explosives.

In single crystals and liquids, heating in a shock-initiation experiment seems to be homogeneous. However, as if nature deliberately attempted to make our study difficult, even these materials develop hot spots in their detonation reaction

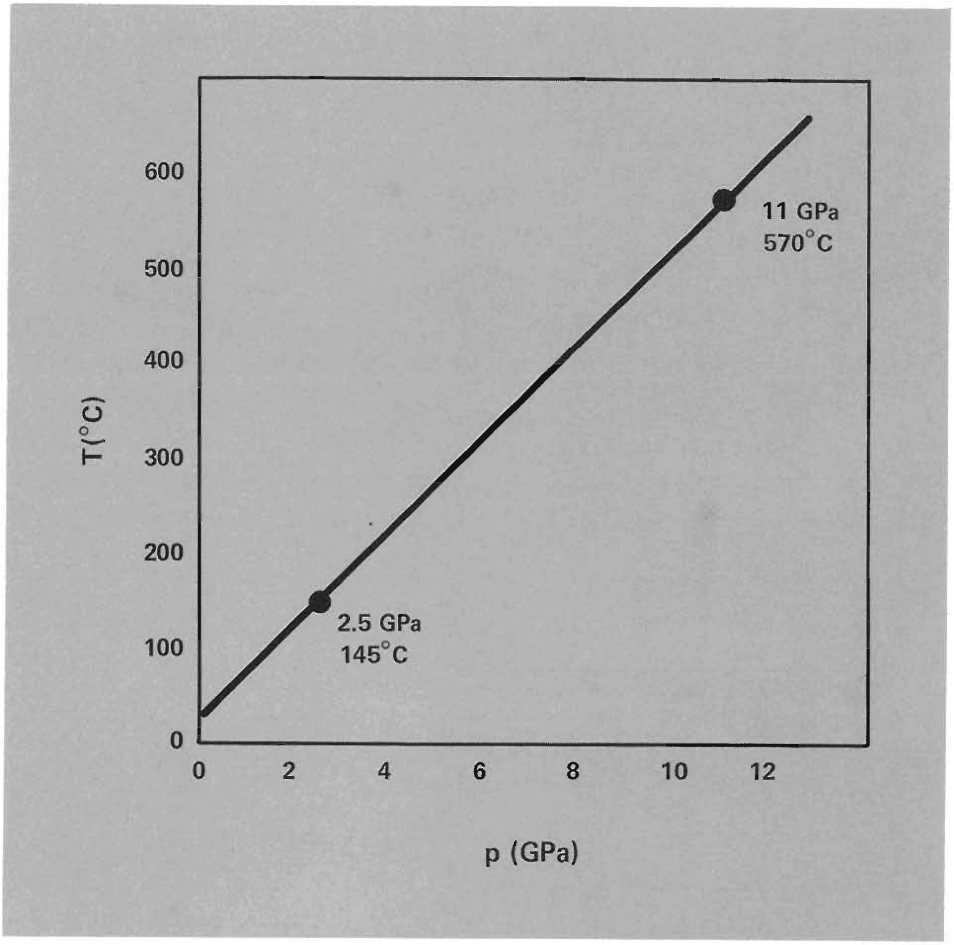


Fig. 26. Plot of shock temperature versus shock pressure for the explosive PETN. A single crystal of PETN is initiated in 1 μs when it is shocked to 11 GPa; the bulk temperature, 570°C, causes rapid reaction. A pressed charge of PETN, made from fine particles of PETN pressed to almost the same bulk density, is initiated in 1 μs when shocked to 2.5 GPa; the bulk temperature, only 145°C, is far too low to cause rapid reaction. In this case the reaction takes place at hot spots and spreads from there.

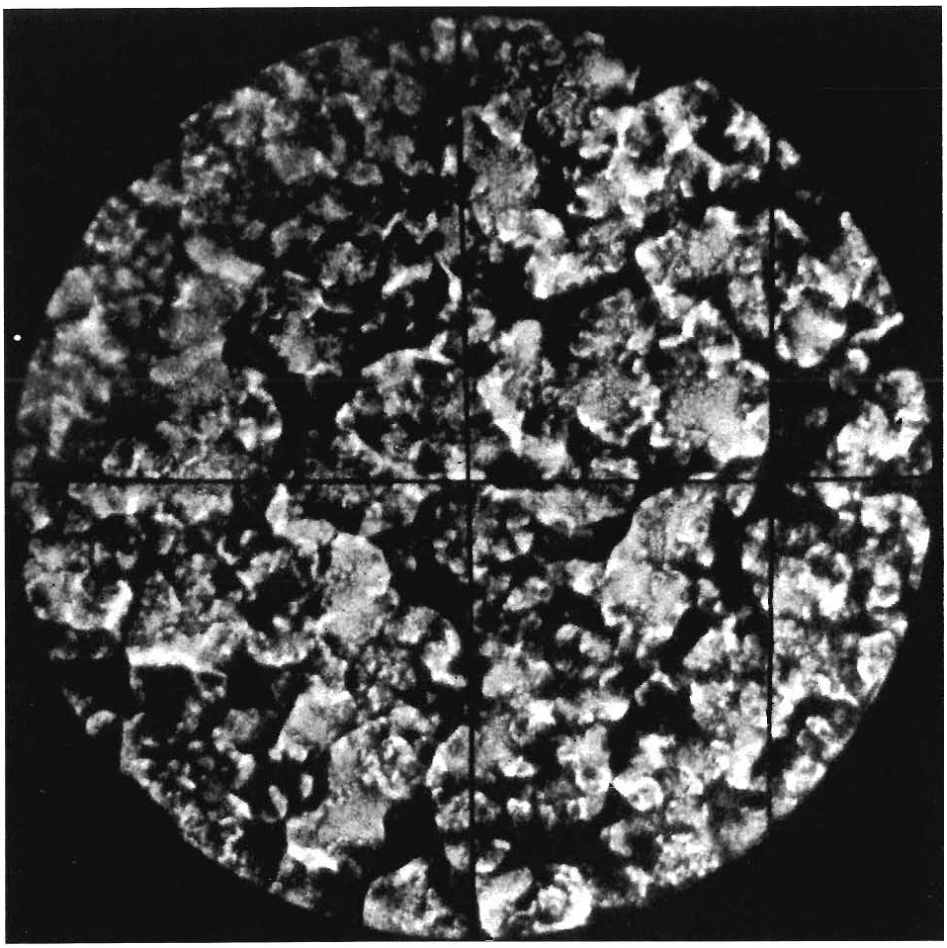


Fig. 27. Image intensifier photograph of detonation light from an 85 vol% nitromethane 15 vol% acetone mixture. The photograph was taken looking straight down at a detonation in a 19-mm-i.d. brass tube. The exposure time was 20 ns. The explosive is water white, and the light can be seen through the unreacted explosive. The transverse wave structure is clearly visible. The apparent brightness temperature is certainly nonuniform, and probably the reaction zone temperature also is nonuniform. Thus, even a completely homogeneous explosive, a liquid in this case, makes its own hot spots.

zones. Figure 27 shows the luminous reaction zone in a transparent liquid detonated in a brass tube. The reaction zone was photographed through the unreacted explosive as the wave moved toward the camera. If the reaction zone were locally smooth and uniform, we would expect the light to be slightly brighter at the center and a little dimmer at the edge, but nearly uniform. The pronounced transverse structure, clearly visible in the figure, is evidence that even in this homogeneous liquid hot spots are generated by the instability of the reaction zone. Sequential photographs of the detonation show that the edges move and the spots transform. Similar structures have been photographed in various liquid and solid explosives, and they probably exist in most explosives.

About 20 years ago, Donald R. White, of the General Electric Research Laboratory in Schenectady, studied detonation structure in gases. He found that the final state of the reaction zone was not a CJ point, but was more like the weak point W in Fig. 23. He analyzed his results to show that the transverse structure in gases, known for many years, could lead to weak detonations. Bdzil also has studied the problem, and his results lead to the same conclusion.

Obviously, we need a theory of detonations with transverse structure or with turbulent reaction zones before we can make any real advance. In the meantime, not only do we have no theory, but we are unable to estimate the size of the effects to be expected. The trouble is that the details of the reaction process (its effective reaction rates) determine the position of the reaction zone end point on the complete reaction Hugoniot curve in Fig. 23. All the conservation condi-

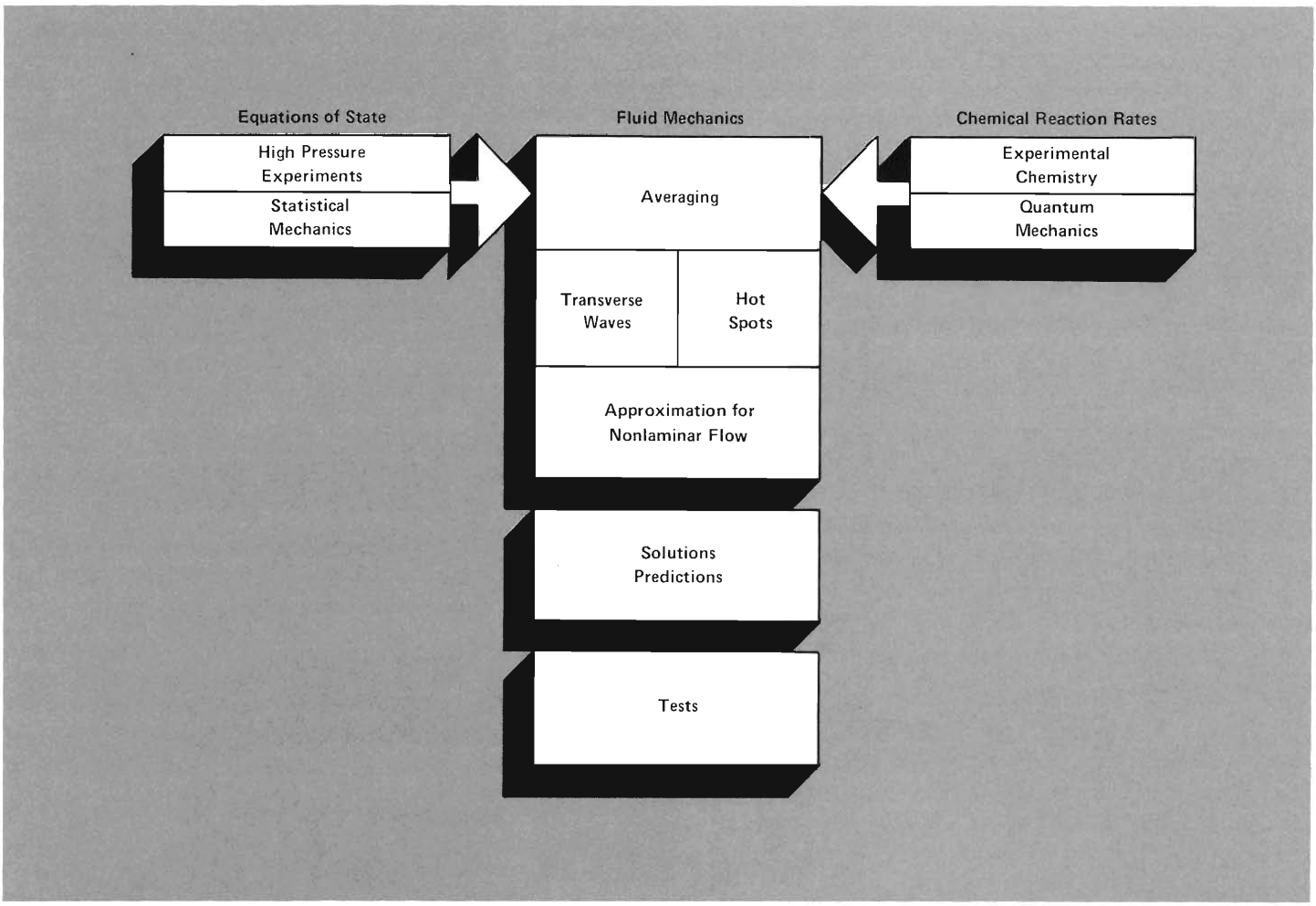


Fig. 28. Summary of detonation physics.

tions are satisfied by any intersection of Rayleigh line and Hugoniot curve. Apparently, no simple energy or momentum arguments can be made to limit the departure from the CJ state.

Summary

The science of explosives (diagrammed in Fig. 28) is an exciting interplay of chemistry and fluid mechanics. To design an explosive or a safe explosive system, we must understand the chemical reaction rates and the equations of state in the detonation reaction zone. To predict an explosive's performance, we must find some way to average the effects of transverse waves and hot spots in the reaction zone.

In the diagram, the boxes at top left and right represent the input of the material properties, the equations of state and the chemical reaction rates. Down the center, we see that fluid mechanics is used to solve interesting problems for explosives with various boundary and initial conditions. The solutions lead to predictions that can be compared with experiment. Usually, independent input from the equation of state and chemical reaction rate boxes is not available, and fluid mechanics for laminar flow is used. The results from experiments are used to infer equations of state and chemical reaction rates that make the predictions agree with the experiments.

New work in detonation chemistry and quantum mechanics should lead us

to ways to fill in the chemical reaction rate box, and new work in statistical mechanics and new capabilities for static high-pressure experiments should tell us more about the equations of state.

Another subject that must be treated is the averaging needed to take account of the inhomogeneities in the reaction zone. The hot spots present in most of the solid materials we use in shock-wave physics usually can be neglected because they cause only small effects compensated for by slight adjustment of the constitutive relations. In detonation, hot spots cannot be neglected because they are the main effect; reaction begins in their neighborhood, and large composition changes take place there. Until we can take account of the inhomogeneities, we cannot use real equations of state and real chemical rates. However, a relatively new mathematical technique called "homogenization" may yield insight into this problem. Homogenization has been applied to systems with great internal complexity to derive rigorous equations that describe the average behavior of the system. For example, a nuclear reactor has many fuel elements and many control rods, and a detailed description of the action of each element and rod combines to give a very complex set of equations that describe the whole system. Homogenization gives a single equation that describes the large-scale behavior of the system, simple enough to be useful, but still reflecting the detailed small-scale variations of its many parts. Perhaps using similar techniques for explosives will allow us to describe the average behavior of an inhomogeneous reaction zone with the real properties rigorously included in the description. Bdzil is studying homogenization of the equations that describe detonation.

For the most part, I have described attempts to develop scientific understanding. Another topic is practical application. Does detonation research have value for engineering? It is always hard to guess what new things might come from research, but we can try to extrapolate. A safe explosive system has all its explosive parts near failure size, so that they operate well when they are initiated intentionally, but will always fail if they should be initiated accidentally. In such a system, the reaction zone is important, and we must understand it before we can make one. We are already a long way in that direction with the insensitive explosives now in use. Further, inhomogeneities control the failure radius and the reaction zone, and we must understand them if the manufacture of explosives is to be controlled satisfactorily. Any systems designed for safety will need knowledge of transients, edge effects, and failure.

The next step beyond good engineering of systems using explosives is designing the explosive itself. If we could fill out the boxes of Fig. 28 in detail, we wouldn't even have to make samples of the explosive molecules for testing or experiment to find the best particle-size distribution. Certainly we can't go that far very soon, but every step toward explosive design is important.

To close, I would like to address any newcomer to the business who feels diffident about speaking up in disagreement with the majority when so many of us have been working so long with explosives. Let me quote from E. Bright Wilson, Jr., who says, "No one can be so obstructive of progress as the 'expert' who has worked all his life on a single subject" ■

THE AUTHOR

William C. Davis has been a Staff Member at Los Alamos since 1954, the year he received his Ph.D. in physics from Johns Hopkins University after earning his Bachelor of Science in engineering from Tufts University in 1949. He has served as a Visiting Professor at the University of California, Riverside, and is the author of more than 40 reports and papers and the co-author of *Detonation*, one in the Los Alamos Series in Basic and Applied Sciences published by the University of California Press. At Los Alamos his field includes the details of detonation in high explosives, measurement of pressure in detonation, experimental methods for testing the theories of detonation, and development of instruments useful for detonation studies. Davis is a member of the American Physical Society, the Optical Society of America, the American Association of Physics Teachers, and the American Association for the Advancement of Science.

Further Reading

Ya. B. Zeldovich and S. A. Kompaneets, *Theory of Detonations* (Academic Press, New York, 1960).

C. H. Johansson and P. A. Persson, *Detonics of High Explosives* (Academic Press, New York, 1970).

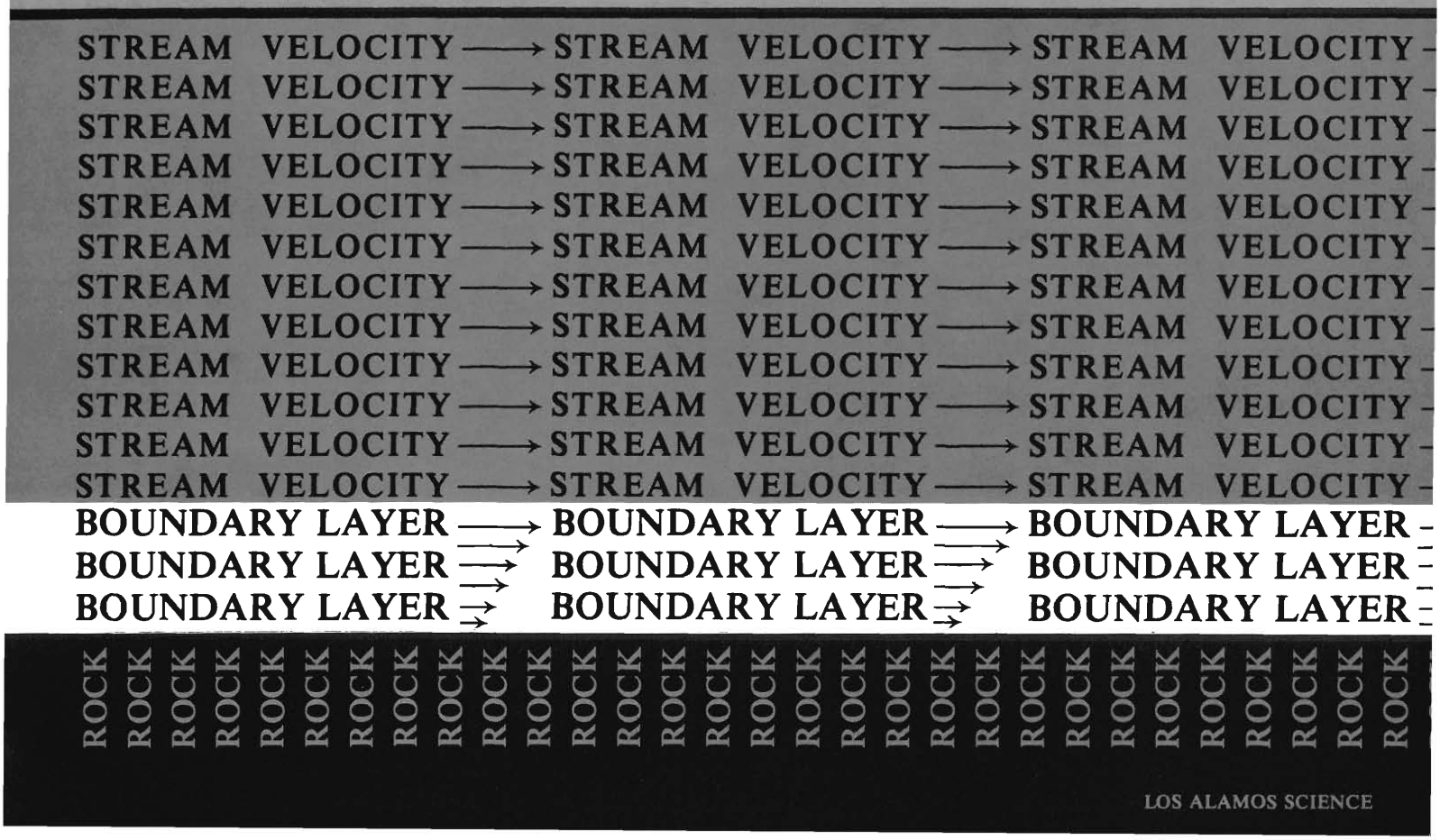
Wildon Fickett and William C. Davis, *Detonation* (University of California Press, Berkeley, 1979).

Charles L. Mader, *Numerical Modeling of Detonations* (University of California Press, Berkeley, 1980).

a novel approach to BOUNDARY LAYER PROBLEMS

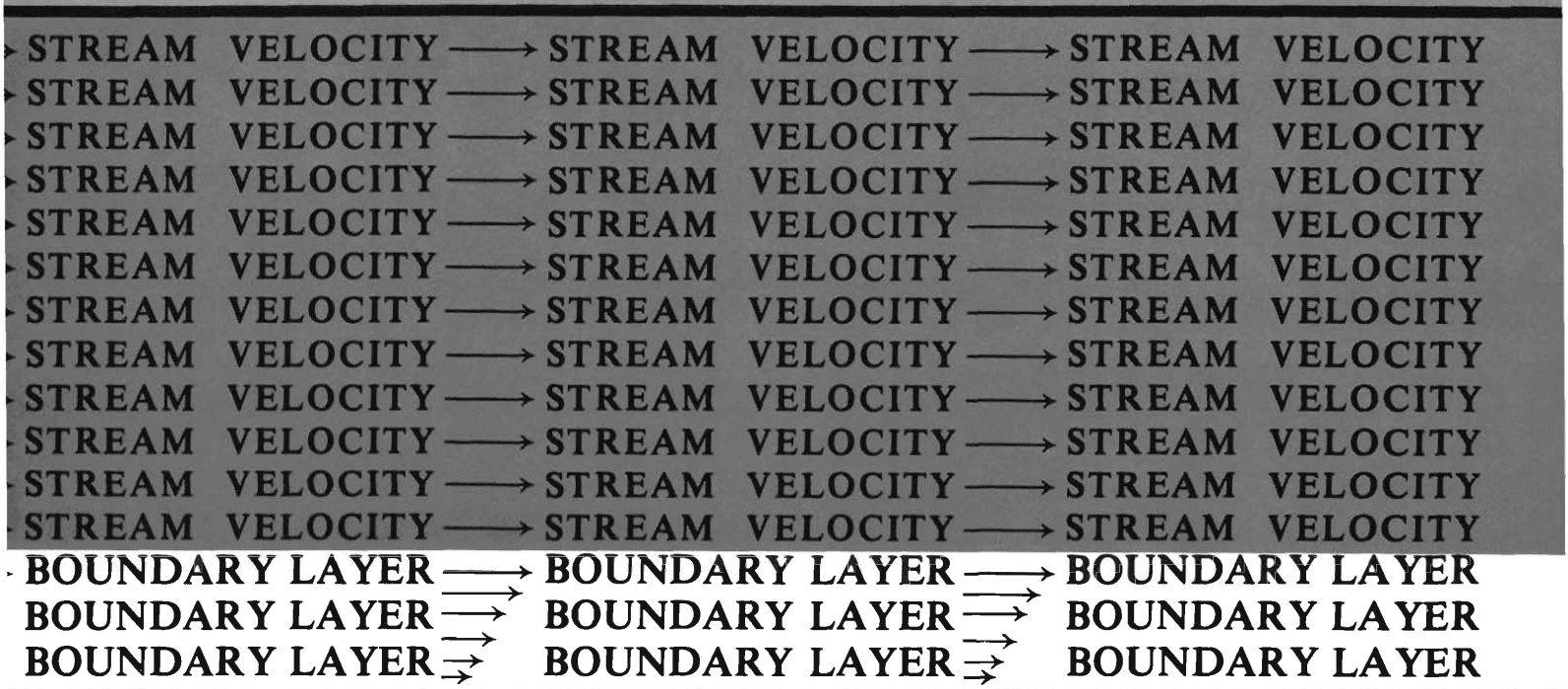
by Carl M. Bender

A radically new method for solving boundary-layer problems illustrates the cross-disciplinary nature of basic research. The method, which is applicable to singular perturbation problems in many fields, was developed in the context of particle physics.



Problems involving a rapid change in the value of a physical variable over a limited region of space or time are common in physics. They are referred to as boundary-layer problems. The most familiar examples occur in the motion of fluids. In these cases, a boundary layer refers to the very narrow region of fluid moving along a solid surface or boundary such as the inside of a pipe or a river bed; the fluid velocity in the boundary layer varies rapidly from zero at the boundary to the free-stream velocity away from the boundary. The boundary layer becomes narrower as the viscosity η decreases. This kind of dependence on viscosity means that the differential equation describing the fluid exhibits singular behavior and cannot be handled adequately by conventional perturbation methods when η is small.

Our new, very general approach to singular perturbation problems enables us to solve nonlinear and high-order boundary-layer problems without resorting to large-scale computer codes to study the systems numerically. The central idea consists of replacing the differential equation for the problem by a discrete equation defined on a lattice. After solving the discrete problem using regular perturbation techniques, we invoke special techniques to recover the continuum answers in the limit as the lattice spacing vanishes.



Perturbation Theory

Perturbation theory is a vast collection of mathematical methods used to obtain approximate solutions to problems that have no closed-form analytical solutions. The methods work by reducing a hard problem to an infinite sequence of relatively easy problems that can be solved analytically. Often, solving the first few of these provides an accurate approximation to the solution of the original hard problem.

A simple example illustrates the idea of perturbation theory. Consider the quintic polynomial

$$x^5 + x - 1 = 0 \ . \tag{1}$$

By drawing a graph of this equation and showing that its slope is always positive, we see that Eq. (1) has just one positive root. We cannot determine its numerical value analytically because there is no algebraic formula for the roots of a quintic polynomial. However, a perturbative approach reduces the problem to a sequence of almost trivial problems.

First, we introduce a perturbation parameter ε, which we specify to be small.

$$x^5 + \varepsilon x - 1 = 0 \ . \tag{2}$$

With ε small, we assume that the roots $x(\varepsilon)$ have a Taylor expansion in powers of ε.

$$x(\varepsilon) = a + b\varepsilon + c\varepsilon^2 + d\varepsilon^3 + \ldots \ . \tag{3}$$

Using this expansion we represent the terms in Eq. (2) as expansions:

$$x(\varepsilon)^5 = a^5 + 5a^4b\varepsilon + (5ca^4 + 10a^3b^2)\varepsilon^2 + (5a^4d + 20a^3bc + 10a^2b^3)\varepsilon^3 + \ldots$$

and

$$\varepsilon x(\varepsilon) = a\varepsilon + b\varepsilon^2 + c\varepsilon^3 + \ldots \ .$$

Substituting these expansions into Eq. (2), collecting the coefficients of like powers of ε, and setting them equal to zero gives a sequence of equations that are easily solved for the coefficients of Eq. (3).

ε^0) $a^5 - 1 = 0 \rightarrow a = 1$ (here we take the real root),
ε^1) $5a^4b + a = 0 \rightarrow b = -1/5$,
ε^2) $5ca^4 + 10a^3b^2 + b = 0 \rightarrow c = -1/25$,
ε^3) $5a^4d + 20a^3bc + 10a^2b^3 + c = 0 \rightarrow d = -1/125$,

and so on. We thus have determined that

$$x(\varepsilon) = 1 - \varepsilon/5 - \varepsilon^2/25 - \varepsilon^3/125 + \ldots \ .$$

The perturbation parameter ε plays a crucial role because it organizes the hard problem in Eq. (1) into a sequence of much easier problems, which in this case involve finding the coefficients a, b, c, d, ...

If we set $\varepsilon = 1$, we recover the original Eq. (1) from Eq. (2) and obtain a rapidly convergent series representation for its perturbative solution.

$$x(1) = 1 - 1/5 - 1/25 - 1/125 + \ldots \doteqdot 0.752 \ . \tag{4}$$

The exact value for x(1) found by solving Eq. (1) numerically on a computer is 0.754878 ... . Thus, our perturbation solution is impressively accurate. We could have found perturbation series for the other four (complex) roots by starting with any of the other four solutions to $a^5 - 1 = 0$.

Singular Perturbation Problems

The problem just considered is a *regular* perturbation problem; that is, its solutions vary smoothly as the perturbation parameter ε approaches zero. The problems of interest here are *singular* perturbation problems; their solutions change abruptly in some way as ε reaches zero. Some or all solutions either might cease to exist or might become infinite or degenerate. To illustrate, we can make Eq. (1) into a singular perturbation problem by introducing ε in a different way, for example,

$$\varepsilon x^5 + x - 1 = 0 \ .$$

This problem is singular because ε multiplies the highest power of x in the equation. As a result, when $\varepsilon \rightarrow 0$, the degree of the polynomial suddenly changes from 5 to 1. Moreover, a fifth-degree polynomial has five roots and a first-degree polynomial has one root, so in the limit $\varepsilon \rightarrow 0$ four roots disappear entirely. As $\varepsilon \rightarrow 0$, the roots move off to infinity in the complex plane according to $|x| \sim \varepsilon^{-1/4}$. (See Fig. 1.) The fact that the character of this problem changes abruptly as $\varepsilon \rightarrow 0$ identifies it as a singular perturbation problem.

The series expansions for singular perturbation problems are more complicated than the Taylor series in Eq. (3). Often they are not in Taylor form (a series in integer powers of ε), and usually they are divergent series. These problems sound so formidable that one might be tempted to avoid them completely. However, avoiding them is often not possible or even

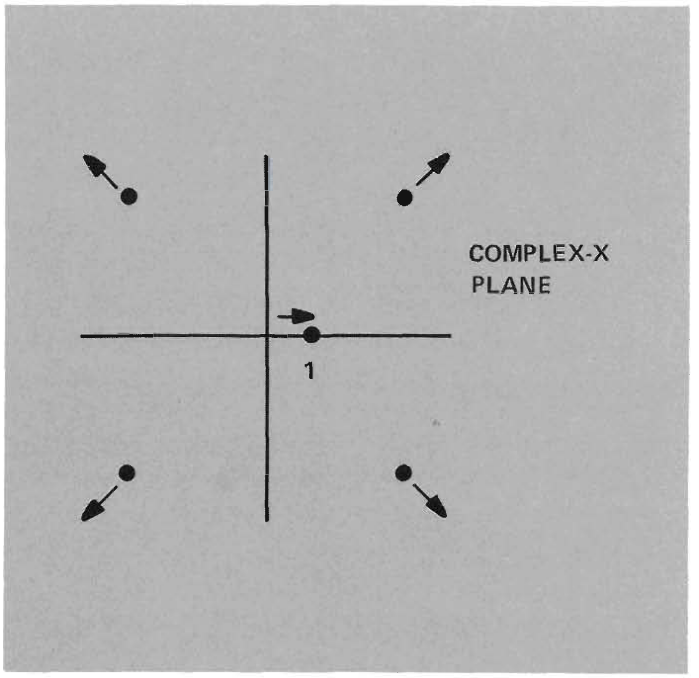


Fig. 1. The dots indicate the position of the five roots x(ε) of $\varepsilon x^5 + x - 1 = 0$ *as* $\varepsilon \to 0$. *One root approaches 1 while the other four move out to infinity according to* $|x| \sim \varepsilon^{-1/4}$.

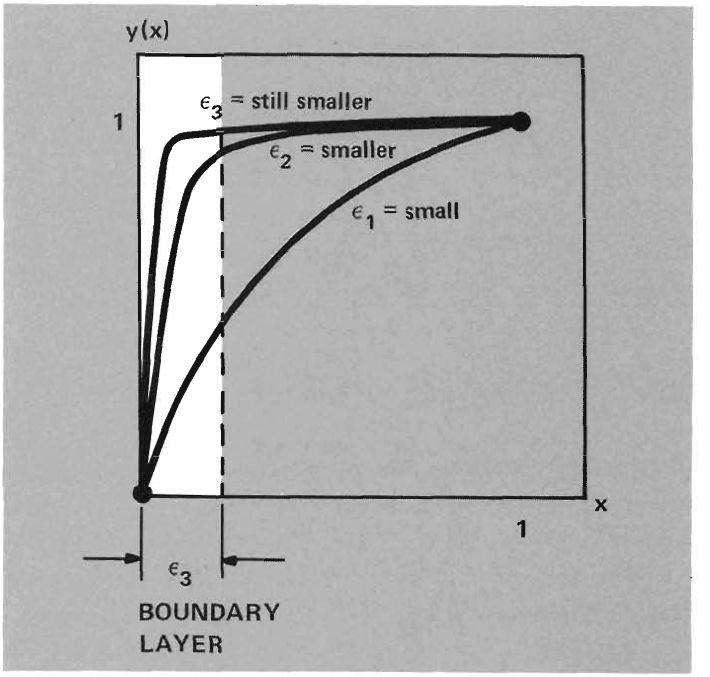


Fig. 2. A plot of the solution to the boundary-value problem $\varepsilon y'' + y' = 0$, $y(0) = 0$, $y(1) = 1$. *This is a singular perturbation problem because the curve in the boundary-layer region becomes steeper as* ε *approaches zero and becomes discontinuous when* ε *reaches zero. The boundary-layer region extends from* $x = 0$ *to* $x \simeq \varepsilon$.

desirable. In many physical problems, a natural small parameter ε occurs in such a way that the problem is automatically singular. For example, the time-independent Schrödinger equation,

$$-\frac{\hbar^2}{2m}\nabla^2\,\psi(\vec{x}) + [V(\vec{x}) - E]\,\psi\,(\vec{x}) = 0\ ,$$

contains the natural small parameter $\hbar^2/(2m)$. In the classical limit of the quantum system defined by $\hbar^2/2m \to 0$, this equation becomes a singular perturbation problem because its character changes abruptly from a differential equation to an algebraic equation.

Even when physics doesn't dictate that a perturbation problem be singular, we may want to introduce a perturbation parameter ε to make the problem singular. Although the resulting perturbation series may be divergent, it is usually asymptotic. (Roughly speaking, an asymptotic series $\Sigma\, a_n\varepsilon^n$ is one that diverges but also has the remarkable property that the first few terms provide an accurate approximation to the sum for sufficiently small ε.) Moreover, using modern summation techniques that have come into wider use over the last 20 years (Pade approximants and the Borel summation), we may obtain a more accurate numerical result from 3 or 4 terms of the divergent series than from 10 or 20 terms of a convergent series even when the perturbation parameter ε is large.

Boundary-Layer Theory

Boundary-layer problems, a special class of singular perturbation problems, provide the simplest context for introducing our new perturbative techniques. Nearly all boundary-layer problems are differential equation problems in which the highest derivative term is multiplied by a small parameter. A simple mathematical example will explain the appearance of a boundary layer.

Consider the boundary-value problem

$$\varepsilon\frac{d^2y}{dx^2} + \frac{dy}{dx} = 0\ ,\ y(0) = 0\ ,\ y(1) = 1\ . \qquad (5)$$

The exact solution to this problem is

$$y(x) = \frac{1 - e^{-x/\epsilon}}{1 - e^{-1/\epsilon}}\ ;$$

its graph for various values of ε is shown in Fig. 2. As ε decreases the curve becomes steeper in the region from $x = 0$ to $x \cong \varepsilon$. This region, in which y(x) exhibits transient behavior or rapid variation, is called a boundary layer. In fact, y(x) becomes discontinuous as ε decreases to zero. Equation (5) is a singular perturbation problem because the order of the differential equation changes abruptly from 2 to 1 as $\varepsilon \to 0$. Since a first-order differential equation cannot satisfy two independent boundary conditions, the solution ceases to exist when $\varepsilon = 0$ and a discontinuity appears in y(x).

Conventional Solutions

A review of the well-known conventional methods for obtaining the first approximation (the first term in the perturbation series) to the solution of a very general boundary-layer problem provides a contrast to our radically new approach.

We consider a boundary-value problem of the form

$$\varepsilon \frac{d^2y(x)}{dx^2} + a(x) \frac{dy(x)}{dx} + b(x)y(x) = 0 \ , \ y(0) = A \ , \ y(1) = B \ , \tag{6}$$

where ε is a small positive parameter. This equation is a generalization of Eq. (5). A and B are arbitrary numbers, and a(x) and b(x) are completely arbitrary continuous functions of x. From boundary-layer theory, we know that if a(x) is nonzero, the term a(x) [dy(x)/dx] acts as a "friction" force, which results in a boundary layer. If $a(x) > 0$ for $0 \leq x \leq 1$ (positive damping), the boundary-layer occurs immediately at $x = 0$. If $a(x) < 0$, the boundary layer occurs at $x = 1$. In general, the thickness of a boundary layer is determined by a scaling transformation on the differential equation. For Eq. (6) with a(x) positive, the boundary layer has thickness ε and lies in the region $x = 0$ to $x \cong \varepsilon$. (See Fig. 3.) Although Eq. (5) [a special case of Eq. (6)] is simply solvable, in general Eq. (6) has no closed-form analytic solution. As with the quintic polynomial in Eq. (1), we must use perturbation methods to obtain an approximate solution.

To solve the boundary-value problem in Eq. (6) approximately for small ε, we make several inspired observations, which allow us to approximate the solutions inside and outside the boundary-layer region and to match these solutions. First, we consider the boundary-layer region where the solution y(x) varies rapidly and is very steep. To be precise, scaling arguments referred to above tell us that a(x) [dy(x)/dx] is much larger than b(x)y(x) in the boundary-layer region (order $1/\varepsilon$ compared with order 1). (See Fig. 2.) Moreover, since the boundary layer from $x = 0$ to $x \cong \varepsilon$ is very narrow, we can approximate the function a(x) by a(0). Hence, inside the boundary-layer region, we may replace Eq. (6) by the much simpler equation

$$\varepsilon \frac{d^2y(x)}{dx^2} + a(0) \frac{dy(x)}{dx} = 0 \ \ (0 \leq x \lesssim \varepsilon) \ . \tag{7}$$

Equation (7) is easy to solve because it is a constant-coefficient equation. The solution that passes through A at $x = 0$ is

$$y(x) = (A - C) \exp[-a(0)x/\varepsilon] + C \ \ (0 \leq x \lesssim \varepsilon) \ , \tag{8}$$

where C is an arbitrary constant.

Second, we consider the region outside the boundary layer, where y(x) varies very slowly. Thus, $\varepsilon d^2y(x)/dx^2$ is small compared with $a(x)\, dy(x)/dx + b(x)y(x)$, and we may replace Eq. (6) by the first-order equation

$$a(x) \frac{dy(x)}{dx} + b(x)y(x) = 0 \ \ (\varepsilon \lesssim x \leq 1) \ . \tag{9}$$

The solution that passes through B at $x = 1$ is

$$y(x) = B \exp\left[\int_x^1 dt \ b(t)/a(t)\right] \ \ (\varepsilon < x < 1) \ . \tag{10}$$

To determine C in Eq. (8) we use the sophisticated perturbative method called asymptotic matching. Roughly speaking, we demand that Eqs. (8) and (10) agree in the region just to the right of the boundary layer, say at $x = \sqrt{\varepsilon}$. This gives

$$C \cong B \exp\left[\int_0^1 dt \ b(t)/a(t)\right] \ .$$

A simple and elegant expression combines the results in Eqs. (8) and (10) to give a good uniform approximation to y(x) over the entire region $0 \leq x \leq 1$.

$$y(x) = \left(A - B \exp\left[\int_0^1 dt \ b(t)/a(t)\right]\right) \exp[-a(0)x/\varepsilon] + B \exp\left[\int_x^1 dt \ b(t)/a(t)\right] \ . \tag{11}$$

The expression in Eq. (11) differs from the exact solution to Eq. (6) by terms of order ε. (See Fig. 3.)

This conventional approach to the solution of boundary-

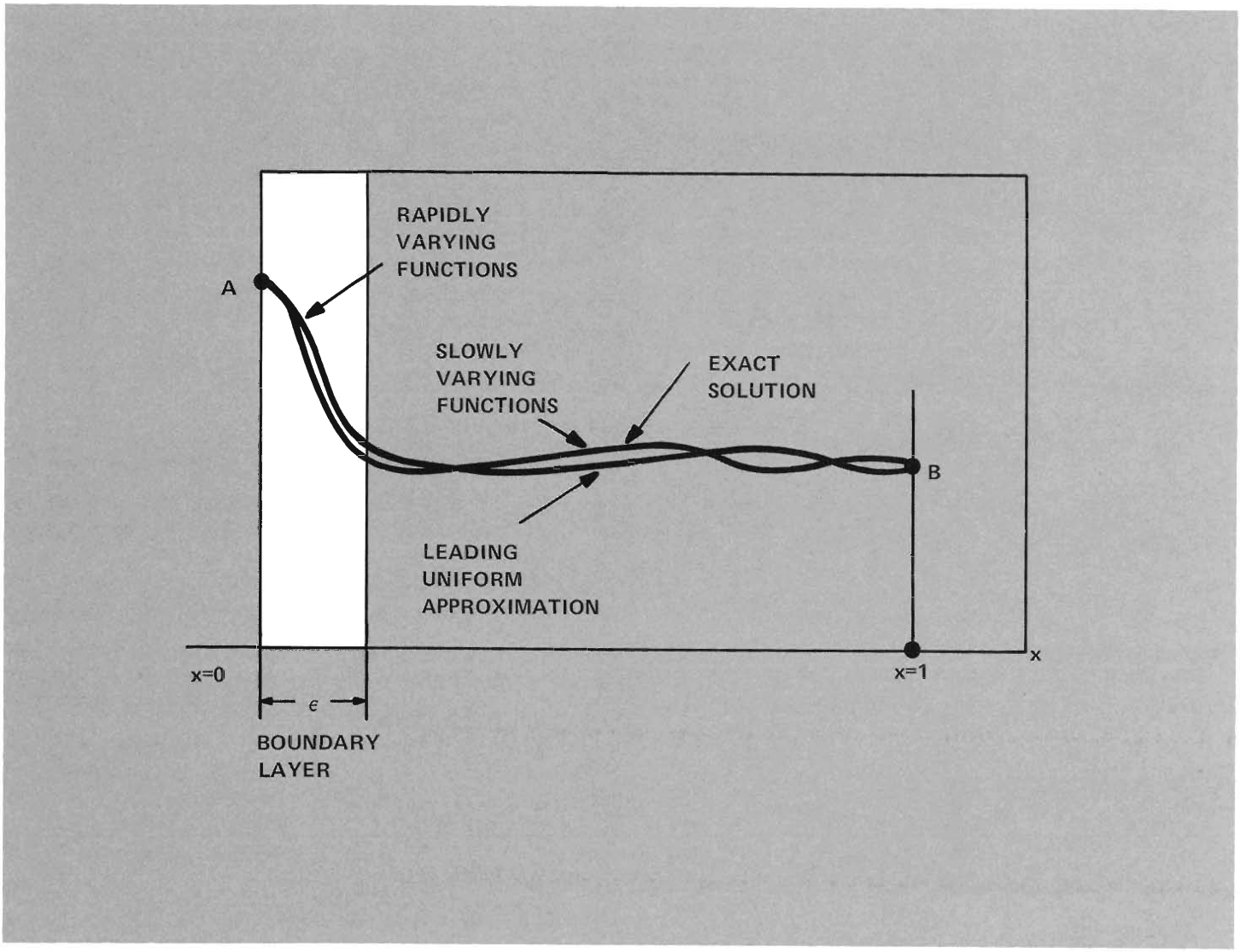


Fig. 3. A schematic comparison between the exact solution of the general linear boundary-layer problem in Eq. (6) and the leading-order uniform approximation to the exact solution in Eq. (11). The approximate solution differs from the exact solution by terms of order ε. *The boundary-layer region extending from* $x = 0$ *to* $x \cong \varepsilon$ *is shown.*

layer problems is widely known and well used. However, even the approximate equations that must be solved inside or outside the boundary layer cannot be solved analytically in some cases. Differential equations that are nonlinear or higher than second order present such difficulties; ordinarily they are solved numerically on a computer. With our new methods, these problems can be solved without recourse to large computer codes.

And Now For Something Completely Different

Our approach to boundary-layer problems is quite different; we will actually solve a singular boundary-layer problem as a regular perturbation problem. The approach has two parts. First, we replace the differential equation by a difference equation on a lattice. The replacement allows us to express the solution as a regular perturbation series in powers of ε as long as the lattice spacing a is held fixed and finite. Second, we take the lattice spacing a to zero to recover the answer to the original problem in the continuum. However, this limit is very peculiar because, as the lattice spacing goes to zero, the perturbation parameter ε goes to infinity. The perturbation series originally obtained in the limit of small ε is now infinite term by term. We use clever summation techniques to determine a finite answer—an answer that approximates the behavior in the boundary layer to surprisingly good accuracy.

We present the new perturbative methods in an application to the nonlinear two-dimensional wave equation for the function u(x,t).

$$\varepsilon^2 \frac{\partial^2 u}{\partial x^2} - \frac{\partial^2 u}{\partial t^2} + u - u^3 = 0 .$$

The equation has a static solution, called a kink, whose analytical form is

$$u(x) = \tanh \frac{x}{\varepsilon\sqrt{2}} . \tag{12}$$

The kink solution solves the boundary-value problem

$$\varepsilon^2 \frac{d^2u}{dx^2} + u - u^3 = 0,\ u(0) = 0,\ u(+\infty) = 1 . \tag{13}$$

We know that Eq. (13) is a singular perturbation problem because the highest derivative in the equation is multiplied by ε [as are the highest derivatives in Eqs. (5) and (6)]. Also, the kink solution Eq. (12) exhibits typical boundary-layer structure; u(x) varies slowly except in the boundary-layer region $-\varepsilon \leq x \leq \varepsilon$, where it rapidly goes from -1 to 1. (See Fig. 4.)

To find u(x) we begin by changing the differential equation to a difference equation. That is, we consider space to be made up of a lattice of distinct points with the spacing a between them held fixed. The points x are denoted by na (n = 0, 1, 2, ...) and u(x) becomes u(na) = u_n. The derivatives of u on the lattice become finite differences; in particular,

$$\frac{d^2u(x)}{dx^2} = \frac{1}{a^2}(u_{n+1} - 2u_n + u_{n-1}) .$$

The difference equation for Eq. (13) is thus

$$\frac{\varepsilon^2}{a^2}(u_{n+1} - 2u_n + u_{n-1}) + u_n - u_n^3 = 0,\ u_0 = 0,\ u_\infty = 1 . \tag{14}$$

In the first part of our approach, we hold the lattice spacing a fixed while we solve Eq. (14) perturbatively for small ε. With a held fixed and ε small, the natural expansion parameter for the problem is

$$\delta = \varepsilon^2/a^2 .$$

We therefore expand u_n for each value of n as a power series in δ,

$$u_n = a_n + b_n\delta + c_n\delta^2 + d_n\delta^3 + \ldots , \tag{15}$$

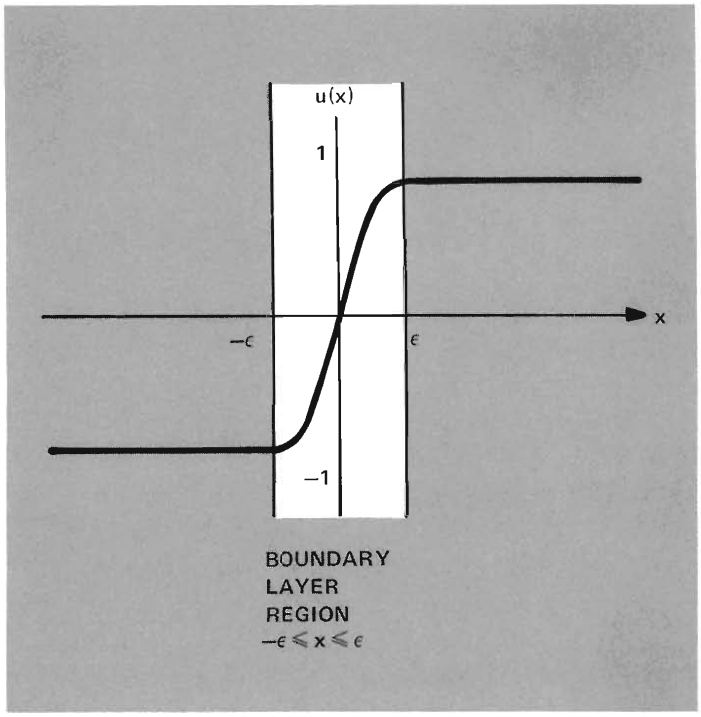


Fig. 4. The so-called kink solution to the equation $\varepsilon^2 u_{xx} + u - u^3 = 0$, $u(0) = 0$, $u(+\infty) = 1$. Observe that u(x) rapidly goes from -1 to 1 in the boundary-layer region when ε is small.

just as we did in Eq. (3) for the regular perturbation problem in Eq. (2).

We impose the initial conditions by taking

$$a_n = \begin{cases} 0 \text{ if } n = 0 , \\ 1 \text{ if } n \geq 1 . \end{cases}$$

Note that these conditions solve the unperturbed problem ($\delta = 0$) and thus follow the usual approach in regular perturbation problems. Substituting Eq. (15) into Eq. (14) and comparing powers of δ, just as we did in solving Eq. (2), routinely gives the perturbation coefficients. The u_n at the first few points on the lattice near x = 0 are

$$\begin{aligned} u_0 &= 0 \\ u_1 &= 1 - \tfrac{1}{2}\delta + \tfrac{1}{8}\delta^2 + \tfrac{11}{28}\delta^4 \\ u_2 &= 1 - \tfrac{1}{4}\delta^2 + \tfrac{5}{16}\delta^3 - \tfrac{15}{32}\delta^4 \\ u_3 &= 1 - \tfrac{1}{86}\delta^3 + \tfrac{9}{32}\delta^4 \\ u_4 &= 1 - \tfrac{1}{16}\delta^4 , \end{aligned} \tag{16}$$

where we have calculated u_n to the fourth order in δ and have lined up the contributions according to the order of the perturbation expansion in which the terms appear. (If k is the order of the expansion, the matrix $u_n^{(k)}$ is triangular for $k \geq 1$.)

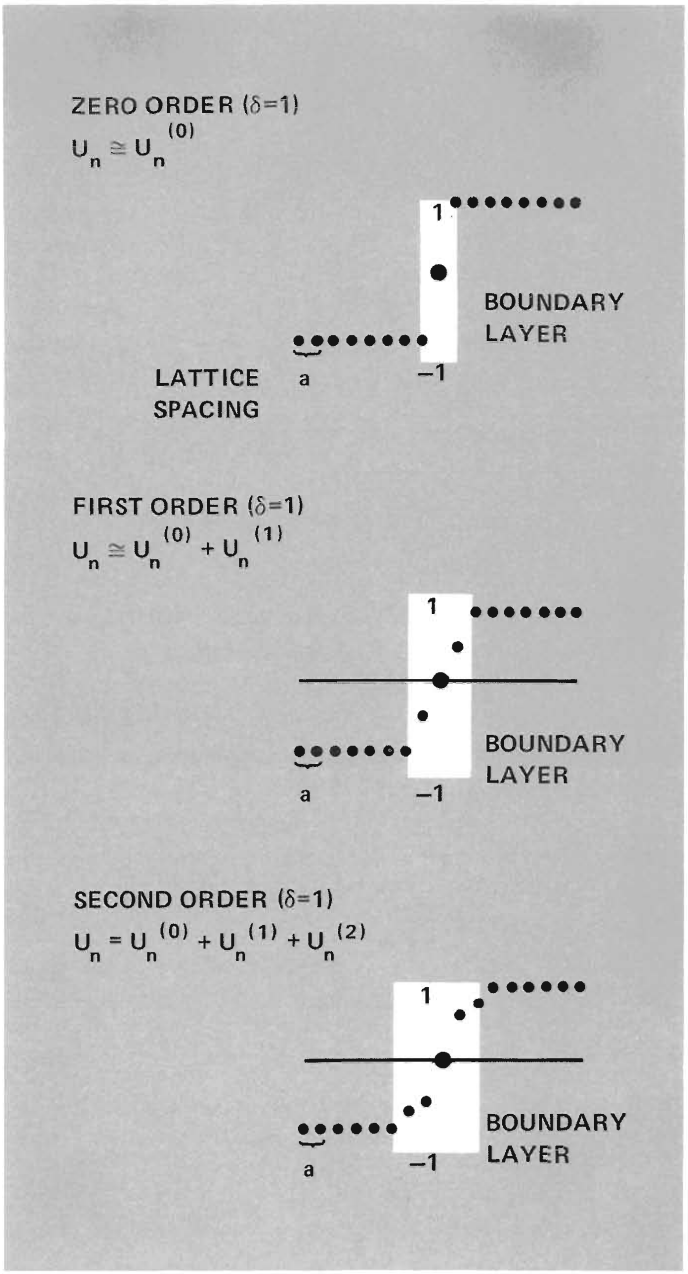


Fig. 5. The development of the boundary-layer structure as the order of the perturbation expansion increases. Results for u_n are plotted in zeroth-, first-, and second- order perturbation theory before taking the limit $\delta \to 0$. (For this plot we chose $\delta = 1$ and $a = \varepsilon = 1/4$.) Compare these results to the exact solution plotted in Fig. 4.

The boundary structure develops as we go to higher orders of the perturbation expansion (higher powers of δ). We determine one point in the boundary layer in the first order, two points in the second order, three points in the third order, and so on. (See Fig. 5.)

It is a peculiarity of our method that the thickness of the boundary layer is na in nth-order perturbation theory, and the boundary layer vanishes in the limit of zero-lattice spacing ($a \to 0$). Since all the points u_n shrink to the origin as $a \to 0$, we cannot determine the value of u(x) in the boundary layer by taking the continuum limit of Eq. (15) directly. However, the derivatives of u(x) at the origin are defined in this limit and can be determined from the usual definitions,

$$u'(0) = \lim_{a\to 0} \frac{u_1 - u_0}{a},$$

$$u''(0) = \lim_{a\to 0} \frac{u_1 + u_{-1} - 2u_0}{a^2},$$

and so on. In many cases, the derivatives themselves are of physical interest, and further, we can reconstruct the function in the boundary layer from its Taylor series,

$$u(x) = \sum_{n=0}^{N} u^{(n)}(0) \frac{x^n}{n!},$$

where

$$u^{(n)}(x) = \frac{d^n u(x)}{dx^n}.$$

How well can we determine the first derivative at the origin u′(0) from Eq. (14)? From the exact solution given by Eq. (12) we know that

$$u'(0) = \frac{1}{\varepsilon\sqrt{2}} = 0.70711/\varepsilon . \qquad (17)$$

Using the expansions in Eq. (16), we have the following result for the first derivative at the origin.

$$u'(0) = \lim_{a\to 0} \frac{1}{a}\left(1 - \frac{\delta}{2} + \frac{\delta^2}{8} + \frac{11}{128}\delta^4 + \ldots\right).$$

In the limit as $a \to 0$, δ is no longer small; in fact, $\delta = \varepsilon^2/a^2$ goes to infinity and

$$u'(0) = \frac{1}{\varepsilon} \lim_{\delta\to\infty} \sqrt{\delta}\left(1 - \frac{\delta}{2} + \frac{\delta^2}{8} + \frac{11}{128}\delta^4 + \ldots\right). \qquad (18)$$

The factor $1/\varepsilon$ in Eq. (18) agrees with the $1/\varepsilon$ in the exact result in Eq. (17). But does the limit of the series in Eq. (18) give $1/\sqrt{2}$? This limit should be understood as follows: our perturbation series has been derived by assuming that a is fixed and that ε and therefore δ are small. Now we must return to the continuum problem in Eq. (13) by taking $a \rightarrow 0$ and keeping ε fixed; therefore we must take the limit $\delta \rightarrow \infty$. Apparently we have a disaster—all terms in Eq. (18) diverge, and what is more, each new term approaches infinity faster than the preceding term because it has two additional powers of a in its denominator.

Everything has been routine until now; why do such unpleasant difficulties arise? The answer is that we have been treating what is fundamentally a singular perturbation problem as a regular perturbation problem by seeking perturbation series in the form of Taylor series [Eq. (5)], and the mathematics is trying to make us pay for our naivety. We will return to Eq. (18) after we have resolved this difficulty.

How To Sum The Series $\infty + \infty + \infty + \infty + \ldots$ And Get a Finite Answer

An elementary example will explain how the series in Eq. (18) can have a finite sum in the limit $a \rightarrow 0$. Suppose we are asked to expand $\sqrt{x}$ in a Taylor series about $x = 0$. The request seems unreasonable because this function does not have such a series (all derivatives of $\sqrt{x}$ are singular at $x = 0$). Nevertheless, we have an idea. We introduce a lattice spacing a and consider the function $(x + a)^{1/2}$, which does have a Taylor series about $x = 0$.

$$(x + a)^{1/2} = a^{1/2} + \frac{x}{2a^{1/2}} - \frac{x^2}{8a^{3/2}} + \ldots \quad . \tag{19}$$

In the limit $a \rightarrow 0$ every term after the first becomes infinite, just as in Eq. (18), but the *sum* of the series is not necessarily infinite. Indeed, if we sum the series first, we obtain $(x + a)^{1/2}$; then we can take the limit $a \rightarrow 0$ and obtain the finite answer $\sqrt{x}$.

Thus we would like to sum the series in Eq. (18) before we take the limit $a \rightarrow 0$ or $\delta \rightarrow \infty$. Although we know all the terms in the series in Eq. (19), we know only a finite number, say $N + 1$, of terms in Eq. (18). If we know only $N + 1$ terms, how can we trick the series in Eq. (18) into revealing its approximate finite sum?

We have developed a general procedure for summing a series with $N + 1$ terms. Suppose we calculate a quantity $Q(\varepsilon)$ as a series in powers of ε.

$$Q(\varepsilon) = \varepsilon^{\alpha} \sum_{n=0}^{\infty} a_n \varepsilon^n \quad , \tag{20}$$

where α is an arbitrary nonnegative exponent. Here we have generalized the notion of a Taylor series slightly to include the possibility of an overall multiplicative fractional power of ε. This series is just like the series in Eq. (18) if we set $\varepsilon = \delta$ and $\alpha = 1/2$. Now we can calculate $Q(\infty)$ knowing only $N + 1$ terms in the series, if we know that $Q(\infty)$ is finite.

We proceed to manipulate the series in Eq. (20) until we create a structure that has a finite limit as $\varepsilon \rightarrow \infty$. First, we raise both sides of Eq. (20) to the power $1/\alpha$.

$$Q(\varepsilon)^{1/\alpha} = \varepsilon \sum_{n=0}^{\infty} (a_n \varepsilon^n)^{1/\alpha} \quad .$$

Using the usual rules for exponentiating a Taylor series, we can rewrite $Q^{1/\alpha}$ in the following form.

$$Q(\varepsilon)^{1/\alpha} = \varepsilon \sum b_n \varepsilon^n$$

$$= \frac{\varepsilon}{\sum c_n \varepsilon^n} \quad .$$

Finally, we raise this equation to the integer power N.

$$Q(\varepsilon)^{N/\alpha} = \frac{\varepsilon^N}{\sum_{n=0}^{\infty} C_n^{(N)} \varepsilon^n} \quad . \tag{21}$$

For all these Taylor series manipulations, we assume ε is small.

Since we know only $N + 1$ terms in the series in Eq. (20), we must work consistently and truncate the series in Eq. (21) at $n = N$. Now we have a structure that is well-defined in the limit $\varepsilon \rightarrow \infty$; in fact, as $\varepsilon \rightarrow \infty$, the only term that survives is $1/C_n^{(N)}$.

$$\lim_{\varepsilon\to\infty} Q(\varepsilon)^{N/\alpha} \sim \lim_{\varepsilon\to\infty} \frac{\varepsilon^N}{\sum_{n=0}^{N} C_n^{(N)}\,\varepsilon^n} = \frac{1}{C_N^{(N)}} .$$

Thus we have obtained an approximate value for $Q(\infty)$, which we call Q_N.

$$Q_N = [C_N^{(N)}]^{-\alpha/N} . \tag{22}$$

We refer to Q_N as the Nth approximant to $Q(\infty)$; in many problems Q_N tends to $Q(\infty)$ as $N \to \infty$. More importantly, in most problems Q_N rapidly gets very close to $Q(\infty)$ when N is still quite small. Through a sequence of simple series manipulations we have converted the series in Eq. (20), which is meaningless as $\varepsilon \to \infty$, into a series of well-defined extrapolants Q_N.

An Example

The convergent properties of the extrapolants Q_N are so surprising that we must demonstrate them with a specific example. The transcendental equation

$$\ln x + \frac{1}{1-x} = 0 \tag{23}$$

has one root between 0 and 1:

$$x = 0.2592463566483 \ldots .$$

Let us see if we can obtain this root perturbatively. We begin by introducing a perturbation parameter ε in a most unusual way.

$$\ln x + \frac{1}{1+\varepsilon}\left(\frac{\varepsilon}{1-x} + 1\right) = 0 . \tag{24}$$

We have chosen to introduce ε this way for two reasons. First, the original equation in (23) is recovered in the limit $\varepsilon \to \infty$. Second, the unperturbed problem [Eq. (24) with $\varepsilon = 0$] is easy to solve: $\ln x + 1 = 0 \to x = 1/e$. With this solution for the unperturbed problem we proceed to find the rest of the perturbation series for $\varepsilon \neq 0$ by the iterative methods discussed earlier for regular perturbation problems. The series for the root $x(\varepsilon)$ to Eq. (24) is

$$x(\varepsilon) = \frac{1}{e}(1 - 0.58198\varepsilon + 1.28714\varepsilon^2 - 3.16768\varepsilon^3 \ldots) . \tag{25}$$

We would like to sum this series as $\varepsilon \to \infty$ using the general formula in Eq. (22), which requires that $\alpha \neq 0$. To convert the series to a form in which $\alpha = 1$, we multiply Eq. (25) by e, take the natural logarithm, and multiply by -1.

$$\begin{aligned} Q(\varepsilon) &= -\ln[e\,x(\varepsilon)] \\ &= \varepsilon(0.58198 + 1.11779\varepsilon + 2.48430\varepsilon^2 \\ &\quad - 6.26489\varepsilon^3 + \ldots) . \end{aligned} \tag{26}$$

The extrapolants for Eq. (26) defined by Eq. (22) yield a very rapidly convergent sequence for the roots of the transcendental equation in Eq. (23).

$x_1 = 0.271713639$,
$x_2 = 0.255145710$,
$x_3 = 0.260300667$,
$x_4 = 0.258935592$,
$x_5 = 0.259336423$,
$x_6 = 0.259219343$,
$x_7 = 0.259254556$,
$x_8 = 0.259243826$,
$x_9 = 0.259247147$,
$x_{10} = 0.259246107$,
$x_{11} = 0.259246436$,
$x_{12} = 0.259246331$,
$x_{13} = 0.259246365$, and
$x_{14} = 0.259246353$,

which is now correct to one part in 10^8.

Note that the extrapolants Q_N, which are all positive, are derived from a perturbation expansion that has both positive and negative terms. The remarkable fact that the first N coefficients in the series raised to the Nth power are all positive ensures that the extrapolants Q_N are always real.

a spin-off from PARTICLE PHYSICS RESEARCH

Our new techniques for singular perturbation problems were developed to solve mathematical problems that arise in a field-theoretic treatment of strongly interacting particles. Strong forces mean that the parameter describing the interactions between particles is very large compared to other parameters in the problem. This problem is very different from the weak-coupling problems usually studied (such as those arising in quantized electromagnetic interactions), in which the interaction force is small (weak) and thus can be treated as a perturbation on a system of freely moving, noninteracting particles. The answers to weak-coupling problems are expressed as power series in the strength of the small interaction force.

In the strong-coupling case, we cannot treat the interaction term as a perturbation since it dominates the dynamics. Instead, we treat the kinetic energy term that determines the motion of free particles as the perturbation. The unperturbed system now consists of particles that interact through strong forces but are motionless or frozen in space time.

The strong-coupling or kinetic energy expansion is singular because the kinetic energy can become arbitrarily large. As a result, the terms in the perturbation expansion in inverse powers of the coupling strength are not well defined. To proceed we use an artifice: we model space-time as a lattice of discrete points rather than as a continuum. This trick, commonly used in particle physics, prevents the momentum or kinetic energy from becoming arbitrarily large. Particle motion is no longer continuous but rather consists of hopping from site to site. Although the dynamics on the lattice is unfamiliar, the strong-coupling expansion of physical quantities becomes well defined and very easy to compute, so easy that the computations are purely algebraic and we can program computers to do the manipulations to a very high order in the perturbation expansion. However, the introduction of a space-time lattice has a major drawback; it introduces into the problem an artificial length, namely, the lattice spacing a between lattice sites. To obtain physically meaningful results, we must return to the continuum by taking the lattice spacing a to zero. Performing this singular and difficult limit was a central problem we solved in our research.

We now realize that the introduction of a lattice and the subsequent continuum limit $a \rightarrow 0$ is a powerful mathematical tool that can be applied to many singular perturbation problems outside the realm of quantum field theory. In general, it has the advantage of converting *singular* perturbation problems that require a great deal of mathematical subtlety and ingenuity into *regular* perturbation problems whose iterative solutions are straightforward and routine. We have used these methods to solve a variety of singular perturbation problems such as boundary-layer problems, and we have even used them to elucidate the statistical mechanics of randomly driven nonlinear oscillators ■

Why Equation (22) Works

We can develop some intuition for why Eq. (22) defines a convergent series of approximants by doing a saddle-point evaluation of a complex integral. For those not interested in this argument, skipping to the next subheading will not break the continuity of our presentation.

If f(z) is an analytic function in a region of the complex-z plane containing the origin, the nth term in the Taylor series for

$$f(z) = \sum_{n=0}^{\infty} A_n z^n$$

is given by a contour integral,

$$A_n = \frac{1}{2\pi i} \int_C \frac{f(z)}{z^{n+1}} dz ,$$

where the contour C encircles the origin. Hence, after making some reasonable analyticity assumptions, we may use this formula to solve Eq. (21) exactly for $C_N^{(N)}$.

$$C_N^{(N)} = \frac{1}{2\pi i} \int_C \frac{d\varepsilon}{\varepsilon} Q(\varepsilon)^{-N/\alpha} . \qquad (27)$$

We would like to examine the behavior of $C_N^{(N)}$ for large N to see if the sequence of approximants converges. To do so, we rewrite the integrand as

$$\exp[-N \ln Q(\varepsilon)/\alpha] .$$

For large N, we apply the saddle-point method. A saddle point ε_0 is defined by the condition

$$Q'(\varepsilon_0) = 0 . \qquad (28)$$

We shift the contour C in Eq. (27) until it passes through the saddle point ε_0 (assuming for now that there is only one) and for the sake of simplicity we ignore any contributions that might come from passing the contour through a singularity in the complex-s plane. The saddle-point method tells us that the most important contribution to the integral in Eq. (27) is found by evaluating the integrand at the saddle point ε_0. (We will ignore any terms that depend on N algebraically because, as we will see, we are going to take the Nth root of the result and let N tend to infinity; in this limit all algebraic terms approach 1.) Evaluating the integral at the saddle point gives

$$C_N^{(N)} \cong \frac{1}{2\pi i} \frac{1}{\varepsilon_0} Q(\varepsilon_0)^{-N/\alpha} .$$

Using the definition of the extrapolants Q_N in Eq. (21), we have

$$Q_N = [\, C_N^{(N)}]^{-\alpha/N} = \left(\frac{1}{2\pi i} \frac{1}{\varepsilon_0}\right)^{-\alpha/N} Q(\varepsilon_0) .$$

Thus we have $Q_N \cong Q(\varepsilon_0)$, where the error approaches zero as $N \to \infty$. Hence,

$$\lim_{N\to\infty} Q_N = Q(\varepsilon_0) .$$

This result may seem very disappointing at first because we were hoping that

$$\lim_{N\to\infty} Q_N = Q(\infty) .$$

However, we assumed at the start that Q(m) was finite. This is roughly the same as assuming that $Q(\varepsilon)$ is level at $\varepsilon = \infty$, which means that it smoothly approaches the constant $Q(\infty)$] and thus satisfies the saddle-point condition Eq. (28) at $\varepsilon_0 = \infty$. In effect, we have assumed at the start that $\varepsilon_0 = \infty$ is a saddle point. Now we understand why the formula Eq. (22) can produce a series of approximants that actually approach Q(a)).

This argument also explains why Eq. (22) sometimes fails; it can fail if there is another saddle point in the complex-c plane that gives a larger contribution than the one at ∞.

Here is the surprising part: although there are many problems for which the sequence of approximants Q_N doesn't converge to $Q(\infty)$, in all 25 or so problems that we have investigated, both in quantum-field theory and in boundary-layer theory, the Q_N sequence still gives a remarkably accurate approximation to the exact answer. Sometimes the approximants Q_N come very close to the exact answer and then veer away, just like the partial sums of an asymptotic series. We do not really understand yet why our method works so well.

Back To The Kink ...

Now that we have described a general method for extrapolating series like the one in Eq. (18) to the continuum, we will return to the kink problem and see how well our method works. The exact answer for the series in Eq. (18) is $1/(\varepsilon\sqrt{2})$, given by Eq. (17). Thus, we would like the series in powers of δ to approach $1/\sqrt{2}$ as $\delta \rightarrow \infty$. We use the general formula in Eq. (22) with $\alpha = 1/2$ to obtain the following sequence of approximants.

$Q_1 = 1.0$,
$Q_2 = 0.84090$,
$Q_3 = 0.78193$,
$Q_4 = 0.75724$,
$Q_5 = 0.74076$,
$Q_6 = 0.73121$,
$Q_7 = 0.72393$,
$Q_8 = 0.71905$,
$Q_9 = 0.71515$, and
$Q_{10} = 0.71231$.

The extrapolants are already very close to the exact answer 0.70711. The higher extrapolants continue to decrease monotonically until a surprising thing happens; they undershoot the exact answer and continue decreasing until they reach a minimum in 24th order.

$Q_{24} = 0.70198$.

The relative error between this value and the exact answer is less than 1%. The extrapolants gradually rise monotonically until they recross the exact answer in 41st order; we believe that the extrapolants will continue to rise from here on.

Unlike the example in Eq. (23), the sequence of approximants is not convergent, probably because of the effect of a saddle point. Nevertheless, the sequence is asymptotic in nature; like Stirling's series for the gamma function and other asymptotic series, early terms in the series comprise a good approximation to the answer until some optimal order is reached. Afterwards, the direct approximants from these series diverge.*

In the same way that we determined $u'(0)$, we can determine all the derivatives $u^{(n)}(0)$ for Eq. (18) by extrapolating series expansions of the form

$$u^{(n)}(0) = \left(\frac{\delta}{\varepsilon^2}\right)^{n/2} \sum a_m^{(n)} \delta^m \ .$$

Then, we can use the Taylor series for $u(x)$ for $x < x_0$ where $u(x_0) = 1$, and we can set $u(x) = 1$ for $x \geq x_0$ to get a reasonable global reconstruction of $u(x)$. However, we need to know at least 10 terms in the Taylor series to perform this reconstruction.

The method we have described to determine $u(x)$ relies on obtaining local information for the differential or classical field equation at the origin. In his paper, "Singular Perturbation-Strong Coupling Field Theory," Carlos R. Handy, a Postdoctoral Fellow at Los Alamos, developed an alternative approach, which allows for global reconstruction of the field solution. The method combines two mathematical tools. The first is the lattice expansion for the given field equation; an example is Eq. (15), from which the power moments are determined as an expansion in inverse powers of the lattice spacing. Handy uses Padé approximant techniques to obtain approximate continuum limit power moments. The second relates to the traditional, mathematical "moments problem." After obtaining a sufficient number of the approximate continuum moments, he reconstructs the corresponding approximate global field solution. This procedure gave excellent results for both the ϕ^4-classical field theory kink and the Sine-Gordon equation kink solutions. It is equivalent to a momentum space formulation of the problem in which the long-range, large-scale behavior of the fields is determined by the small-momentum infrared domain.

The perturbation series derived from conventional boundary-layer methods are also asymptotic divergent series.

Other Boundary-Layer Problems

Boundary layers or transient phenomena arise in many diverse physical settings. Here we apply the new solution techniques to three problems.

1. Blasius Equation

The Blasius equation describes the boundary-layer structure of fluid flow across a flat plate.

$$2\varepsilon\, y'''(x) + y(x)y''(x) = 0,\ y(0) = y'(0) = 0,\ y'(+\infty) = 1\ .$$

A quantity of physical interest is $y''(0)$, which determines the stress on the plate apart from dimensional parameters.

To solve for $y''(0)$ perturbatively, we rewrite the Blasius equation on a lattice.

$$2\delta(f_{n+1} - 3f_n + 3f_{n-1} - f_{n-2}) + f_n(f_{n+1} - 2f_n + f_{n-1}) = 0\ ,$$

where $f_n = y(na)/a$, $\delta = \varepsilon/a^2$, and a is the lattice spacing. As usual, we solve for f_n as a series in powers of δ and obtain from the solution a sequence of extrapolants for $y''(0)$. The exact value for $y''(0)$ obtained numerically is $(0.33206...)\ \sqrt{\varepsilon}$. The first few extrapolants obtained by our new techniques are

$$Q_1 = 0.5/\sqrt{\varepsilon}\ ,$$
$$Q_2 = 0.4204/\sqrt{\varepsilon}\ ,$$
$$Q_3 = 0.3948/\sqrt{\varepsilon}\ ,$$
$$Q_4 = 0.3819/\sqrt{\varepsilon}\ , \text{ and}$$
$$Q_5 = 0.3742/\sqrt{\varepsilon}\ .$$

As N increases, Q_N becomes very flat.

$$Q_{25} = 03502/\sqrt{\varepsilon}\ ,$$
$$Q_{26} = 0.3500/\sqrt{\varepsilon}\ ,$$
$$Q_{37} = 0.3485/\sqrt{\varepsilon}\ , \text{ and}$$
$$Q_{38} = 0.3484/\sqrt{\varepsilon}\ .$$

The relative error between the exact answer and Q_{38} is about 5%. At present, we do not know whether the sequence Q_N approaches the exact answer as $N \to \infty$. In fact, there are many ways to extrapolate Q_N to its limiting value Q_∞, but we will not discuss them here.

2. Damped Linear Oscillator

An initially quiescent spring-mass system subject to an impulse I_0 at $t = 0$ satisfies the equation

$$m\frac{d^2y}{dt^2} + \beta\frac{dy}{dt} + ky = I_0\delta(t)\ ,\ y(0\_) = y'(0\_) = 0\ ,$$

where β is the damping coefficient, m is the mass, and k is the spring constant. For small m, the solution y(t) exhibits a boundary layer of thickness m/β situated at $t = 0$. The exact solution satisfies $y'(0^+) = I_0/m$.

The lattice version of the differential equation is

$$\varepsilon(y_{n+1} - 2y_n + y_{n-1}) + y_{n+1} - y_n + kay_n/\beta = I_0\delta_{n,0}/\beta\ ,$$

where $y(an) = y_n$ and $\varepsilon = n/(\beta a)$. On the lattice the perturbation series for the damping term at $t = 0$, $y'(0)$, is very simple.

$$y'(0) = \lim_{\varepsilon\to\infty} \varepsilon\frac{I_0}{m}(1 - \varepsilon + \varepsilon^2 - \varepsilon^3\ ...)\ .$$

Using Eq. (22), the formula for the Nth approximant Q_N, we obtain for all N

$$Q_N = \frac{I_0}{m}\ .$$

Thus, our perturbative approach gives the exact answer to all orders.

3. Green's Function for the Diffusion Equation

Our lattice techniques work for partial as well as for ordinary differential equations. To illustrate, we consider a heat diffusion equation with a point source in the space and time variables.

$$u_t = \upsilon u_{xx} + \delta(t)\delta(x)\ ,\ u(x,t) = 0\ (t < 0)\ ,$$

where υ is the thermal diffusivity. The solution to this equation u(x,t) describes the temperature distribution in a one-dimensional system like a wire or rod.

The exact solution is the Green's function,

$$u(x,t) = \frac{\theta(t)}{(4\pi\upsilon t)^{1/2}} \exp(-x^2/4\pi\upsilon t) , \qquad (29)$$

where

$$\theta(t) = \begin{matrix} 1 \text{ if } t \geq 0 \\ 0 \text{ if } t < 0 \end{matrix} .$$

Note that for small υ and fixed t, the temperature distribution u(x,t) has a boundary layer of thickness $(\upsilon t)^{1/2}$ at x = 0. We wish to calculate the temperature at x = 0 and time t. The exact result from Eq. (29) is

$$u(0,t) = \frac{\theta(t)}{(4\pi\upsilon t)^{1/2}} = (0.282095)\frac{\theta(t)}{(\upsilon t)^{1/2}} .$$

To apply our solution methods, we introduce a discrete lattice in the spatial variable to obtain the differential-difference equation

$$\frac{d}{dt} u_n = \varepsilon(u_{n+1} - 2u_n + u_{n-1}) + \frac{\delta(t)\delta_{n,0}/a}{a} ,$$

where $\varepsilon = \upsilon/a^2$. The perturbation series solution for u_0, the temperature at x = 0, is

$$u_0(t) = \lim_{a\to 0} \frac{1}{a} \theta(t) \sum_{k=0}^{\infty} \frac{(-\varepsilon t)^k (2k)!}{(k!)^3}$$

$$= \frac{\theta(t)}{(\upsilon t)^{1/2}} \lim_{\varepsilon t\to\infty} (\varepsilon t)^{1/2} \sum_{k=0}^{\infty} \frac{(-\varepsilon t)^k (2k)!}{(k!)^3} .$$

This series can be summed exactly to give

$$u_0(t) = \theta(t)\,(\varepsilon/\upsilon)^{1/2} \exp(-2t\varepsilon)\, I_0(2t\varepsilon) ,$$

where I_0 is an associated Bessel function. Using the asymptotic relation $I_0(x) \sim e^x/(2\pi x)^{1/2}$ $(x \to +\infty)$, we can take the limit $\varepsilon \to \infty$ to obtain the exact answer in Eq. (29).

However, we are more interested in finding out what happens when we extrapolate the perturbation series term-by-term to the limit using Eq. (22). We obtain a sequence of extrapolants, which appear to converge rather slowly after they have been divided by $\theta(t)/(\upsilon t)^{1/2}$.

$Q_1 = 0.5$,
$Q_2 = 0.435$,
$Q_3 = 0.408$,
$Q_4 = 0.393$, and
$Q_5 = 0.384$.

The sequence continues to approach the exact answer but becomes very flat as N increases.

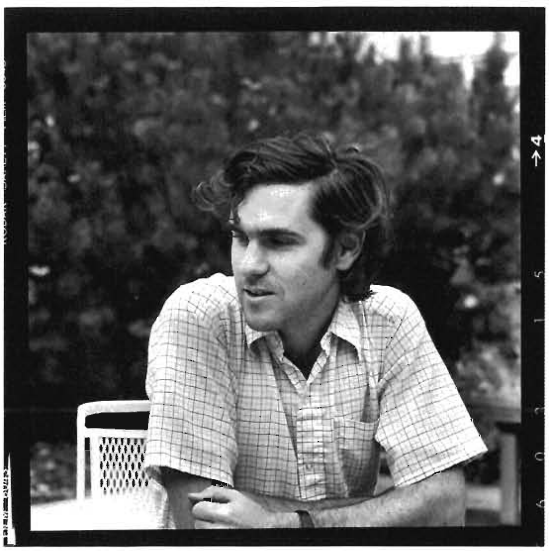
$Q_{10} = 0.362$,
$Q_{15} = 0.354$,
$Q_{20} = 0.349$,
$Q_{25} = 0.346$,
$Q_{30} = 0.344$,
$Q_{35} = 0.343$, and
$Q_{40} = 0.342$.

The last approximant differs from the exact answer by about 18%. This example gives the poorest results we have found so far; in most problems we have studied, we can predict the answer to within a few percent.

Conclusions

Our new way of doing perturbation theory, in which the perturbation parameter ε initially is assumed to be small and eventually is extrapolated to infinity, appears to be a powerful tool in boundary-layer theory and in many other areas. Although the method involves taking unusual limits, the computations are purely algebraic and therefore relatively simple. There is much work to be done in determining the method's full range of applicability. We have returned to its application in field theory and hope that investigators in other areas will find ways to exploit the methods in new physical contexts ■

THE AUTHOR



Carl Bender, Visiting Staff Member in the Laboratory's Theoretical Division, is a Professor of Physics at Washington University. He received his A.B. from Cornell University in 1964, his Ph.D. in Theoretical Physics from Harvard University in 1969, and held a one-year postdoctoral appointment at the Institute for Advanced Study in Princeton. Until 1977 he was an Associate Professor of Applied Mathematics at Massachusetts Institute of Technology where he received the Graduate Teaching Award in 1976. He was awarded a Sloan Foundation Fellowhip in Theoretical Physics and Applied Mathematics. Previously, he was a visiting Research Scientist at Imperial College, London, and at California Institute of Technology. He has published widely in the fields of Applied Mathematics and Theoretical Physics. Currently he is on the Editorial Board of Advances in Applied Mathematics. He is coauthor of *Advanced Mathematical Methods for Scientists and Engineers*, which is used as a standard text for graduate scientists.

Further Reading

1. C. M. Bender, F. Cooper, G. S. Guralnik, and D. H. Sharp, "Strong-Coupling Expansion in Quantum Field Theory," Phys. Rev. D **19**, 1865 (1979).

2. C. M. Bender, F. Cooper, G. S. Guralnik, R. Roskies, and D. H. Sharp, "Improvement of an Extrapolation Scheme for Strong-Coupling Expansion in Quantum Field Theory," Phys. Rev. Lett. **43**, 537 (1979).

3. C. M. Bender, F. Cooper, G. S. Guralnik, R. Roskies, and D. H. Sharp, "Effective Potential for a Renormalized d-Dimensional $g\phi^4$ Field Theory in the $g \rightarrow \infty$ Limit," Phys. Rev. Lett. **45**, 501 (1980).

4. C. M. Bender, F. Cooper, G. S. Guralnik, H. Rose, and D. H. Sharp, "Strong-Coupling Expansion for Classical Statistical Dynamics," J. Stat. Phys. **22**, 647 (1980).

5. C. M. Bender, F. Cooper, G. S. Guralnik, E. Mjolsness, H. A. Rose, and D. H. Sharp, "A Novel Approach to the Solution of Boundary-Layer Problems," Advances in App. Math. **1**, 22 (1980).

6. C. M. Bender and S. A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers* (McGraw-Hill, New York, 1978), Chapter 7 for a general discussion of the ideas of perturbation theory and Chapter 9 for an extensive introduction to boundary-layer theory.

7. J. D. Cole, *Perturbation Methods in Applied Mathematics* (Blaisdell, Waltham, Massachusetts, 1968).

8. A. H. Nayfeh, *Perturbation Methods* (Wiley, New York, 1973).

9. H. Schlichting, *Boundary-Layer Theory* (McGraw-Hill Book Co., New York, 1968) p. 129, for a discussion of the Blasius equation.

ACKNOWLEDGMENTS

The work presented here was done at Los Alamos in collaboration with Fred Cooper, Harvey A. Rose, and David H. Sharp of Los Alamos and Gerald Guralnik of Brown University.


RF
POWER
ON

the Neutrino in 1980

Neutrino experiments planned at the Los Alamos Meson Physics Facility and at the Nevada Test Site may hold the key to many advances in particle physics and cosmology.

by Thomas J. Bowles and Margaret L. Silbar

Fifty years ago scientists postulated the existence of the neutrino, a neutral, massless, spinning particle traveling at the speed of light, passing unimpeded through the whole earth and filling the universe in copious numbers. After numerous experiments and reformulations, the properties and behavior of the neutrino still challenge our theories about the forces and symmetries of nature.

The present puzzle concerns the results of a recent and controversial experiment, which indicate that neutrinos oscillate or spontaneously change their "flavor" as they travel through "empty" space—almost as if the apple falling from Newton's tree transformed itself into an orange before it struck him on the head. If this result is proved correct, it will mean that the neutrino, long assumed to be a massless particle, does have a nonzero rest mass. It will also mean that some of the conservation laws, which have been so successful in describing the phenomenology of weak interactions, will no longer apply to all physical processes.

Experiments are now being designed at Los Alamos and elsewhere to search further for evidence of a finite neutrino mass.

The Los Alamos Meson Physics Facility (LAMPF) provides the best testing ground for certain classes of ideas about neutrino oscillation, for it produces neu-

trino fluxes of uniquely high intensity and special composition and time structure. At the Nevada Test Site, the extremely high instantaneous neutrino fluxes available when nuclear devices are detonated provide other unequalled capabilities to test concepts of weak interactions.

The Laboratory has played a central role in the development of neutrino physics, for it was a Los Alamos team—Frederick Reines and the late Clyde Cowan, Jr.,—who first observed the neutrino in reactor experiments. This role will continue as experiments on neutrino oscillation and neutrino masses help test grand unified theories of fundamental particles. New data from these experiments may indicate the existence of totally new interactions in nature. They may even suggest that our expanding universe will come to a halt and finally contract.

Is the Neutrino Massless?

The possibility of a nonzero rest mass for the neutrino has been considered since 1930 when Wolfgang Pauli postulated its existence to save the laws of energy and momentum conservation in beta decay of radioactive nuclei. Originally beta decay was thought to be a two-body process in which a nucleus, such as RaE, changes to a unique final state, in this case RaF, and emits an electron (e^-):

$$\text{RaE } (^{210}_{83}\text{Bi}) \rightarrow \text{RaF } (^{210}_{84}\text{Po}) + e^- .$$

Conservation of energy and momentum implies that all electrons in such a two-body decay reaction come out with the same energy. Instead, the emitted electrons have a wide spectrum of energies, as shown in Fig. 1. Pauli suggested that the missing energy was carried away by a highly penetrating (and hence, undetected) particle with no charge and little or no rest mass. He also noted that, to conserve angular momentum and to preserve the "law of spin and statistics," this neutral particle must be a fermion, a particle with an intrinsic spin of $^1/_2\hbar$.

In 1934, two years after the discovery of the neutron, Enrico Fermi used the neutrino hypothesis to formulate a theory of beta decay. This theory correctly predicted the shape of the electron spectrum and provided the central ideas for what was to become the theory of all weak interactions.

The basic process in nuclear beta decay is the change of a neutron (n) into a proton (p) with the emission of an electron and an antineutrino ($\bar{\nu}$):

$$n \rightarrow p + e^- + \bar{\nu} .$$

Fermi postulated that this process takes place through the direct interaction of these four particles, which are all fermions. He assumed that the neutrino, like the other fermions, has a particle and an antiparticle form, but since the neutrino has no charge, and perhaps no magnetic moment, the distinction between neutrino and antineutrino was not then known.

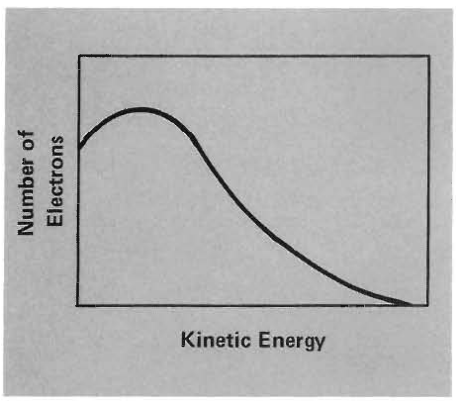


Fig. 1. Typical energy spectrum of electrons emitted in beta decay.

However, the mathematical form of the interaction dictated that fermion number (the number of fermions minus the number of antifermions) be conserved in the reaction. Since the neutron, the proton, and the electron are particles, and by convention are assigned a fermion number of +1, the neutral, undetected particle emitted in beta decay must be the antineutrino with a fermion number of −1. Then the total fermion number is +1 before and after the reaction. (As we will see, this and other number-conservation laws have played an important role in understanding and predicting weak-interaction processes.) But the question of whether or not the neutrino has a mass remained open.

That same year Hans Bethe and Rudolf Peierls pointed out that Fermi's theory allowed the neutrino to induce inverse beta decay:

$$\bar{\nu} + p \rightarrow n + e^+ .$$

This reaction provided a means for verifying the existence of the neutrino by observing the emitted neutron and positron. However, the probability for inverse beta decay is so low (the cross section is about 10^{-44} cm^2) that neutrino detection was not pursued until almost 20 years later when an intense source of antineutrinos became available from fission reactors.

In 1953, Frederick Reines and Clyde Cowan, Jr., reported results consistent

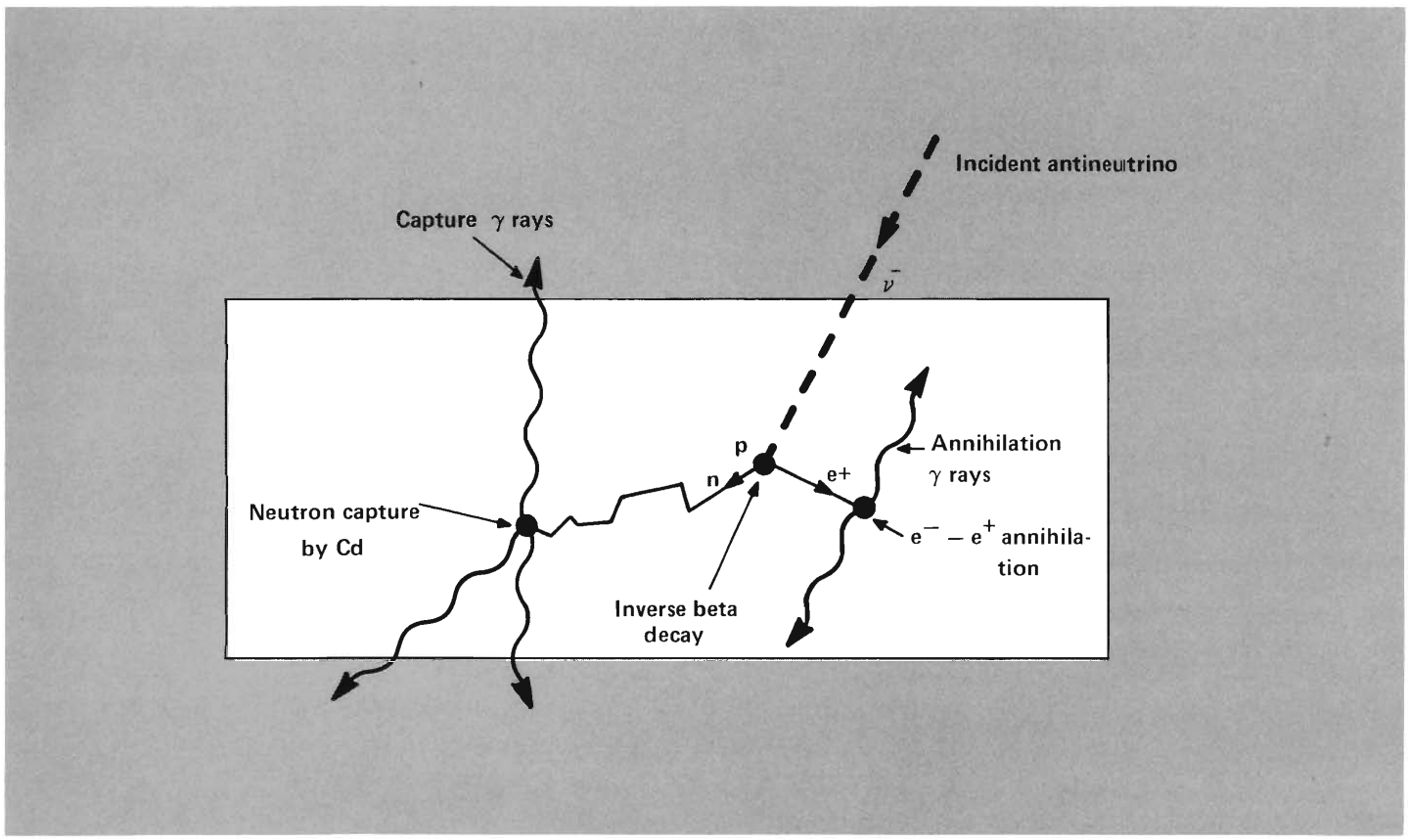


Fig. 2. Neutrinos can be detected by observing the events associated with inverse beta decay, the interaction of an antineutrino with a proton to produce a neutron and a positron. The positron annihilates with an electron to produce two gamma rays. The neutron eventually interacts with a nucleus with a high neutron-capture cross section, such as cadmium, to produce capture gamma rays. The positrons and/or the gamma rays are observed by their interactions with a detecting medium, such as a liquid scintillator.

with the occurrence of inverse beta decay in an experiment at the Hanford reactor. Incontrovertible proof of the neutrino's existence came three years later when Reines and Cowan, together with F. B. Harrison, Herald W. Kruse, and Austin D. McGuire, repeated the experiment at the Savannah River reactor. Figure 2 shows how neutrino-induced inverse beta decay can be observed.

By this time other weak processes were known, in particular the decay of the pion (π) into the muon (μ),

$$\pi \rightarrow \mu + \nu ,$$

and the decay of the muon,

$$\mu \rightarrow e + \nu + \bar{\nu} .$$

Although the neutrinos were not observed, their existence was assumed, this time to preserve lepton-number conservation, a special case of fermion-number conservation. A lepton is a fermion that does not participate in the strong interactions. Each lepton (e^-, μ^-, and ν) is assigned a lepton number of $+1$ and each antilepton (e^+, μ^+, and $\bar{\nu}$) a lepton number of -1. With these assigned numbers, the law of lepton-number conservation says that total lepton number remains constant in weak processes. This law, which emerges naturally from the assumed mathematical form of weak interactions, is consistent with all observed weak processes and explains the nonoccurrence of processes

that would violate it.

Then, in 1957, the startling discovery that parity is not conserved in beta decay led to a major breakthrough in our conception of the neutrino and of weak interactions in general. At the suggestion of T. D. Lee and C. N. Yang, C. S. Wu and her collaborators measured the direction of electrons emitted in the beta decay of polarized Co^{60}. They found that 40% more electrons were emitted along the Co^{60} spin axis than opposite it. This near-maximum violation of parity, or right-left symmetry, could be explained if the antineutrinos emitted in beta decay exist only in a longitudinally polarized form, that is, with spin vector pointing either along or opposite the direction of motion (right- or left-polarized, respectively), but not both.

The property of longitudinal polarization led to a new and very appealing theory of the neutrino. Unlike other known fermions, whose wave functions have four components corresponding to particle and antiparticle in both right- and left-polarized states, the neutrino has only two components: the neutrino is *always* left-polarized, or more precisely, left-handed, and the antineutrino is *always* right-handed.* The other two components (right-handed neutrino and left-handed antineutrino) are missing. Such a two-component theory in which neutrino and antineutrino are distinguished by their handedness implies that the neutrino travels at the speed of light and is therefore a massless particle (see Fig. 3).

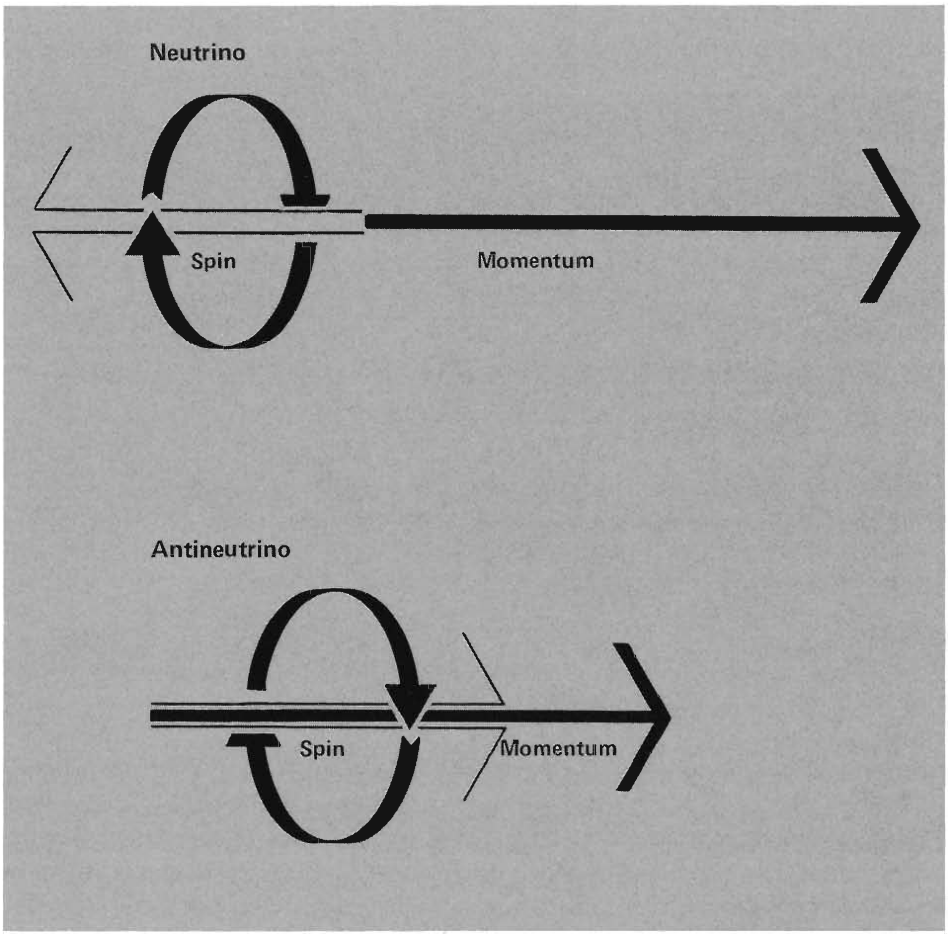


Fig. 3. In the standard two-component neutrino theory, the neutrino is a massless particle with intrinsic spin, or angular momentum, of $^1/_2$ ħ. The neutrino is always left-handed, that is, the direction of its spin vector is opposite that of its momentum; the antineutrino is always right-handed with the direction of its spin vector the same as that of its momentum. This distinction between particle and antiparticle is possible only if the neutrino is traveling at the speed of light (and is therefore massless). Otherwise, transformation to a reference frame moving faster than the neutrino would reverse the momentum and cause the neutrino to appear to be an antineutrino.

The massless two-component neutrino theory had two important consequences for weak interactions. First, since the neutrino and its antiparticle were not the same, $\nu \neq \bar{\nu}$, it was possible to retain the lepton number assignments of $+1$ and -1, respectively. Thus lepton-number conservation remained an exact law for weak interactions. Second, the left-handed character of the neutrino helped to establish a universal left-handed mathematical form for all weak-interaction processes, thus explaining why even massive particles emerging from such processes are partially polarized.

**Right- or left-handed is not exactly the same as right- or left-polarized except for massless particles, but for this discussion we will ignore the difference.*

We must emphasize that although the standard two-component theory implies that the neutrino is massless, it does not prove this as fact. It is possible to formulate a two-component theory for a massive neutrino consistent with the observed polarizations of participants in weak interactions. However, this alternative formulation (in which the two components correspond to the right- and left-handed neutrino) does not produce a lepton-number conservation law because the neutrino and the antineutrino are the same particle, that is, $\nu = \bar{\nu}$. The physics community favored the massless two-component theory. It was simple. It was aesthetically appealing. And, most important, it preserved lepton-number conservation.

This simple picture of the neutrino became more complex when, in 1963, experiments showed that neutrinos come in two "flavors." The quantum number flavor was postulated to explain the fact that antineutrinos from beta decay never produce positive muons. That is, if we call the antineutrino from beta decay $\bar{\nu}_e$, then

$$\bar{\nu}_e + p \rightarrow n + e^+$$

but

$$\bar{\nu}_e + p \not\rightarrow n + \mu^+ .$$

Thus weak interactions separate the leptons into two families, the muon family consisting of the muon, the muon neutrino, and their antiparticles, and the electron family consisting of the electron, the electron neutrino, and their antiparticles. Further these two families are not "mixed" by weak interactions: a member of the muon family cannot by itself change into a member of the electron family. For example, a negative muon cannot decay into an electron unless a muon neutrino is emitted in the decay. Thus the decay $\mu^- \rightarrow \nu_\mu + e^- + \bar{\nu}_e$ does occur, but the decay $\mu^- \rightarrow e^- + \gamma$ does not.

Lepton-number conservation was thus separated into muon-number conservation and electron-number conservation. Again, these conservation laws reflect the observed occurrence or nonoccurrence of various weak interactions.* Evidence for a third lepton family, the tau and the tau neutrino, has now produced a tau-number conservation law. The existence of three neutrino flavors did not, however, change the notion that each one was a massless, two-component particle.

That brings us to the present status of theories and observations about the neutrino. The hypotheses of a massless two-component neutrino and lepton-number conservation appear correct for all observed weak interactions and are incorporated into the Glashow-Weinberg-Salam model that unifies the electromagnetic and the weak forces.

Although all experiments to date support this simple picture of the neutrino, there is no fundamental reason why the neutrino should be massless. In fact, modern views of fundamental forces which incorporate the Glashow-Weinberg-Salam model into a larger model suggest that neutrinos have a small rest mass and that effects of this rest mass can be observed through the phenomenon of neutrino oscillation between members of different lepton families.

Experiments to search for reactions that violate muon-number conservation are discussed by Cy Hoffman and Minh Duong-Van in Los Alamos Science, 1, No. 1, 62-67 (1980).

Neutrino Oscillation— A Recurring Idea

The first suggestion that neutrinos might oscillate—periodically change from one form to another—came from Bruno Pontecorvo in 1957. He noted that if the massless two-component neutrino theory was wrong, that is, if neutrinos were massive and lepton-number conservation was violated, then the neutrino, like the neutral K meson, might oscillate between its particle and antiparticle forms. This possibility was not explored because at that time the physics community was just beginning to accept the massless two-component theory of the neutrino.

In 1963, as a result of the discovery of the muon neutrino, the idea of oscillation surfaced again—this time between the electron neutrino and the muon neutrino—and in 1969 gained some respectability as a possible explanation for the solar neutrino puzzle. (See Kolb's commentary, "Neutrinos in Cosmology and Astrophysics.") It nonetheless remained on the fringes of theoretical physics because it violated the empirically established conservation laws of muon number and electron number.

Only in the last few years has the possibility of neutrino oscillation been taken more seriously. Indeed, this phenomenon is a natural result of radically new schemes that attempt a unified description of all fundamental interactions. (See Goldman's commentary, "Neutrinos and Grand Unified Theories.") To achieve unification, these theories must postulate that at least some of

the empirical number-conservation laws are not exact. The theories suggest that, at very high energies (or, equivalently, at very small distances), neutrinos and the other leptons become indistinguishable from the quarks, which are the constituents of those particles (such as the proton and the pion) that interact through the strong force. In this framework neutrinos acquire a nonzero rest mass just as do all the other fundamental constituents of matter. Thus, the major philosophical objections to oscillation have been swept away.

In fact, the structure of unified theories suggests that neutrino oscillation does occur. The terms in these theories that could give the neutrino its mass are not likely to have the same symmetry as those describing weak interactions; these mass terms might mix members of the different lepton families and thereby produce oscillations between different neutrino flavors ($\nu_e \leftrightarrow \nu_\mu$) or they might produce oscillations between particle and antiparticle forms ($\nu_{e,\mu} \leftrightarrow \bar{\nu}_{e,\mu}$) as originally suggested by Pontecorvo. Further, if the neutrino is massive, it might actually be a four-component fermion whose two unseen components take part in "superweak" interactions that have yet to be observed. Thus the discovery of massive neutrinos through oscillation experiments or direct measurements would open up a Pandora's box of new possibilities for the elusive and still puzzling neutrino.

New Results and New Plans

We can now understand the excitement created by the announcement in 1980 of the controversial experiment mentioned at the beginning of this article. The experiment, performed at the Savannah River reactor by a group from the University of California at Irvine, provided evidence for neutrino oscillation, evidence that implies a nonzero rest mass for at least one neutrino type. Oscillation experiments determine not a value for the mass of a neutrino, but rather a value for $m_1^2 - m_2^2$, where m_1 and m_2 are the eigenmasses of the neutrino mass eigenstates defined by the lepton-mixing mass terms referred to above. This difference is said by the Irvine group to be about 1 $(eV)^2$.

The implied existence of massive neutrinos is supported by another recent experiment. A group at the Institute of Theoretical and Experimental Physics in Moscow measured the neutrino mass directly from a careful analysis of the spectrum of electrons emitted in the beta decay of tritium, $H^3 \rightarrow He^3 + e^- + \bar{\nu}_e$. The deviation between their experimentally determined spectrum and that expected for a massless neutrino indicates a value for the rest mass of the electron antineutrino of at least 14 eV, with a most probable value of ~35 eV. (The rest mass of the electron is 511 keV.) Although no one has yet found fault with this experiment, there is concern that the observed deviation may be due to interactions of the electrons as they leave the sample, which in this case was solid (tritium in an amino acid deposited on an aluminum backing).

To eliminate this uncertainty, a Michigan State University-Los Alamos group is planning a similar experiment at Los Alamos, in which the source consists of a cold (10 K), monatomic tritium beam. As the beam passes through a 1-m long decay region, the spectrum of electrons from beta decay of tritium will be measured with a magnetic spectrometer and analyzed for evidence of a nonzero neutrino mass. Recent work by solid-state physicists at the University of Amsterdam and at the Massachusetts Institute of Technology promises atomic beams in the near future of sufficient intensity to permit verification of the Moscow result.

Physicists at Los Alamos National Laboratory also plan to address the question of neutrino mass by searching for evidence of neutriino oscillation both at LAMPF and at the Nevada Test Site.

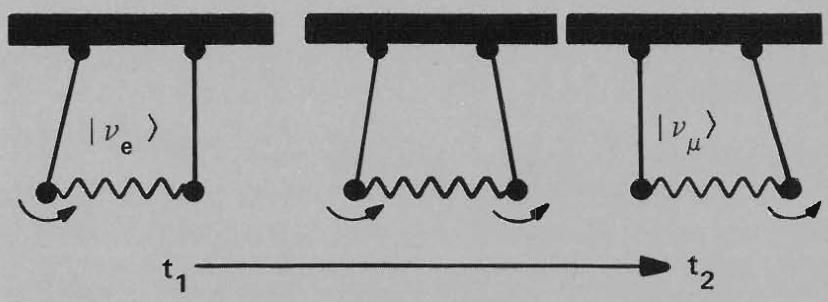
The Mechanism of Oscillation

Oscillation of one neutrino flavor into another and back again seems very mysterious because it conjures up the image of matter suddenly disappearing and just as suddenly reappearing. But this quantum-mechanical phenomenon becomes more intelligible when we realize that it is analogous in many ways to the classical phenomenon of energy transfer between the two (identical) bobs of a double pendulum system (see Fig. 4).

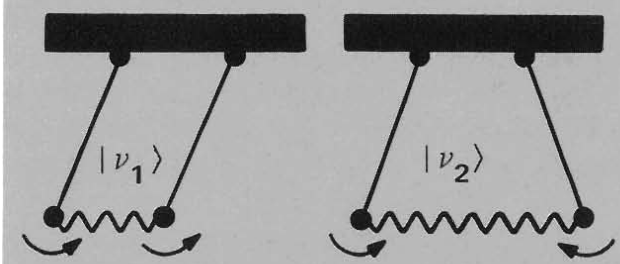
We know that if we start the pendulum by swinging only one bob, after some time the second bob will be swinging and the first will be stationary; still later the second bob will be stationary and the first bob will again be swinging. The change of an electron neutrino into a muon neutrino and back again is analogous to this periodic transfer of energy between the two bobs.

Why does this transfer occur? In the case of the double pendulum it occurs because the pendulums are coupled by a spring between the bobs. In the case of massive neutrinos it would occur if different neutrino flavors are coupled by

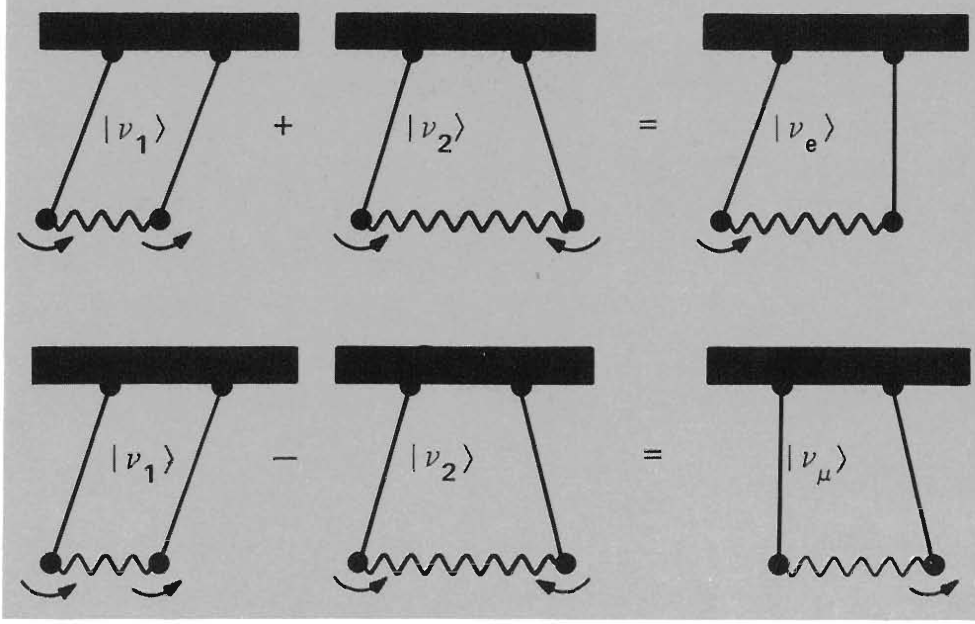
Fig. 4. An analogy between neutrino oscillation and periodic energy transfer in a coupled double pendulum.



At time t_1, the left pendulum swings alone. Coupling of the two pendulums by a spring between the bobs causes energy to be transferred until, at some time t_2, all the energy has been transferred to the right pendulum. Then the process reverses. This periodic energy transfer is analogous to the quantum-mechanical oscillation between an electron neutrino and a muon neutrino in which the probability of finding an electron neutrino gradually decreases from unity to a minimum (determined by the mixing angle θ) and then increases again to unity.



The coupled double pendulum has two stationary states, or normal modes. The mode in which the two pendulums swing in phase has a lower oscillation frequency than the mode in which the pendulums swing completely (180°) out of phase.



Addition and substraction of the normal modes (with appropriate normalizing factors) yield motion of the right pendulum and motion of the left pendulum, respectively.

In this normal mode description, energy transfers from the left to the right pendulum because the different oscillation frequencies of the normal modes cause the relative phase between them to change with time. Similarly the mass difference between the neutrino mass eigenstates causes the combination of mass eigenstates representing the electron neutrino to evolve with time into that representing the muon neutrino.

the proposed lepton-mixing mass terms.

The time it takes for energy to transfer from one pendulum bob to the other depends on the frequency difference between the double pendulum's two normal modes of oscillation; similarly, the time it takes for one neutrino flavor to oscillate into another depends on the difference between the squares of the two neutrino eigenmasses $(m_1^2 - m_2^2)$.

To understand this dependence we must continue our analogy on a mathematical level. Consider the possibility that the electron and muon neutrinos produced in weak processes acquire mass as a result of new interactions that mix members of the muon and electron families.* Then although these neutrinos do not have definite masses, their wave functions $|\nu_e\rangle$ and $|\nu_\mu\rangle$ can be described by linear combinations of two other neutrino wave functions $|\nu_1\rangle$ and $|\nu_2\rangle$ that do have definite eigenmasses m_1 and m_2, respectively.

$$|\nu_e\rangle = \cos\theta\,|\nu_1\rangle + \sin\theta\,|\nu_2\rangle \qquad (1)$$

and

$$|\nu_\mu\rangle = -\sin\theta\,|\nu_1\rangle + \cos\theta\,|\nu_2\rangle\,, \qquad (2)$$

where θ, the mixing angle, describes the extent to which ν_e and ν_μ are mixed by the new interaction. If m_1 does not equal m_2, oscillation between ν_e and ν_μ will occur. That is, as the state $|\nu_e\rangle$ produced at $t = 0$ propagates in time, the relative phase between its $|\nu_1\rangle$ and $|\nu_2\rangle$ components will change so that the state $|\nu_e\rangle$ will begin to disappear and the state $|\nu_\mu\rangle$ will begin to appear.

This is exactly analogous to the time evolution of the double pendulum. We can identify $|\nu_e\rangle$ and $|\nu_\mu\rangle$ with the states describing the swing of one or the other pendulum bob, and $|\nu_1\rangle$ and $|\nu_2\rangle$ with the normal modes, or stationary states, of the pendulum corresponding to the two bobs swinging exactly in phase and exactly out of phase, respectively. The swing of one bob or the other corresponds to addition or subtraction of equal parts of the two normal modes, respectively [Eqs. (1) and (2) with $\theta = \pi/4$]. But the two normal modes have different frequencies. (The mode in which the bobs swing in opposite directions has a higher frequency because the spring restricts the separating motion.) Consequently the relative phase between the two normal modes changes with time so that the additive combination describing the swing of one bob [Eq. (1)] eventually changes into the subtractive combination describing the swing of the second bob [Eq. (2)].

Analysis of the electron neutrino's time evolution is similar and is given in the accompanying note "Derivation of Neutrino Oscillation Length." Here we simply state the results. As an electron neutrino produced at $t = 0$ travels through space at almost the speed of light, c, it will periodically disappear and reappear over a characteristic length $L \cong cT$ where T is the period of neutrino oscillation. L is called the oscillation length. Thus the probability $P_{ee}(x)$ that an electron neutrino has not oscillated into another type of neutrino at a distance x from the source is given by

$$P_{ee}(x) = 1 - (\tfrac{1}{2}\sin^2 2\theta)\,[1 - \cos(2\pi x/L)]\,,$$

where θ, the mixing angle that appears in Eqs. (1) and (2), determines the amplitude of the oscillation. The oscillation length L in meters is given by

$$L = 2.5\,E_\nu/\delta m^2\,,$$

where E_ν is the neutrino energy in MeV and δm^2 is $m_1^2 - m_2^2$ in $(eV)^2$.

Designing Oscillation Experiments

The equation for $P_{ee}(x)$ says that oscillation effects are maximum at half-integer multiples of L. Therefore, to detect oscillation the distance between the neutrino source and the detector must be at least the same order of magnitude as L. However, δm^2 and hence L are unknown.

Theoretical considerations suggest that δm^2 may be very small and hence L may be very large. Consequently, to be sensitive to low values of δm^2, detectors should be placed at the farthest distance from the source consistent with obtaining a measurable signal. In addition, to minimize the oscillation length as much as possible, experiments should exploit sources of low-energy neutrinos. The most convincing experiments will be those that demonstrate, by changes in the source-detector distance, the variation with distance of the number of neutrinos of a particular type.

Neutrinos of a particular type are detected by observing the reactions they induce in a detecting medium. For example, if protons are the medium and electron antineutrinos from beta decay

In actuality, we may be dealing with the mixing of more than just two neutrino types. The formalism, however, remains the same. Ultimately, the question of which neutrinos mix with which no doubt depends on how many different neutrinos exist.

Derivation of Neutrino Oscillation Length

Some simple algebra can show how neutrino oscillation effects depend on the mass difference of the neutrino mass eigenstates. We express the quantum-mechanical wave function for an electron neutrino produced at t = 0 as a mixture of the mass eigenstates $|\nu_1\rangle$ and $|\nu_2\rangle$ with masses m_1 and m_2.

$$|\nu(0)\rangle = |\nu_e\rangle = \cos\theta\,|\nu_1\rangle + \sin\theta\,|\nu_2\rangle ,$$

where θ, the mixing angle, characterizes the extent of mixing of the mass eigenstates in the weak-interaction eigenstate. At a later time t, the wave function is

$$|\nu(t)\rangle = \cos\theta\,\exp(-iE_1t)\,|\nu_1\rangle + \sin\theta\,\exp(-iE_2t)|\nu_2\rangle ,$$

where E_1 and E_2 are the energies of $|\nu_1\rangle$ and $|\nu_2\rangle$. For relativistic neutrinos ($E_\nu \gg m$), we can approximate E_1 and E_2 by

$$E_k = (p^2 + m_k^2)^{1/2} \cong p + m_k^2/2p .$$

After substituting these energies, $|\nu(t)\rangle$ becomes

$$|\nu(t)\rangle = \exp[-it\,(p + m_1^2/2E_\nu)][\cos\theta\,|\nu_1\rangle + \sin\theta|\nu_2\rangle\,\exp(i\delta m^2t/2E_\nu)] ,$$

where $\delta m^2 = m_1^2 - m_2^2$ and $E_\nu \cong p$. Since these neutrinos are traveling almost at the speed of light, we can replace t by x/c, where x is the distance from the source of electron neutrinos. Then we calculate $P_{ee}(x)$, the probability of finding an electron neutrino at x.

$$\begin{aligned} P_{ee}(x) &= |\langle\nu(0)|\nu(t = x/c)\rangle|^2 \\ &= |\cos^2\theta + \sin^2\theta\,\exp(i\delta m^2x/2E_\nu c)|^2 \\ &= \cos^4\theta + \sin^4\theta + 2\cos^2\theta\sin^2\theta\cos(\delta m^2x/2E_\nu c) . \end{aligned}$$

Adding and substracting $2\cos^2\theta\sin^2\theta = \frac{1}{2}\sin^2 2\theta$ and setting $L = 2\pi cE_\nu/\delta m^2$, we obtain

$$P_{ee}(x) = 1 - (\tfrac{1}{2}\sin^2 2\theta)\,[1 - \cos(2\pi x/L)] .$$

Thus the probability oscillates with distance and the oscillation length L is given in meters by

$$L = 2.5\,E_\nu/\delta m^2 ,$$

where E_ν is in MeV and δm^2 is in $(eV)^2$ ■

are the source, a reaction to observe is inverse beta decay, $\bar{\nu}_e + p \rightarrow n + e^+$. The signature of this reaction is the delayed coincidence of photons from annihilation of the positron and from capture of the neutron. A decrease in the rate of this reaction from that predicted for massless neutrinos would be evidence for neutrino oscillation. (An experiment of this kind is known as a disappearance experiment.)

To determine such a change in an already low reaction rate is a tremendous challenge. The experiments require extensive shielding to minimize background events and sophisticated detectors to differentiate background from events of interest. Detectors are similar in concept to that used by Reines and Cowan, but they are more sensitive. Made up of multiple units monitored by modern electronics, these detectors obtain fine-grained information about the sources of the measured signals.

Despite the size of present detectors, the expected count rates are still quite low. For example, at LAMPF with a typical neutrino flux of 10^{10} neutrinos per second incident on a 50-ton detector (a not abnormally large size) located 100 m from the LAMPF beam stop, one can expect to observe an interaction of interest in the detector only once every three or four hours.

As difficult as these experiments may be, their implications for grand unified theories and cosmology provide a compelling motive to attempt them.

Oscillation Experiments So Far

The Irvine group, who have claimed positive results, used neutrinos from the Savannah River reactor to search for the

transition $\bar{\nu}_e \rightarrow$ anything. They measured the cross sections for both a "charged current" and a "neutral current" reaction* induced by an electron antineutrino on a deuteron:

$$\bar{\nu}_e + d \begin{cases} \nearrow n + n + e^+ \text{ (charged current reaction).} \\ \searrow n + p + \bar{\nu}_e \text{ (neutral current reaction).} \end{cases}$$

The detecting medium, located 11.2 m from the reactor core, consisted of 268 kg of heavy water in which were placed 10 ^{3}He-filled proportional counters for measuring coincident and single neutron counts. Liquid-scintillator anticoincidence counters surrounded the detector, which was enclosed in a lead and cadmium shield.

The cross section for the neutral current reaction is independent of the type of neutrino that initiates the reaction. The charged current reaction, in contrast, can be induced only by the electron antineutrino. If neutrino oscillation occurs, the count rate for two-neutron events will be reduced relative to the count rate for one-neutron events.

From their data, the Irvine group derived a "ratio of ratios" (the ratio of the experimental cross sections for the two reactions divided by the ratio of the theoretical cross sections). For massless neutrinos this ratio of ratios should be unity. The Irvine group's currently reported value of 0.38 ± 0.21 led them to claim that neutrino oscillation does occur.

The controversy associated with this experiment is due to the fact that its results are not confirmed by another experiment performed in Grenoble at the Laue-Langevin Institute's reactor. A group from the California Institute of Technology, the Nuclear Sciences Institute (Grenoble), and the Technical University of Munich also searched for the transition $\bar{\nu}_e \rightarrow$ anything by determining the rate of inverse beta decay. The detecting medium, located 8.7 m from the reactor core, consisted of 375 liters of proton-rich liquid scintillator for photon detection and neutron moderation and ^{3}He-filled proportional counters for neutron detection. Using a neutrino energy spectrum calculated by Davis and Vogel, the group found no deviation from the expected rate of inverse beta decay and therefore concluded that oscillation at the level seen by the Irvine group does not occur. For maximum mixing ($\theta = \pi/4$), the Grenoble experiment set* an upper limit on δm^2 of 0.16 $(eV)^2$. The Irvine group claims** $\delta m^2 = 1$ $(eV)^2$ and $\sin^2 2\theta = 0.5$ $(2\theta = \pi/4)$.

The largest uncertainty in the results of both these experiments lies in the energy spectra of neutrinos from the reactors, which for the original analyses were calculated theoretically. The Irvine group claims that its ratio of ratios is insensitive to both experimental and theoretical uncertainties, but both groups are now performing experiments to check calculations of the neutrino spectra. Preliminary data from a direct measurement by the Grenoble group support their calculated spectrum. This group has since moved its experiment to a reactor in Gösgen, Switzerland, where they plan to obtain data at source-detector distances of 35-80 m. The Irvine group, in turn, has begun direct spectrum measurements at the Savannah River reactor and is obtaining further oscillation data at distances of 11-38 m.

Other information on oscillation exists from an experiment performed at LAMPF by a group primarily from Yale University and Los Alamos. This experiment set an upper limit on δm^2 of 0.9 $(eV)^2$ for the transition $\bar{\nu}_\mu \rightarrow \bar{\nu}_e$ and 3.0 $(eV)^2$ for the transition $\nu_e \rightarrow$ anything.*

Figure 5 shows the limits on δm^2 set by these experiments.

Whether the evidence for neutrino oscillation is confirmed or eventually refuted, these experiments have stimulated tremendous interest in the scientific community and many experimental groups are planning even more sensitive experiments to determine neutrino masses and mixing angles.

Neutrinos at LAMPF

As a source of neutrinos, LAMPF offers two great advantages: neutrino fluxes much greater than those of any other accelerator facility, and low neutrino energies (20-53 MeV) particularly suited to oscillation experiments. It also offers a unique opportunity to study muon neutrino oscillations at low

Charged current or neutral current refers to the charge of the hypothetical intermediate vector boson that mediates the reaction.

*As reported (at the 90% confidence level) by F. Boehm et al. in "Neutino Oscillation Experiment at the ILL Reactor at Grenoble," presentation at the International Conference on Neutrino Physics and Astrophysics: Neutrino 80, Erice, Sicily, June 23-27, 1980 (to be published).

**As reported (at the 68% confidence level) by F. Reines et al. in "Evidence for Neutrino Instability," presentation at the Spring Meeting of the American Physical Society, Washington, D.C., April 28-May 1, 1980 and submitted to Physical Review Letters.

*Unless otherwise noted, all upper limits on δm^2 are at the 90% confidence level.

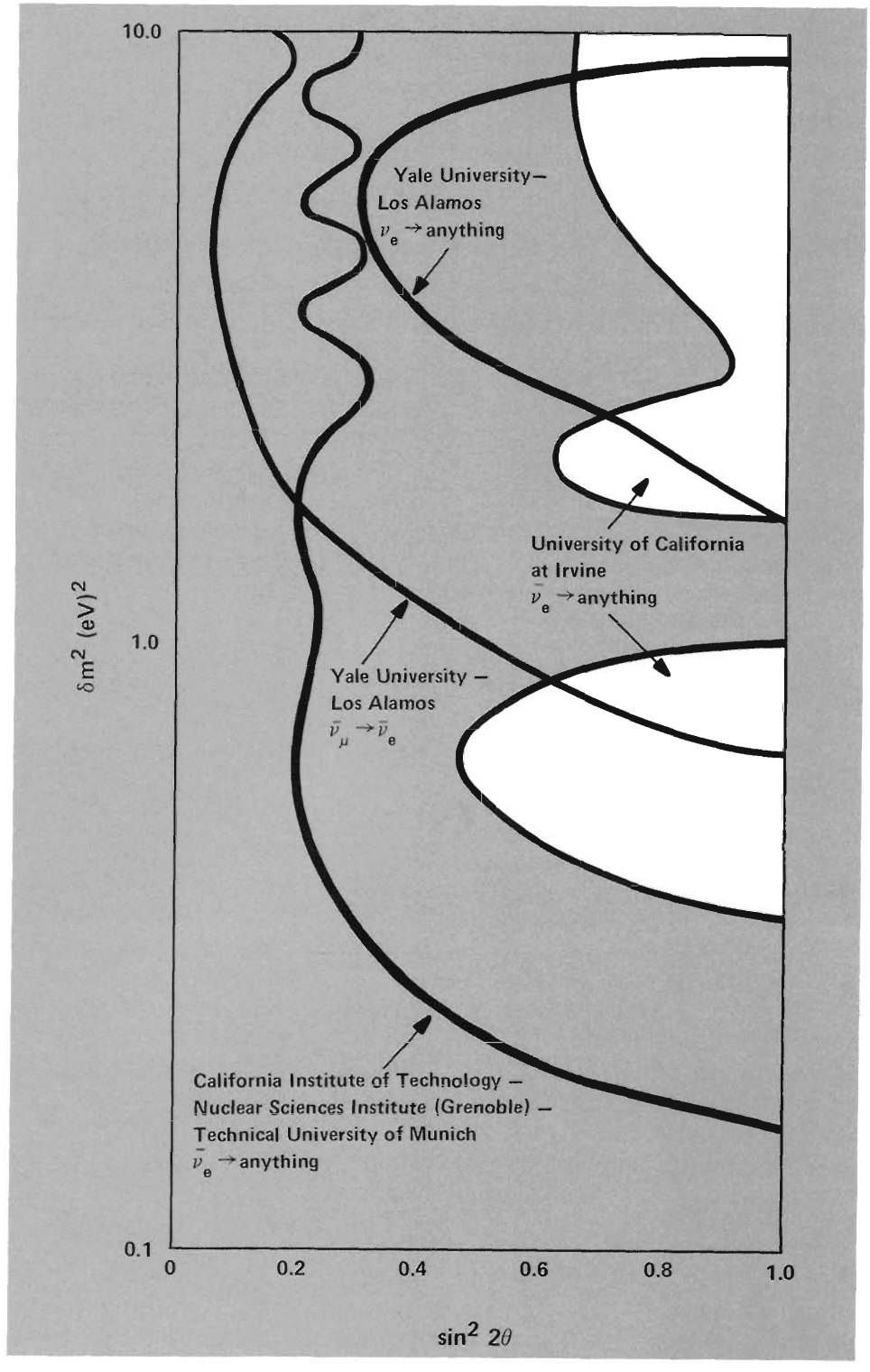
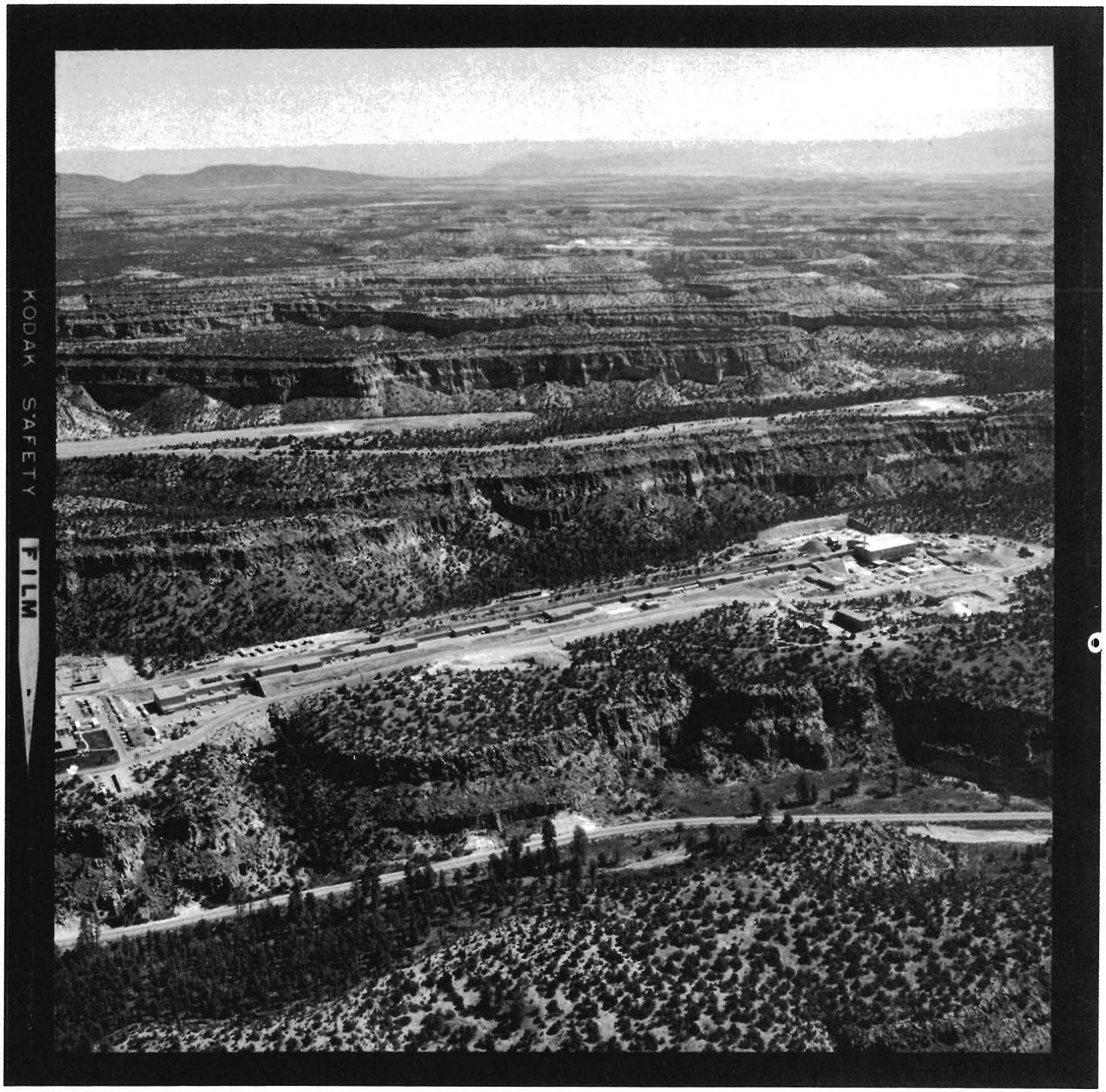


Fig. 5. Results of completed neutrino experiments appear to be contradictory. The Irvine group's positive results for δm^2 as a function of mixing angle θ (claimed at the 68% confidence level) are shown by the unshaded region. Negative results from other experiments established upper limit curves for δm^2 as a function of θ, shown here at the 90% confidence level. According to these latter experiments, possible values of δm^2 and $\sin^2\theta$ lie to the left and below these upper limit curves.



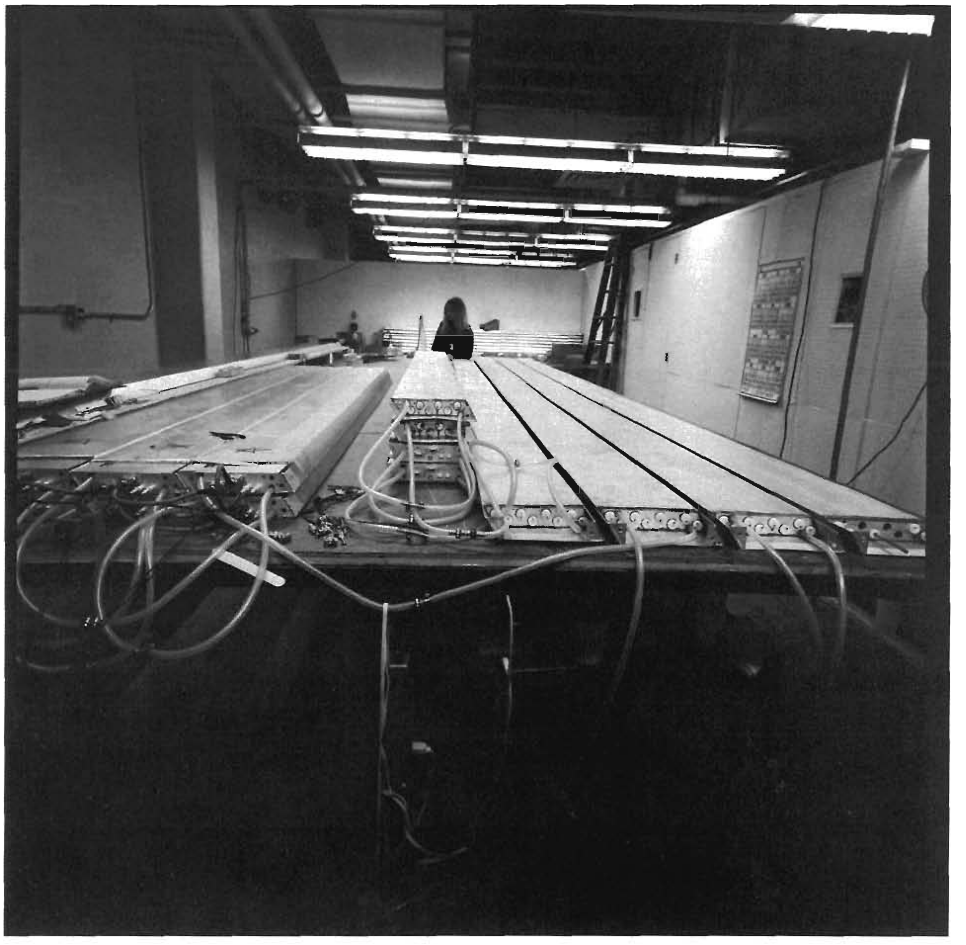
Aerial view of the Clinton P. Anderson Meson Physics Facility (LAMPF). The beam stop and neutrino facility, dwarfed by other experimental areas, are located at the far right.

energies.

Neutrinos at LAMPF are produced by the decay of pions and muons. Protons that remain after passing through the upstream targets are brought to rest in the beam stop at the end of the accelerator. In the process, they collide with atoms in the beam stop to produce the three charge states of the pion: π^+, π^0, and π^-. The neutral pions decay into photons and the negative pions are almost completely absorbed in nuclear reactions. The positive pions come to rest and decay by the reaction $\pi^+ \rightarrow \mu^+ + \nu_\mu$. The positive muon in turn decays by the reaction $\mu^+ \rightarrow e^+ + \nu_e + \bar{\nu}_\mu$. Note that these two decays yield no electron antineutrinos.

The two oscillation experiments scheduled to begin at LAMPF in early 1981 take advantage of this fact. Both are designed to search for the transition $\bar{\nu}_\mu \rightarrow \bar{\nu}_e$ by determining the rate of inverse beta decay. Occurrence of this reaction at a rate significantly higher than expected from background events will be clear evidence of neutrino oscillation. (Such experiments are called appearance experiments because electron antineutrinos appear in a beam that originally had none.)

One experiment, an Irvine-Los Alamos collaboration, is an outgrowth of a long-planned study of the elastic scattering of electron neutrinos by electrons.The heart of this experiment is a 14-ton "sandwich" detector to be located inside the neutrino facility, an "iron house" about 10 m from the beam stop. The detector, which contains 10 tons of plastic scintillator and 4 tons of polypropylene flash chambers, will measure the energy, position, and direction of motion of the recoil electrons



Marilyn Barlett is shown assembling several of the 600 drift tube modules to be used in the anticoincidence counter of the Irvine-Los Alamos neutrino oscillation experiment at LAMPF.

TABLE I

NEUTRINO OSCILLATION EXPERIMENTS

Experiment	E_ν (MeV)	Source-Detector Distance (m)	δm^2 for $\sin^2 2\theta = 1$[a] $(eV)^2$	Transition	Observed Reaction
University of California at Irvine SITE: Reactor, Savannah River Plant STATUS: Completed	1-8	11.2	~1[b]	$\bar{\nu}_e \rightarrow$ anything	$\bar{\nu}_e + d \rightarrow n + n + e^+$ $\bar{\nu}_e + d \rightarrow n + p + \bar{\nu}_e$
California Institute of Technology Nuclear Sciences Institute (Grenoble) Technical University of Munich SITE: Reactor, Laue-Langevin Institute STATUS: Completed	1-8	8.7	<0.16	$\bar{\nu}_e \rightarrow$ anything	$\bar{\nu}_e + p \rightarrow n + e^+$
Yale University Los Alamos National Laboratory SITE: LAMPF STATUS: Completed	20-53	9	<0.9 <3.0	$\bar{\nu}_\mu \rightarrow \bar{\nu}_e$ $\nu_e \rightarrow$ anything	$\bar{\nu}_e + p \rightarrow n + e^+$ $\nu_e + d \rightarrow p + p + e^-$
University of California at Irvine SITE: Reactor, Savannah River Plant STATUS: In progress	1-8	11-38[c]	<0.1	$\bar{\nu}_e \rightarrow$ anything	$\bar{\nu}_e + p \rightarrow n + e^+$
California Institute of Technology Nuclear Sciences Institute (Grenoble) Technical University of Munich SITE: Reactor, Gösgen, Switzerland STATUS: In progress	1-8	35-80[c]	<0.1	$\bar{\nu}_e \rightarrow$ anything	$\bar{\nu}_e + p \rightarrow n + e^+$
University of California at Irvine Los Alamos National Laboratory SITE: LAMPF STATUS: Approved for 1981	20-53	10	<0.4	$\bar{\nu}_\mu \rightarrow \bar{\nu}_e$	$\bar{\nu}_e + p \rightarrow n + e^+$
Los Alamos National Laboratory SITE: LAMPF STATUS: Approved for 1981	20-53	33[d]	<0.15	$\bar{\nu}_\mu \rightarrow \bar{\nu}_e$	$\bar{\nu}_e + p \rightarrow n + e^+$
Los Alamos National Laboratory SITE: Nevada Test Site STATUS: Approved for 1981-82	1.8-8	200-1200[e]	<0.0005, <0.005[e]	$\bar{\nu}_e \rightarrow$ anything	$\bar{\nu}_e + p \rightarrow n + e^+$
Rice University University of Houston Los Alamos National Laboratory SITE: LAMPF STATUS: Proposed	20-53	50-75[e]	<0.06, <0.12[f]	$\bar{\nu}_\mu \rightarrow \bar{\nu}_e$	$\bar{\nu}_e + p \rightarrow n + e^+$
Ohio State University Argonne National Laboratory Louisiana State University California Institute of Technology SITE: LAMPF STATUS: Proposed	20-53	25-60[e]	<0.2	$\nu_e \rightarrow$ anything	$\nu_e + d \rightarrow p + p + e^-$
Los Alamos National Laboratory SITE: LAMPF STATUS: Proposed	60-250	40-300[e]	<0.025 <0.2	$\nu_\mu \rightarrow \nu_e$ plus $\bar{\nu}_\mu \rightarrow \bar{\nu}_e$ $\nu_\mu \rightarrow$ anything plus $\bar{\nu}_\mu \rightarrow$ anything	$\nu_e + n \rightarrow p + e^-$ $\bar{\nu}_e + p \rightarrow n + e^+$ $\nu_\mu + n \rightarrow p + \mu^-$ $\bar{\nu}_\mu + p \rightarrow n + \mu^+$

[a] *Unless otherwise noted, all δm^2 values are at the 90% confidence level.*

[b] *The Irvine group claimed $\delta m^2 = 1\ (eV)^2$ for $\sin^2 2\theta = 0.5$ (at the 68% confidence level).*

[c] *Several source-detector distances in this range will be investigated.*

[d] *If positive results are observed at this distance, the detector will be moved.*

[e] *The smaller upper limit on δm^2 is that projected for use of several detectors on each of several weapon tests.*

[f] *The smaller and larger upper limits on δm^2 are those projected with and without neutron detection, respectively.*

from elastic neutrino-electron scattering. In the search for neutrino oscillation, the same quantities will be measured for the positrons from inverse beta decay reactions. The Irvine-Los Alamos group expects to place an upper limit on δm^2 of 0.4 $(eV)^2$, assuming maximum neutrino mixing.

The other scheduled experiment, a Los Alamos effort, is in some sense a preliminary experiment: its first objective is to test a 5-ton detector being assembled at LAMPF but planned for eventual use in neutrino oscillation experiments at the Nevada Test Site where tests of fission weapons provide an extremely brief but extremely intense pulse of neutrinos. The detector, containing 4470 liters (~5 ton) of gadolinium-loaded liquid scintillator and surrounded by an anticoincidence shield, will be located 33 m from the LAMPF beam stop in a hole drilled in tuff and will be covered with about 8 m of sand and iron for shielding. If cosmic-ray backgrounds are found to be sufficiently low, the group will search for the transition $\bar{\nu}_\mu \rightarrow \bar{\nu}_e$. Both the positrons and the neutrons from inverse beta decay will be detected. During a 110-day run with one detector (the possibility of more detectors is being considered), the group hopes to set an upper limit on δm^2 of 0.15 $(eV)^2$. If warranted by observation of positive results, the detector can, with modest effort, be relocated at other source-detector distances. Background information garnered by the group will also be used by later experiments.

The group also believes that it may be possible to search for the transition $\nu_e \rightarrow$ anything by determining the rate of the reaction $\nu_e + Pb \rightarrow (Bi^* \rightarrow Bi + n) + e^-$. If so, then their experiment would be simultaneously an appearance and a disappearance experiment (and could thus serve to refute or verify the Irvine group's results).

Neither of these experiments in their currently approved forms constitutes what has been called the definitive test: detection of the same transition at the same neutrino energy but at varying source-detector distances. However, when the 5-ton Los Alamos detector is moved to the Nevada Test Site, it will, again if warranted by observation of positive results, be positioned at various source-detector distances.

Three new proposals for experiments at LAMPF involve the search for oscillation effects as a function of distance. We will briefly discuss the Nevada Test Site experiment and three proposed LAMPF experiments.

Table I summarizes the status and basic parameters of all the neutrino oscillation experiments discussed.

Neutrinos at the Nevada Test Site

The Los Alamos experiment at the Nevada Test Site, a disappearance experiment like those performed at reactors, will be a search for a rate of the reaction $\bar{\nu}_e + p \rightarrow n + e^+$ that is lower than expected in the absence of oscillation. Here, sensitivity to small values of δm^2 will be greater than at reactors because of the larger source-detector distances (200-1200 m). Test of a fission weapon produces a short but very intense pulse of electron antineutrinos. Signal-to-noise ratios will therefore be high. On the other hand, ground motion from the blast creates huge problems and the total number of counts expected from each test is quite small. It may be necessary to use several detectors on each of several tests. The return for such extra effort, however, is the possibility of setting an upper limit on δm^2 of 0.0005-0.005 $(eV)^2$, values that may not be achieved in other ways.

The Next Generation Experiments

The high interest in oscillation phenomena and the need to explore all channels of neutrino mixing have focused continuing attention on LAMPF, where an intense source of muon neutrinos as well as electron neutrinos will enable experimentalists to search for oscillation with increased sensitivity. At this writing, three such proposals have been submitted to and reviewed by the Program Advisory Committee, and its recommendations are awaiting review by Louis Rosen, Director of LAMPF.

Two of these experiments are representative of neutrino experiments that can be done at the LAMPF beam stop. One proposal is by a group from Rice University, the University of Houston, and Los Alamos; the other is by a group from Ohio State University, Argonne National Laboratory, Louisiana State University, and the California Institute of Technology. In both experiments, detectors will be placed at distances from the source that are variable and greater than those to be investigated by the Irvine-Los Alamos group. Variable distance helps to eliminate uncertainties due to beam-associated backgrounds and neutrino flux magnitudes.

The Rice-Houston-Los Alamos group proposes an appearance experiment.

They will search for the transition $\bar{\nu}_\mu \rightarrow \bar{\nu}_e$ by determining if the rate of inverse beta decay, $\bar{\nu}_e + p \rightarrow n + e^+$, is greater than expected from background events. For positron tracking and energy measurement, the group will use a 40-ton array of liquid scintillator modules and drift chambers. Detector-source distances will be in the range 50-75 m with the detector placed in a deep tunnel for cosmic-ray shielding. The projected upper limit on δm^2 is 0.12 $(eV)^2$ if only positrons are detected, and 0.06 $(eV)^2$ if, as is hoped, both neutrons and positrons are detected simultaneously.

In addition, this experiment may provide information about the muon neutrino lifetime, which is of relevance to detecting the neutrino background assumed to exist as a remnant of the Big Bang. If a muon neutrino decays to a lower-mass neutrino plus a monoenergetic photon, the decay could be detected by observing the electron-positron pairs produced by the photon. The group claims they can improve the limit on the muon neutrino lifetime from its present value, $(2.6 \times 10^4)m_{\nu_\mu}$ s/MeV, to $(2.7 \times 10^8)m_{\nu_\mu}$ s/MeV, where m_{ν_μ} is the muon neutrino mass.

The Rice-Houston-Los Alamos group may also perform a disappearance experiment. Replacement of liquid scintillator with heavy water would permit them to observe the reaction $\nu_e + d \rightarrow p + p + e^-$ and thus to search for the transition $\nu_e \rightarrow$ anything.

The Ohio State-Argonne-Louisiana-Cal Tech group proposes both disappearance and appearance experiments. The former will consist of a search for the transition $\nu_e \rightarrow$ anything by determining the rate of the reaction $\nu_e + d \rightarrow p + p + e^-$. Properties of the emitted electrons will be measured with a detector consisting of drift chambers in heavy water. With source-detector distances of 25-60 m, the group expects to set an upper limit on δm^2 of 0.2 $(eV)^2$. Substitution of ordinary water in the detector will permit the group to perform an appearance experiment by searching for the transition $\bar{\nu}_\mu \rightarrow \bar{\nu}_e$ by determining the rate of inverse beta decay. In addition, if the group's proposal to build a more energetic muon neutrino source at the LAMPF beam dump is approved, the group will search for disappearance of muon neutrinos by determining the rate of the reaction $\nu_\mu + d \rightarrow \mu^- + p + p$.

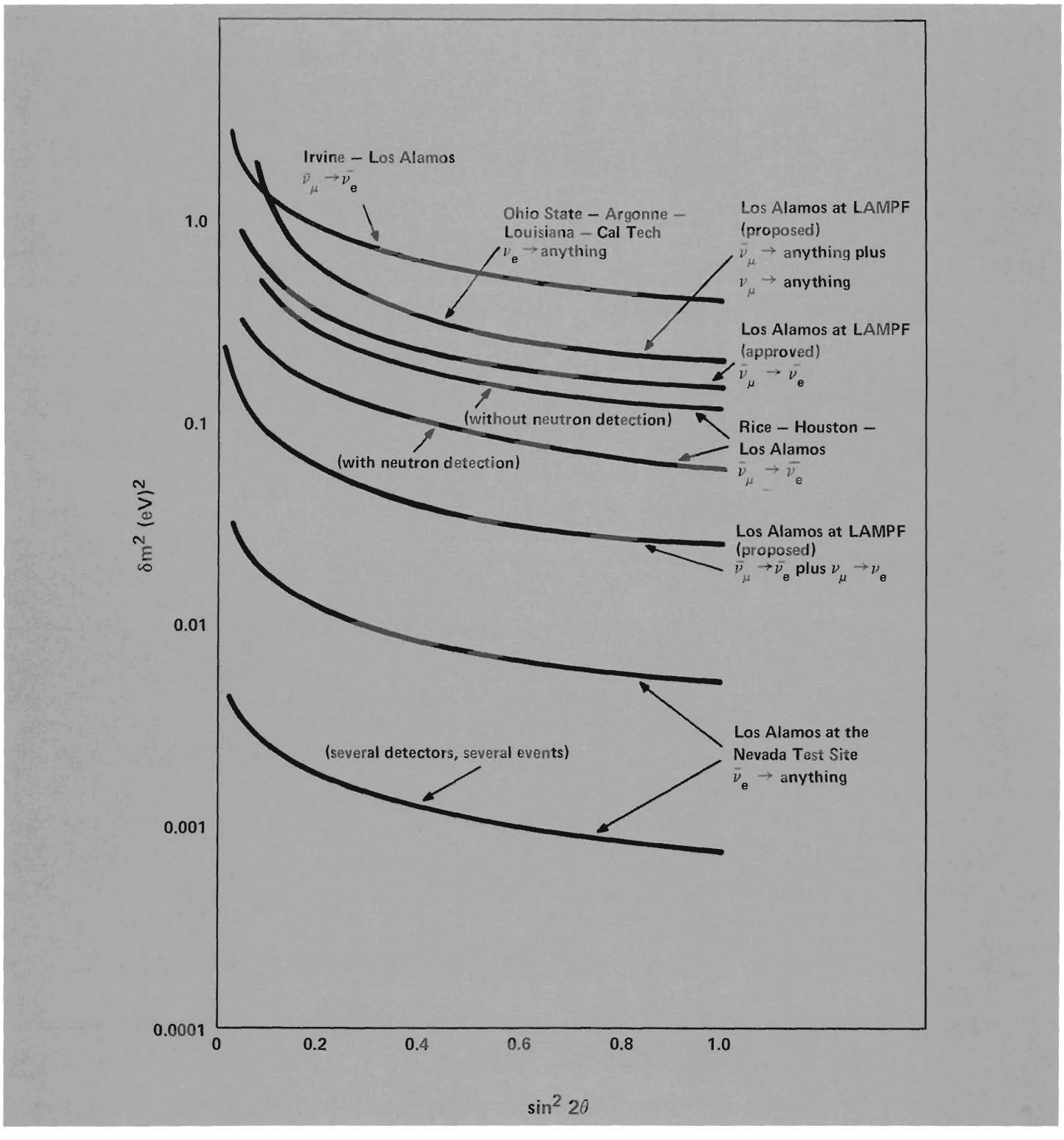
The third proposed experiment was conceived by a Los Alamos group and calls for the construction of a new beam line at LAMPF. Here pions, created by collision of part of the proton beam with a thin target, will decay in flight in a decay region to produce muon neutrinos and antineutrinos with average energies of 150 MeV. At such energies the reactions $\bar{\nu}_\mu + p \rightarrow n + \mu^+$ and $\nu_\mu + n \rightarrow p + \mu^-$ are possible. Observation of these reactions at rates lower than expected will indicate the occurrence of the transitions $\bar{\nu}_\mu, \nu_\mu \rightarrow$ anything. The group will also search for the transitions $\bar{\nu}_\mu \rightarrow \bar{\nu}_e$ and $\nu_\mu \rightarrow \nu_e$ by determining the rates of the inverse beta decay reactions $\nu_e + n \rightarrow p + e^-$ and $\bar{\nu}_e + p \rightarrow n + e^+$. The proposed 50-ton detector, to consist of liquid scintillator and drift chambers, will be positioned at source-detector distances of 40-300 m.

The projected limits on δm^2 for the various proposed experiments are shown in Fig. 6.

Conclusion

The neutrino has been one of the most successful inventions of theoretical physics. In the 50 years since its existence was first postulated, it has profoundly altered our theories of weak interactions and of astrophysical phenomena. Now we realize that if this particle has even a very small mass, it can influence our understanding of *all* fundamental interactions and the dynamics of the cosmos as well. The new experiments to measure neutrino masses and mixing angles will just begin to explore the many possibilities created by the existence of massive neutrinos ■

Fig. 6. The approved and proposed experiments at LAMPF and at the Nevada Test Site may not observe neutrino oscillation effects. Even so, they will establish upper limits on the neutrino mass difference δm^2 as functions of the mixing angle θ. The curves here are the upper limits (at the 90% confidence level) projected for the various experiments.



Thomas J. Bowles,
Los Alamos National Laboratory

Robert L. Burman,
Los Alamos National Laboratory

Herbert H. Chen,
University of California at Irvine

Scientists Working on Current Neutrino Experiments at Los Alamos.

Experimental Collaborations

University of California at Irvine: E. Pasierb, F. Reines (spokesman), and H. W. Sobel

California Institute of Technology: F. Boehm (spokesman), A. A. Hahn, H. E. Henrikson, H. Kwon, and J. L. Vuilleumier
Nuclear Sciences Institute (Grenoble): J. F. Cavaignac, D. H. Koang, and B. Vignon
Technical University of Munich: F. v. Feilitzsch, R. L. Mössbauer, and V. Zacek

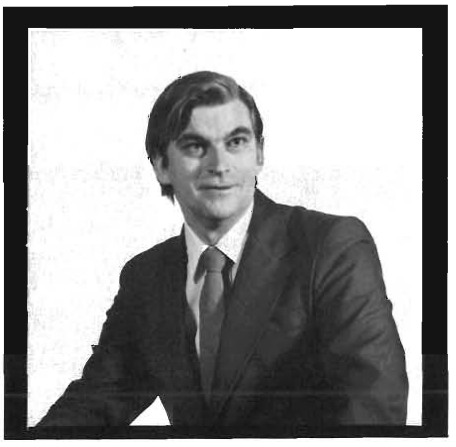
Yale University: V. W. Hughes, P. Nemethy (spokesman), and S. E. Willis
Los Alamos National Laboratory: R. L. Burman, D. R. F. Cochran, J. Frank, and R. P. Redwine
Saclay Nuclear Research Center: J. Duclos
National Research Council of Canada: C. K. Hargrove
Swiss Institute for Nuclear Research: H. Kaspar
University of Berne: U. Moser

University of California at Irvine: R. C. Allen, G. A. Brooks, H. H. Chen (co-spokesman), P. J. Doe, D. Hamilton, F. Reines (co-spokesman), A. N. Rushton, and K. C. Wang
Los Alamos National Laboratory: T. J. Bowles, R. L. Burman, R. Carlini, D. R. F. Cochran, T. Dombeck,* M. Duong-Van, J. Frank, V. Sandberg, and R. Talaga

Los Alamos National Laboratory at LAMPF (approved): D. Bartram, A. Brown, H. W. Kruse (spokesman), J. Mack, C. Smith, and J. Toevs

Los Alamos National Laboratory at the Nevada Test Site: D. Bartram, A. Brown, H. W. Kruse (spokesman), R. Loncoski, J. Mack, C. Smith, and J. Toevs

Rice University: G. C. Phillips (co-spokesman), H. Miettinen, G. S. Mutchler, and J. B. Roberts
University of Houston: A. D. Hancock, B. W. Mayes, and L. S. Pinsky
Los Alamos National Laboratory: J. C. Allred, A. A. Browman, R. L. Burman, D. R. F. Cochran, J. B. Donahue, M. Duong-Van (co-spokesman), M. V. Hynes, and B. W. Noel



Thomas W. Dombeck, Visiting Staff Member, Los Alamos National Laboratory

Minh Duong-Van, Los Alamos National Laboratory

Herald W. Kruse, Los Alamos National Laboratory

Ohio State University: T. Y. Ling (co-spokesman) and T. A. Romanowski (co-spokesman)
Argonne National Laboratory: L. G. Hyman and B. Musgrave
Louisiana State University: R. Imlay and W. J. Metcalf
California Institute of Technology: R. B. McKeown

Los Alamos National Laboratory at LAMPF (proposed): T. J. Bowles, R. L. Burman, and T. Dombeck* (spokesman)

Michigan State University: A. Ledebuhr and R. G. H. Robertson (co-spokesman)
Los Alamos National Laboratory: T. J. Bowles (co-spokesman), J. Browne, R. Hardekopf, and R. Mills

**On leave from the University of Maryland.*

Further Reading

"The Early Days of Experimental Neutrino Physics," F. Reines, Science, **203**, No. 4375, 11-16 (1979).

"Lepton Mixing and Neutrino Oscillations," S. M. Bilenky and B. Pontecorvo, Physics Reports, **41**, No. 4, 225-261 (1978).

"Is The Proton Stable?" M. Goldhaber, P. Langacker, and R. Slansky, Science, **210**, No. 4472, 851-860 (1980).

"Cosmology Confronts Particle Physics," G. Steigman, Annual Review of Nuclear and Particle Science, **29** 313-337 (1979).

Neutrinos in Cosmology and Astrophysics

by Edward W. Kolb, Jr.

It has been over 15 years since Penzias and Wilson discovered the 3 K microwave background radiation. Interpretation of this background as a remnant of the hot Big Bang is the cornerstone of modern theories of the beginning, the present, and the future evolution of the large-scale structure of the universe. Despite the appearance of the clear night sky as viewed from the mountains of northern New Mexico, most of the photons in the universe do not originate in stars, but are present in the invisible 3 K background. Fifteen years of observation have confirmed the thermal nature of the background spectrum. A universe at a temperature of 3 K has about 400 photons per cubic centimeter, or about 10^{88} photons in the visible universe. By comparison, neutrons and protons, which number only about 10^{80}, are but a small contaminant in a vast sea of photons, (Luckily, the nucleons, unlike the photons, are not uniformly distributed.) By observing the background photons, we directly probe the state of the universe when the photons were last scattered, which occurred when the universe was at a temperature of 10^{5} K, or about 10^{6} years after the Big Bang.

In addition to the background photons, there should also be a sea of neutrinos left over from the Big Bang. There should be about as many neutrinos as photons, about 10^{88}. Confirmation of this neutrino background would be in some sense even more fundamental than the discovery of Penzias and Wilson, since the background neutrinos last scattered when the temperature of the universe was 10^{10} K, or about one

second after the Big Bang. Thus, the background neutrinos are a much older relic of the origin of the universe than are the background photons.

Although direct detection of the background neutrinos seems remote, they may nevertheless play a crucial role in cosmology, and may even dominate the mass of the universe. Since the average energy of a nucleon (rest mass energy of 10^9 eV) is about 10^{13} times larger than the average energy of a background photon (10^{-4} eV), the nucleons contribute a factor of 10^5 more mass-energy to the universe than do the more numerous photons. However, if there exists a neutrino with mass greater than about 10 eV, the larger mass of the nucleon would be compensated by the sheer number of neutrinos, and the neutrinos would provide the bulk of the energy density. Although "weighing" the universe is not an easy task (where does one put the scale?), there are observational limits on its total mass. These observations require any neutrinos that survived the Big Bang and contribute to the present energy density of the universe to have a mass less than about 50 eV. If there are neutrinos with mass greater than about 20 eV, they would provide enough mass to gravitationally bind the universe, halt its expansion, and eventually cause it to recontract. It is fascinating to imagine the minuscule neutrino determining the fate of the universe.

A neutrino with a mass of a few eV may also account for the "dark mass" in galaxies. Astronomers can determine the mass of a galaxy by measuring its gravitational potential or by measuring the emitted light. However, these two methods disagree. Galaxies should be much brighter if all their mass is contained in stars. This so-called missing mass problem can be viewed as a missing light problem. Neutrinos with mass of a few eV would be expected to cluster gravitationally with the galaxies and could thus account for the dark mass.

Neutrinos may be important in astrophysics not only because of their mass, but also because of the possibility of neutrino decay, If one species of neutrino has a mass, all species of neutrinos would be expected to be massive. Most theories predict that the heaviest neutrino would be unstable. One possible decay mode would be the decay of a heavy neutrino to a light neutrino and a photon. Although the lifetime of the neutrino would be expected to be extraordinarily long, if it is less than about 10^{18} years, it may be possible to detect the ultraviolet radiation from the decay of 10-eV background neutrinos. There have even been exotic suggestions that the decay of short-lived ($\sim 10^{-14}$ s), heavy (-1 MeV) neutrinos can provide enough energy to power supernovae.

Finally, neutrino oscillation could solve the puzzle of the missing solar neutrinos. We know that a large number of electron neutrinos are produced by the reactions that power the sun. However, experiments measure a much lower solar neutrino intensity than predicted by theory. But if the electron neutrinos oscillate, they could arrive at the earth as muon or tau neutrinos, escape detection, and account for the missing neutrinos.

Because massive neutrinos offer such attractive solutions to a wide variety of astrophysical and cosmological problems, it has been suggested that massive neutrinos *must* exist. Even if one does not subscribe to such an extreme viewpoint, it is easy to understand the importance of determining the properties of the neutrino ■

by Terrence J. Goldman

Neutrinos and Grand Unified Theories

SCIENCE IDEAS

Today many particle theorists believe that three of the known fundamental interactions—the strong, the electromagnetic, and the weak—are but different facets of one "grand unified" theory of interactions. How close we have come to this ancient goal of unification is demonstrated by the fact that there are now only several candidates for such a theory. Which, if any, is correct must still be decided by experiment.

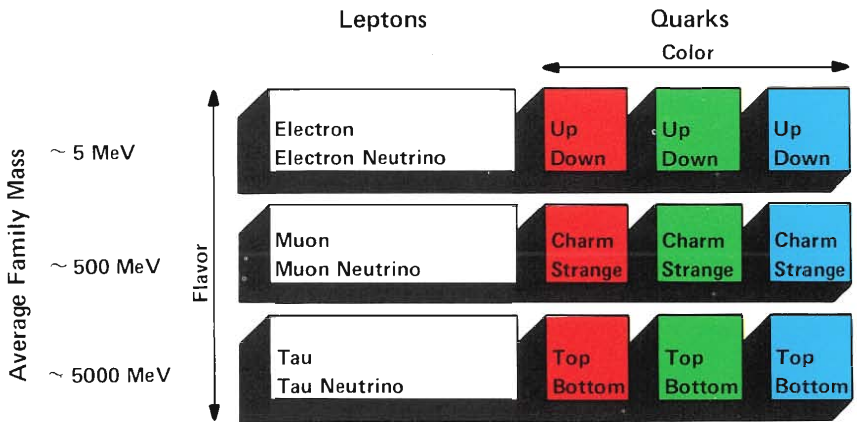
All the proposed grand unified theories build on the observation that the quarks (the components of those particles, such as the neutron and the pion, that interact through the strong force) and the leptons (the particles, such as the electron and the muon neutrino, that do not interact through the strong force) can be arranged by increasing mass into three families. The electron and its neutrino, the muon and its neutrino, and the tau and its neutrino are paralleled by the up and down quarks, the strange and charmed quarks, and the top and bottom quarks. In grand unified theories this parallel structure, or symmetry, between quarks and leptons is given a dynamic origin by postulating a new interaction that can change quarks into leptons and vice versa.

According to these new theories, our division of fundamental forces into strong, electromagnetic, and weak is merely a consequence of the energies at which we have observed particle interactions. At much higher energies, beyond any we can hope to achieve in the laboratory, the distinctions disappear: all three forces become components of a single force, and leptons and quarks become indistinguishable from each other. This symmetry between leptons and quarks is broken at energies accessible in the laboratory, energies at which quarks participate in strong interactions but leptons do not.

The grand unification of interactions and also the mechanism for breaking the symmetry are expressed mathemetically in non-Abelian gauge field theories. These theories are a generalization of the familiar Abelian gauge theory for electromagnetic interactions known as quantum electrodynamics.

As in quantum electrodynamics, particles interact through the exchange of quantum excitations of the gauge fields. In quantum electrodynamics, the exchange particle is the photon. In grand unified theories, the exchange particles are called gauge vector bosons. All these vector bosons would be massless, like the photon, were it not for the fact that the gauge symmetry is partially broken by the effects of Higgs bosons. These massive scalar particles are introduced into the theory to provide a background field that breaks the symmetry just as the magnetic field breaks the rotational symmetry in a ferromagnetic domain. Higgs boson effects are also responsible for nonzero quark and lepton masses.

As a result of the symmetry breaking, the leptoquark gauge bosons that turn quarks into leptons acquire extremely large masses. Theoretical predictions for the leptoquark mass range from 10^5 GeV in Pati-Salam type models to 10^{15} GeV for models that have



Quarks and leptons are now considered to be the fundamental constituents of matter. They are here arranged horizontally according to color quantum number and vertically according to flavor quantum number. Quarks come in three colors—red, green, and blue. Each color represents a different strong-interaction "charge." Leptons do not participate in strong interactions and thus have no color. Quarks and leptons are each assigned one of six flavor quantum numbers. The six flavors have been grouped into three pairs corresponding to three families of quarks and three families of leptons. The families have then been arranged by increasing average rest mass. This arrangement suggests the existence of three larger families, each one made up of quarks and leptons. Our everyday world is composed entirely of the lightest of these three quark-lepton families. The proton, for example, consists of two up quarks and one down quark; the neutron consists of two down quarks and one up quark. The other two families appear to be more massive imitations of the first family.

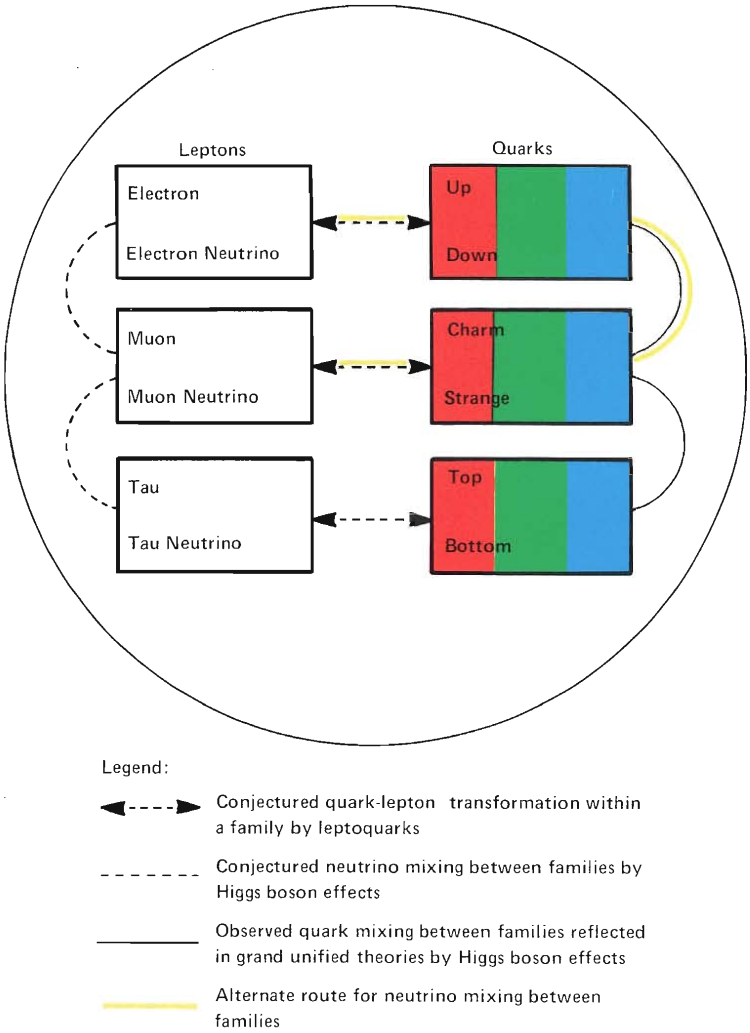
SU(5), 0(10), or E_6 symmetry. At energies greater than the leptoquark mass, the distinction between quarks and leptons vanishes and the symmetry of grand unified theories becomes apparent. Such energies cannot, however, be achieved in foreseeable experiments. The required accelerators based on present technology would be larger than the solar system! Therefore, we must find indirect tests of the symmetries and new interactions predicted by grand unified theories.

Almost all of these theories offer at least one such test: they predict that protons undergo radioactive decay and hence that no element is perfectly stable. (The exceptions have been contrived specifically to avoid this.) Proton decay was previously believed to be impossible; it was prohibited by negative observations that had been elevated to a principle called baryon-number conservation. However, the large value of the leptoquark mass severely suppresses the rate of proton decay and makes grand unified theories consistent with the observed stability of the proton and the apparent conservation of baryon number. At the same time, grand unified theories make fairly precise predictions for the proton decay rate. Experimental searches for this decay, sensitive at the appropriate level, are now underway.

Most of these theories also have as a consequence another previously dismissed possibility—that neutrinos have a nonzero rest mass. Again one can contrive theories that avoid this, but the natural structures that appear produce neutrino masses in a fashion similar to that by which masses are produced for quarks and charged leptons. There is, moreover, simply no reason for neutrinos to be massless, and so the converse is to be expected. Furthermore, since the quarks (down, strange, and bottom) are

SCIENCE IDEAS

Grand unified theories incorporate the observed family structure of fundamental particles in a natural way. Each family is a separate realization of a single representation of the unifying symmetry group. These theories not only organize the particles into families, but also predict new interactions between quarks and leptons within a family. Leptoquarks (massive vector bosons) change leptons into quarks and vice versa. At very high energies, leptons and quarks become indistinguishable. In addition, Higgs bosons effect mixing across family lines. Such mixing has been observed between strange and down quarks and between bottom and strange quarks. In the context of grand unification, similar mixing between lepton families seems likely. Neutrino oscillation would be evidence for such mixing. (An indirect route for neutrino mixing is indicated by the yellow lines.)



known to be mixed by mass effects, it is likely that the different neutrino types undergo a similar mass mixing. The existence of leptoquarks and the known quark mixing provide a route for lepton mixing (albeit by a very small amount).

Explicit calculations for specific models yield neutrino mass values no larger than a few tens of electron volts for the heaviest (tau) neutrino. This is close to the present experimental limit for those neutrinos that accompany electrons in beta decay (but these need not be pure electron antineutrinos if mixing occurs) and very close to the value suggested by astrophysical arguments. The results of a recent experiment on beta decay of tritium have been interpreted as evidence for a neutrino mass between 15 and 45 eV.

Exciting as is the prospect of determining another property of the neutrino, the implications for grand unification add even more impetus to the search. The pattern of mass values and the values of mixing parameters will give vital clues about details of the Higgs bosons in a correct grand unified theory. Although much of the structure of this theory is fixed by the gauge symmetry principle, a number of Higgs bosons are introduced into the theory to provide quarks and leptons with their known masses. There is even one set of Higgs bosons that can only be observed indirectly by determining the nature of the neutrino masses it generates.

Thus, while the enormous energy scales involved most probably rule out direct experimental tests of grand unified theories, low-energy measurements of neutrino masses and mixings in oscillation experiments will provide a window on the most basic laws of our universe ■

The New Los Alamos Center for Nonlinear Studies

By David K. Campbell and Basil Nichols

Linear analysis as a mathematical discipline began in the nineteenth century and in the intervening years has achieved many spectacular successes throughout science. But in fact, most problems posed by nature are nonlinear, and the linear approximations we use to describe them are often a tacit admission that the underlying equations cannot be solved.

Recently, however, a confluence of ideas from diverse branches of science has led to solutions of problems once thought to be intractable. These exciting developments have kindled interest in studying the generic features of nonlinear phenomena.

Of course, when we speak of a given physical system, we must realize that it can behave sometimes in a linear and sometimes in a nonlinear manner. The ripples that distort the surface of a flag as it waves in a gentle, steady breeze are smooth, predictable motions that can be described mathematically by a linear model. When the flag convulses in response to strong, gusty winds, the sharp, intermittent snaps are not uniform; the flag's energy is released in spurts as it adjusts its shape violently to the changes in its environment. The flag has changed from a linear to a nonlinear system at the bidding of the wind.

Similar transitions occur when a steady stream breaks up into turbulent vortices, or when magnetically confined plasmas break up and splatter on the containment walls. These examples suggest the tremendous qualitative differences between the two types of behavior. Linear systems are typically smooth, regular, and predictable, whereas nonlinear systems are often characterized by sharp and unstable boundaries, erratic or chaotic behavior, and dramatic responses to very small changes. Such properties make quantitative description a very difficult task.

Mathematically, what makes linear problems easier to solve than nonlinear ones? Perhaps the most crucial property is that two solutions of a linear equation may be added together to give a new solution to the equation. As a consequence, there exist established methods for solving completely any linear system, even very complicated ones. The methods amount to breaking up the complicated system into many simple pieces and, from solutions to the pieces, "patching together" a solution for the full system. In contrast, two solutions to a nonlinear system do not add together to give a new solution; hence, patching together simple pieces to understand the whole simply does not work. In a sense, nonlinear systems must be treated *in toto* and in their full complexity, and so it is not surprising that there exists no general method for solving them. And yet some methods must be found because for many systems the point at which they enter the nonlinear regime is precisely the point at which they become of interest for technological application.

To link the recent progress in basic research with vital applications to technology, a Center for Nonlinear Studies was founded at Los Alamos last fall. The Center will coordinate common themes, stimulate interchange among those in diverse fields who have similar problems, encourage collaboration between other nonlinear research centers, and bridge

the gap between fundamental research and applied technology. Researchers will no longer have to reinvent the wheel independently for each new nonlinear problem.

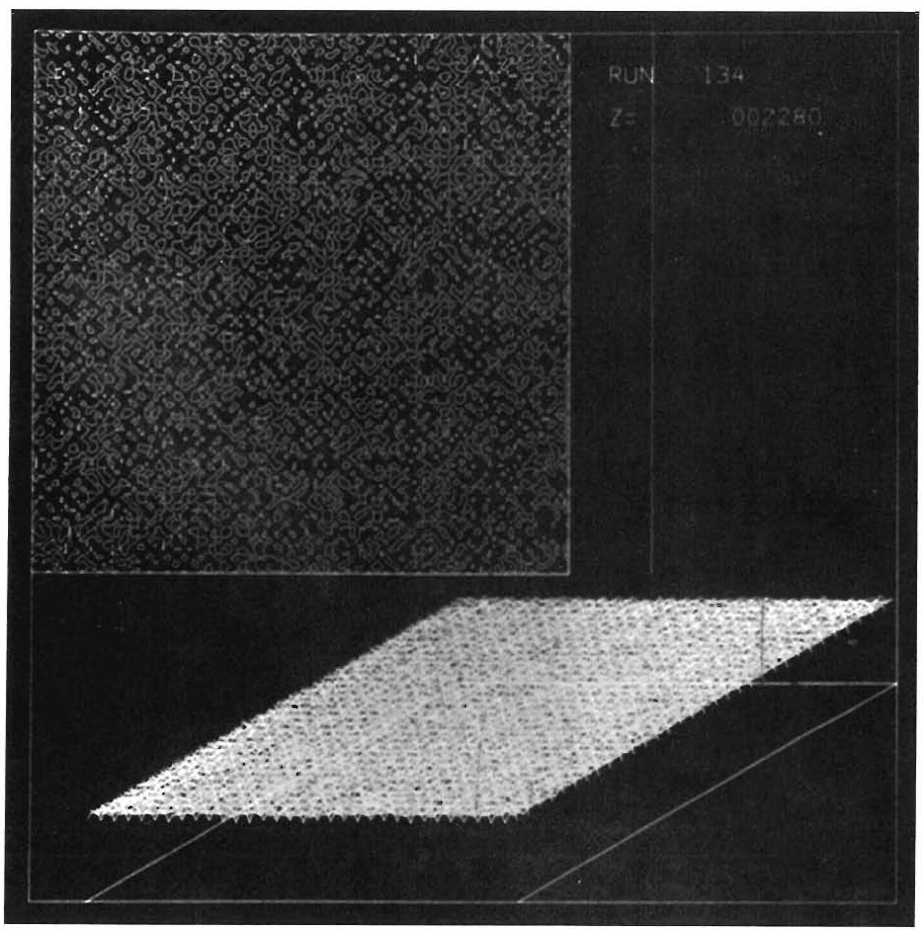
At Los Alamos, important nonlinear problems can be found in nearly every area of research, from fundamental studies of materials properties through combustion models designed to improve engine efficiency to vital questions of reactor safety. Two prime examples of nonlinear systems that are extremely important to our long-term energy goals are lasers and plasmas. Each system can exhibit complicated nonlinear behavior, and when, as in the case of inertial confinement fusion, lasers and plasmas are combined, the problems become even more complex.

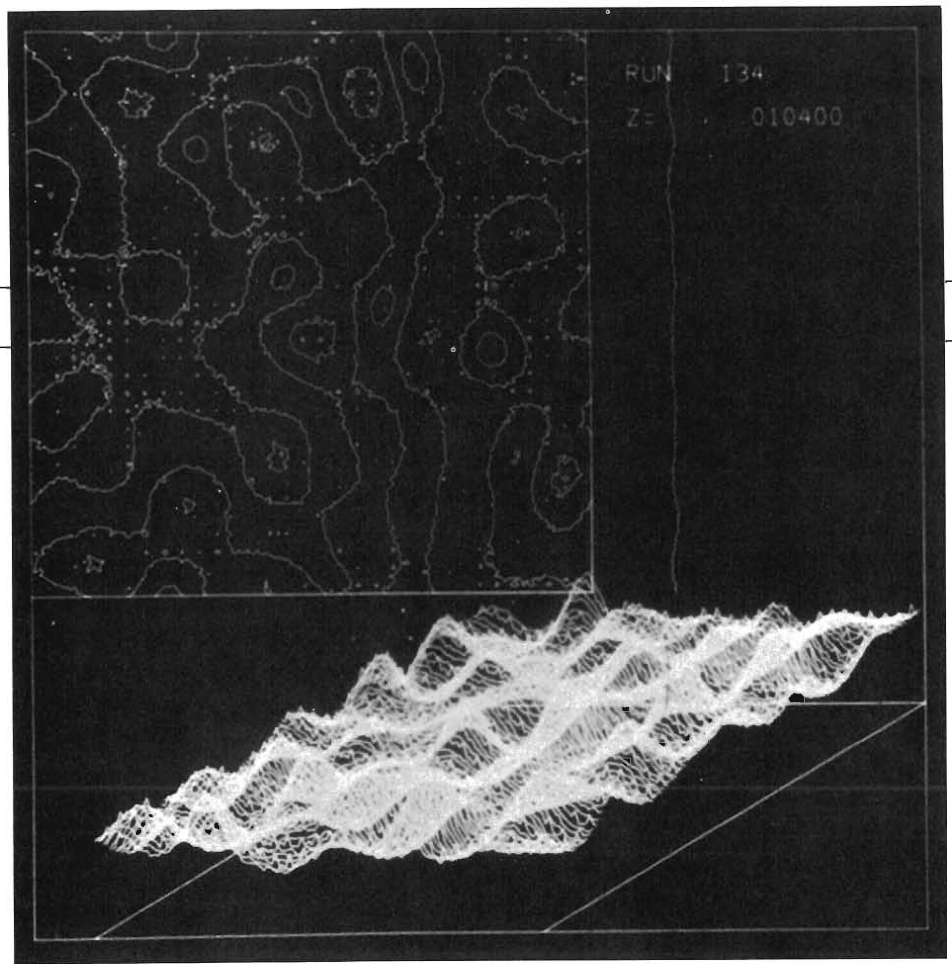
At moderate intensities, laser light passes through a medium—a gas or a solid—without major change; its transit is well characterized by a linear equation. As the intensity of the light increases, a nonlinear phenomenon called self-focusing occurs. The intense laser beam changes the index of refraction of the medium through which it moves, and thus causes the light to bend back upon itself and become more intense. As the cycle repeats, the laser beam can break up into self-focused filaments of extreme intensity; these not only disrupt the uniformity of the laser pulse, but can also destroy the laser system optics.

In plasmas, the central nonlinear problems are connected with instabilities. At low particle density and effective temperature, a plasma discharge can be relatively well controlled and made to move in a predictable, linear fashion through a magnetic field. But to achieve fusion by magnetic con-

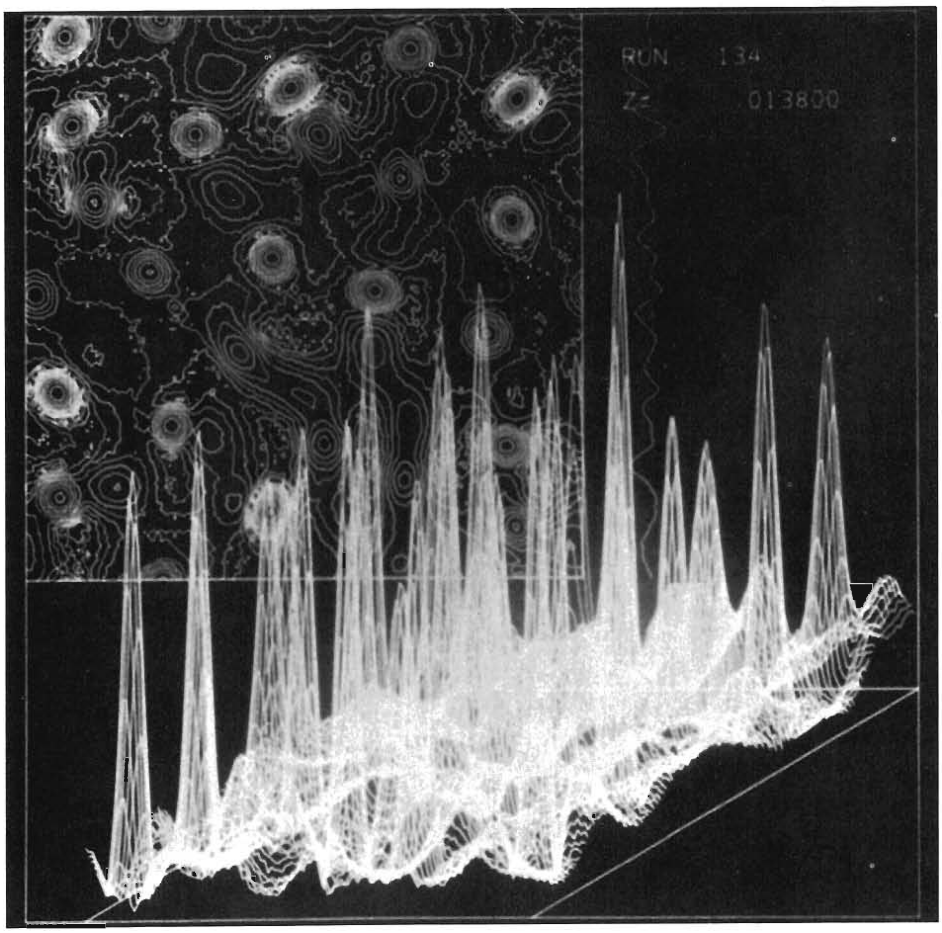
The development of a self-focusing instability in a laser beam passing through a plasma is shown in these three frames from a Los Alamos computer-generated movie. The frames show both a contour plot (upper left) and a projected plot of the laser intensity across the profile of the beam. When the laser beam enters the plasma (a), it is of essentially uniform intensity. As the beam moves into the plasma (b), the self-focusing instability initiates the formation of filaments of high intensity. The self-focused filaments become fully developed as the beam moves even further into the plasma (c). (Courtesy of Fred Tappert, University of Miami and Laboratory Consultant)

(a)



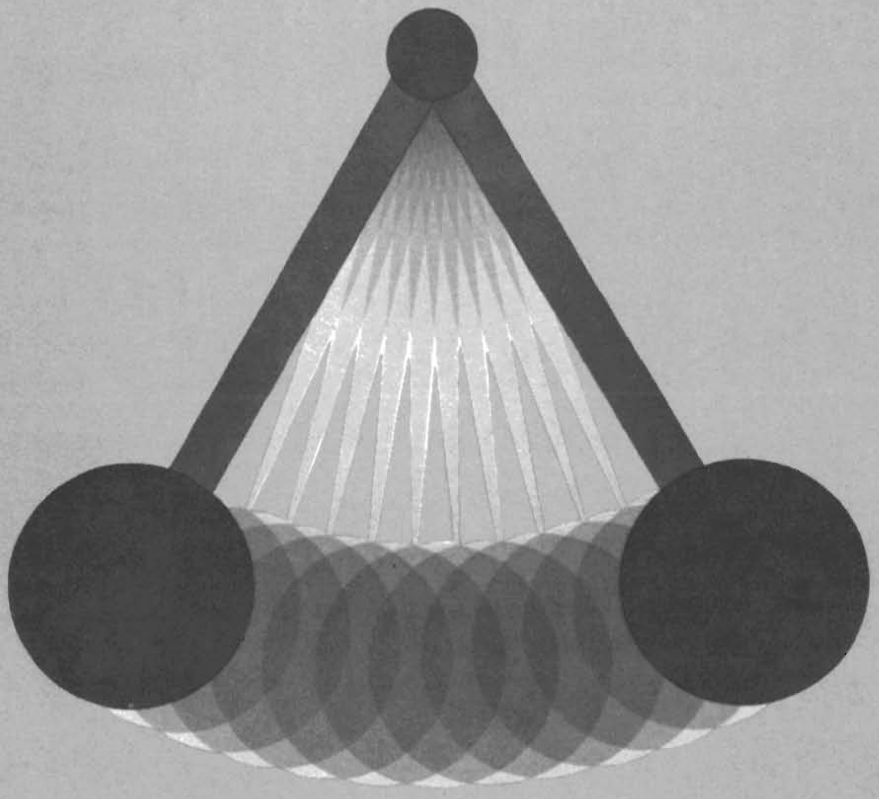


(b)



(c)

Galileo Revisited



One of the simplest examples of a nonlinear phenomenon arises in a system that was deeply involved in the birth of modern physics: the simple plane pendulum. When Galileo observed the swinging chandelier in the cathedral at Pisa and drew the conclusion that the period of its motion was independent of amplitude, he did not realize that his considerations contained a crucial linearization approximation. With the hindsight of 300 years, understanding this approximation has become an elementary application of Newton's Laws.

The force acting on the plane pendulum is that of gravity. The pendulum's motion is completely described by giving θ, the angle that it makes with the vertical, as a function of time. The component of gravitational force that drives the pendulum is, by elementary trigonometry, $-mg \sin \theta$. Thus Newton's law, $\vec{F} = m\vec{a}$, becomes

$$-mg \sin \theta(t) = m\ell \frac{d^2\theta(t)}{dt^2}$$

or

$$\frac{d^2\theta(t)}{dt^2} + \frac{g}{\ell} \sin \theta(t) = 0 .$$

This is a nonlinear equation because the sum of two given solutions, $\theta_1(t)$ and $\theta_2(t)$, is *not* a solution. That is,

$$\frac{d^2}{dt^2}[\theta_1(t) + \theta_2(t)] = \frac{d^2\theta_1(t)}{dt^2} + \frac{d^2\theta_2(t)}{dt^2}$$

but

$$\begin{aligned} \sin[\theta_1(t) + \theta_2(t)] &= \sin \theta_1(t) \cos \theta_2(t) + \sin \theta_2(t) \cos \theta_1(t) \\ &\neq \sin \theta_1(t) + \sin \theta_2(t) . \end{aligned}$$

Thus the motion of even a simple plane pendulum is nonlinear and, in fact, the period of its motion *does* depend on the

amplitude. What Galileo really observed was the behavior of the pendulum for very small values of θ. In this regime, we can use the approximation

$$\sin\theta(t) = \theta(t) - \frac{\theta^3(t)}{3!} + \frac{\theta^5(t)}{5!} + \ldots ,$$

and notice that if θ is very small, then θ^3 is even smaller. Hence keeping only the first term in the expansion of sin θ gives the *approximate* equation for the simple pendulum

$$\frac{d^2\theta(t)}{dt^2} + \frac{g}{\ell}\theta(t) = 0 .$$

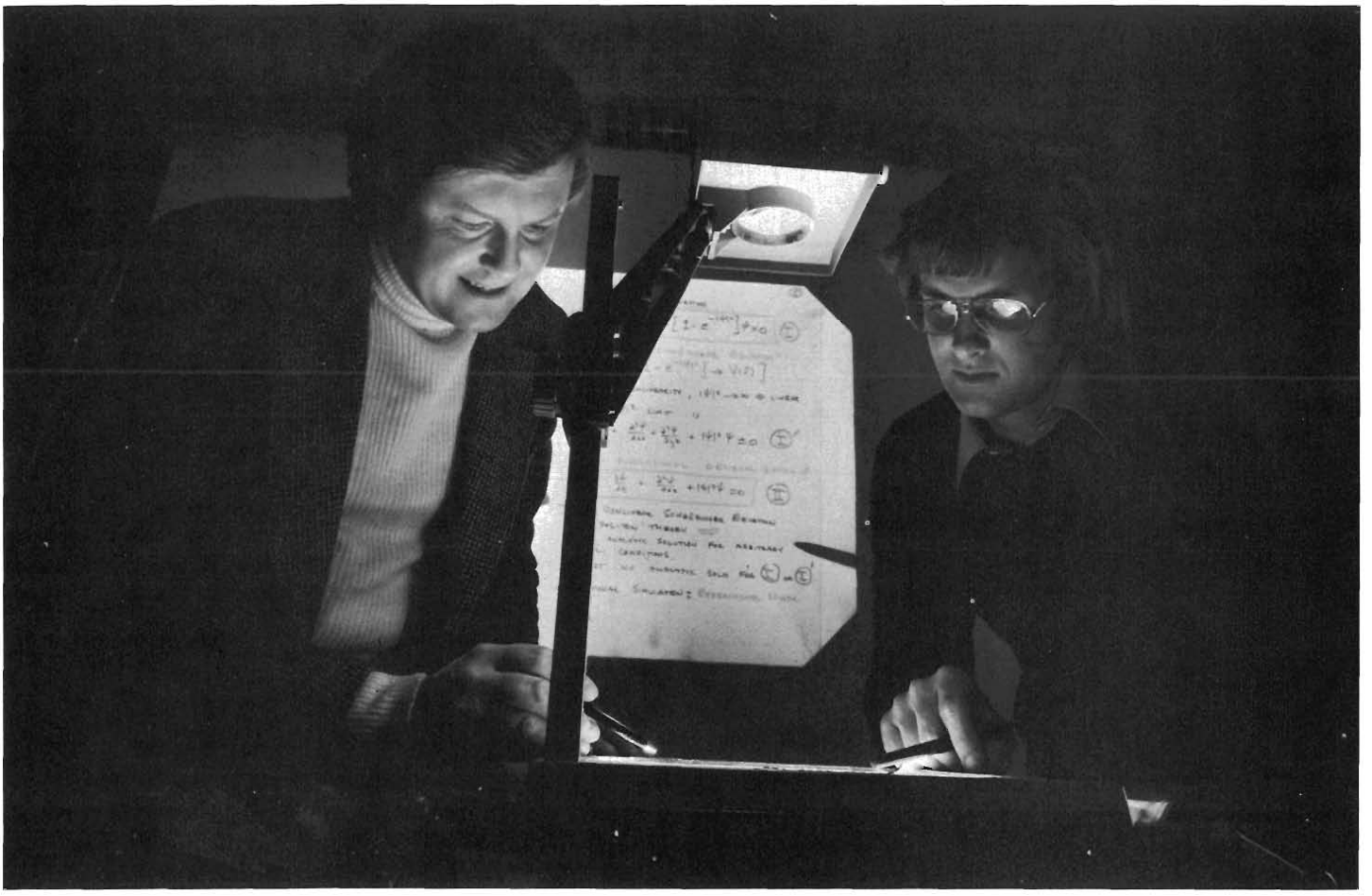
This equation is manifestly linear and its solution demonstrates that the period of the motion is independent of its amplitude. Hence the "familiar" result for the simple pendulum is the consequence of a linearizing approximation applied to a fundamentally nonlinear problem.

It is natural to ask whether the erratic, "unpredictable" behavior often associated with nonlinear systems occurs in the simple pendulum for large amplitude. The answer is a firm "almost": that is, the pendulum by itself, even for large amplitude, is too simple a system to reveal the true complexities arising from nonlinearity. But, if the pendulum is driven—as it would be in a clock—and damped—as it is slightly by air resistance—then, for certain ranges of the driving and damping forces, the pendulum exhibits the full complexity of a nonlinear system. In particular, in a manner that can be analyzed quantitatively by a universality theory [see Los Alamos Science 1, No. 1, 4 (1980)], the pendulum proceeds through a sequence of subharmonic bifurcations leading to essentially turbulent motion ■

finement, the densities and temperatures must be increased until the plasma enters a highly nonlinear regime, one in which every slight change in circumstances tends to drive the system to instability.

New problems arise when plasmas interact with intense laser beams. The laser must compress the plasma to very high densities before fusion will occur, but the interface between the laser light and the plasma becomes highly unstable and fragments into jagged, fingerlike protusions. This "Rayleigh-Taylor" instability plays havoc with the careful estimates of correct plasma density and time factor needed to initiate a fusion chain reaction. Such instabilities stand between us and the goal of controlled fusion—the hoped-for energy technology of the future—and they are not confined to lasers and plasmas. Essentially the same Rayleigh-Taylor phenomenon bedevils attempts at secondary oil recovery: when water is used to force oil to the surface, the interface between the two is unstable. To recover the maximum amount of oil from a given field, this instability must be understood and, if possible, transformed by clever technology from a vice into a virtue.

The occurrence of similar phenomena in such apparently different problems as laser-plasma interactions and oil recovery is one source of great optimism



Basil Nichols (left) and David K. Campbell of the Laboratory's Theoretical Division have charge of the administrative details of the new Los Alamos Center for Nonlinear Studies.

among experts in nonlinear science; it suggests that progress in one problem may quickly translate to others. Then too, today's fast, sophisticated computers are providing knowledge of the types of mathematics necessary to model nonlinear phenomena. Finally, recent progress in developing more explicit methods of solution for limited classes of nonlinear problems and, more generally, the emergence of global methods of qualitative analysis, encourage the hope that a deeper analytic understanding is close at hand.

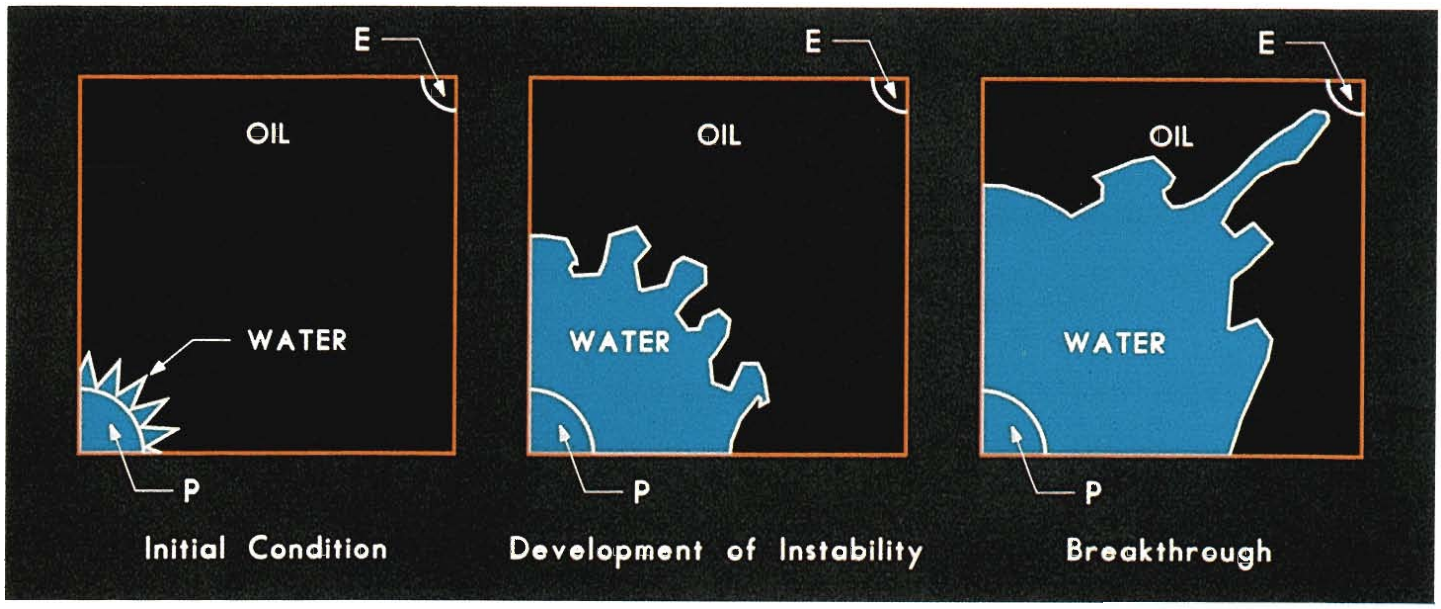
Buoyed by this optimism and driven by the importance of the challenge, the new Los Alamos Center for Nonlinear Studies is mounting a unified attack on nonlinear problems. The Center is designed to foster the interplay of researchers in many fields and will provide a mechanism for correlating recent theories in mathematics and physics, numerical simulations, and physical experiments. The Center will also interact with other centers of nonlinear research, particularly those at University of California campuses.

To stress its interdisciplinary nature, the Center will not be associated with any specific Laboratory technical division but instead will be directed by a chairman on the staff of the Los Alamos Associate Director for Physics and Mathematics. Visiting scientists will work with Laboratory personnel on

projects based on general nonlinear themes; they will be guided by a small permanent Center staff.

To strengthen ties with the academic community, a number of Senior Fellows from universities and research institutions will be appointed. Until a permanent chairman is named, the Center's efforts will be guided by a steering committee chaired by Mark Kac of Rockefeller University and including Rutherford Aris, University of Minnesota; Roger F. Dashen, Institute for Advanced Study, Princeton University; Martin D. Kruskal, Princeton University; Alwyn C. Scott, University of Wisconsin; J. Robert Schrieffer, University of California, Santa Barbara; and I. M. Singer, University of California, Berkeley.

David K. Campbell and Basil Nichols of the Laboratory's Theoretical Division are handling the Center's administrative responsibilities ■



In some advanced methods of oil recovery, water is pumped into reservoirs to force oil to the surface. The diagrams above illustrate how instability of the oil-water interface effectively limits the amount of oil that can be recovered. The quarter circles P and E represent the water-input pump and the oil-extraction pump, respectively. The initial form of the oil-water interface shown here, symmetric and five-fingered, was chosen to model certain aspects of an actual recovery problem. At "breakthrough" the unstable finger-like protrusion of water reaches the extraction pump and no further oil can be recovered. In this modeling, the change in area occupied by water from initial conditions to breakthrough is directly related to the amount of oil recovered. Clearly, maximum oil recovery by this method requires understanding and control of the interfacial instability. (Courtesy of J. Glimm and E. Isaacson, The Rockefeller University, and D. Marchesin and O. McBryan, Courant Institute of Mathematical Sciences)

IGPP Workshop Report: Plasma Physics Near the Earth's Bow Shock

by S. Peter Gary

When the supersonic flow of the solar wind first encounters the earth's magnetic field it creates a shock wave. This interaction compresses the magnetosphere on the dayside and shapes it into an elongated teardrop on the nightside (Fig. 1). The ultimate consequences of this interaction are far-reaching disturbances in our atmosphere, such as magnetic substorms that interfere with power transmission and communications and produce the spectacle of the polar auroras. The wave is called the bow shock, in analogy to the bow wave of a boat, and is a jump in plasma density, temperature, and magnetic field associated with the transition from supersonic to subsonic flow.

So turbulent is the shock that we have been unable to model the highly nonlinear processes that determine its structure. But progress has been made in modeling the foreshock region (upstream of the bow shock) where energetic protons reflected from the shock back toward the sun may help to heat, decelerate, and deflect the solar wind.

New data on energetic protons and two plausible models of their interaction in the foreshock were discussed last September at a Workshop on Space Plasma Physics held at Los Alamos and supported by the University of California's Institute of Geophysics and Planetary Physics. Both models rely heavily on data from space probes mounted by NASA's International Sun-Earth Explorer (ISEE) mission. One model, developed by Los Alamos and the Max Planck Institute for Extraterrestrial Physics (Garching, West Germany), postulates a causal relation between various populations of energetic protons. A second foreshock model, proposed by E. W. Greenstadt, postulates that one type of proton population and its associated large-amplitude magnetic fluctuations are intrinsic parts of the bow shock structure.

Under typical conditions, the bow shock stands at about 15 earth radii from our planet along the Sun-Earth line, but since the solar wind is subject to fluctuations in flow speed, the shock can

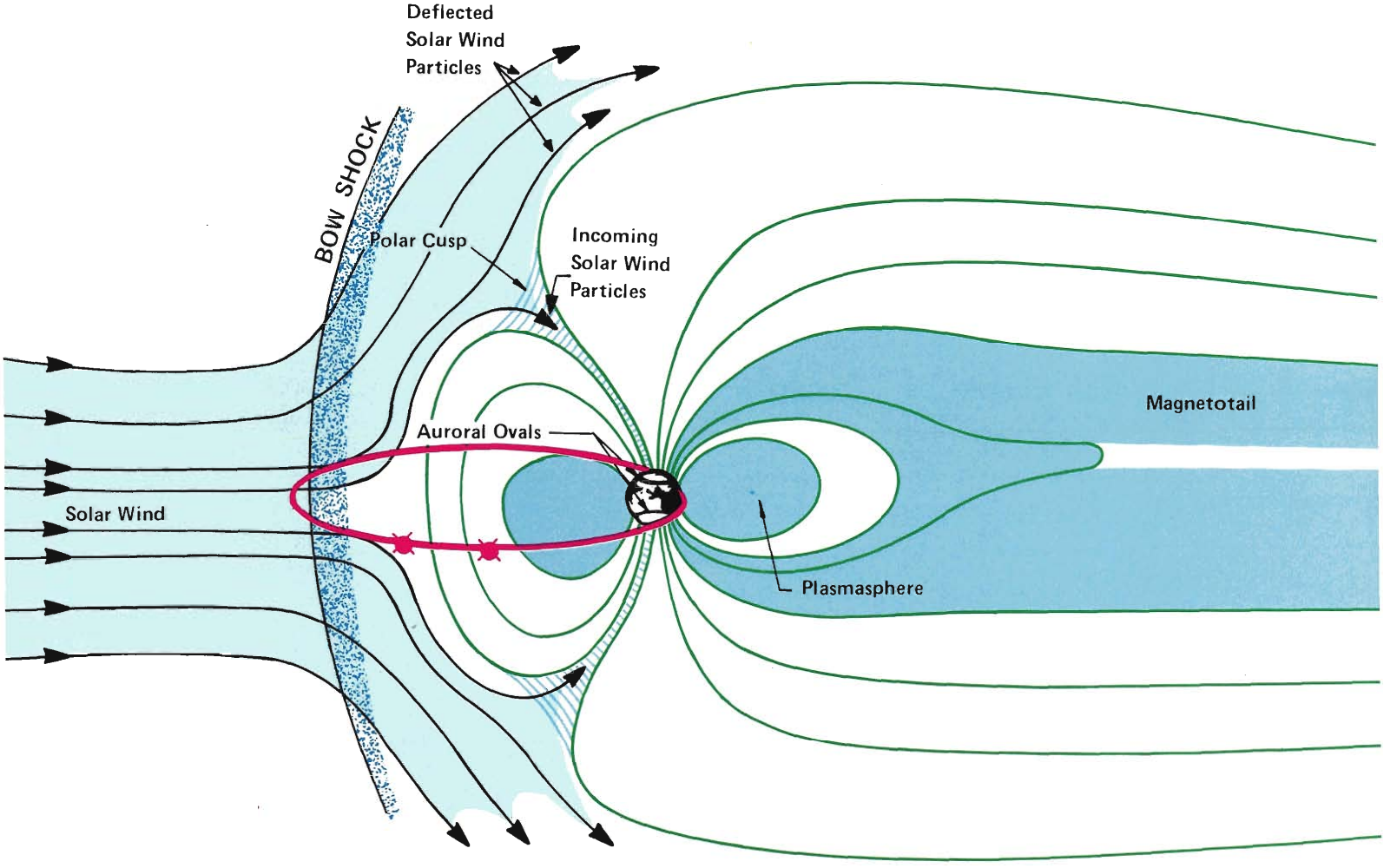


Fig. 1. A current picture of geospace. The solar wind incident on the magnetosphere is less dense than the earth's atmosphere by a factor of ~10^{18} and consists primarily of protons and electrons traveling away from the sun at ~400 km/s with thermal energies of ~10 eV. Particles deflected by the earth's magnetic field are sometimes trapped in the magnetotail. Frequently, magnetic energy stored in the tail is suddenly released, accelerating the plasma toward the earth. Such a magnetic substorm ultimately produces strong magnetic disturbances and spectacular auroras in the earth's polar regions. Two co-orbiting International Sun-Earth Explorer satellites launched in 1977 are gathering important data on the solar wind, bow shock, and magnetosphere.

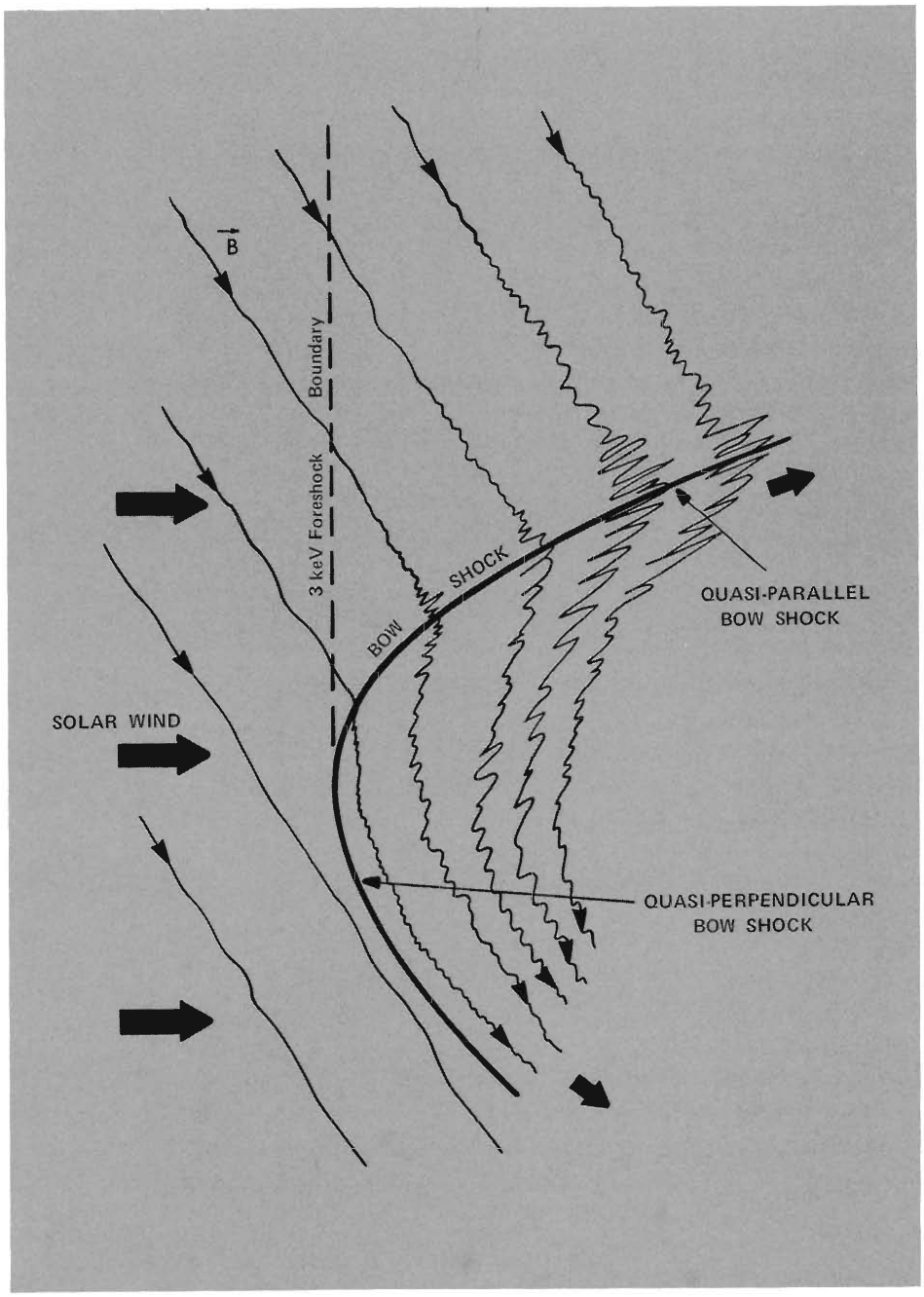


Fig. 2. The earth's bow shock and near upstream region. Also shown is the foreshock boundary for 3-keV protons traveling along $\vec{B}$ against the solar wind. Reflected protons with energies less than 3 keV are found between this boundary and the shock. The detailed behavior of the magnetic fluctuations at the quasi-parallel shock is not well understood and their representation here is only suggestive.

change position quite quickly.

The bow shock's local structure is determined to a great extent by the angle θ between the shock normal and the solar wind magnetic field $\vec{B}$. (Fig. 2). A perpendicular shock ($\theta = 90°$) is a rather abrupt change in plasma properties across a narrow region. A parallel shock ($\theta = 0°$ or $180°$) is hardly a shock in the classic sense but a rather broad transition embedded in large-amplitude fluctuations (turbulence). Between these two extremes are quasi-perpendicular and quasi-parallel shocks.

While the bow shock is capable of reflecting and accelerating charged particles along magnetic field lines back toward the sun, the solar wind continuously sweeps them toward the shock, creating the foreshock boundary that limits the region in which a proton with a given velocity parallel to $\vec{B}$ can be found.

A plasma, like a fluid, can sustain a diversity of fluctuations termed (at small amplitudes) waves, (if growing) instabilities, or (at sufficiently large amplitudes) turbulence. In the case of plasmas, such fluctuations represent temporal and spatial changes of both magnetic field and plasma properties, such as ion densities and velocities. Consequently, to characterize a plasma wave completely requires careful correlation of field and plasma data.

Such information about the solar wind, bow shock, and magnetosphere has been collected by two Explorer spacecraft launched into the same orbit at an apogee of 22.2 earth radii in October, 1977. Since the distance between the two satellites is variable, the data have allowed space scientists, for the first time, to separate temporal and

Los Alamos Establishes Branch of IGPP

More than 20 years ago, the University of California at Los Angeles formed an Institute of Geophysics and Interplanetary Physics (IGPP) to close the hiatus that had existed between physics and geology into which fell such fields as rock physics and high-temperature mineralogy. The program was so successful that IGPP branches were established at the University of California at San Diego and Riverside and in September last year, a branch was formally initiated at Los Alamos.

Each IGPP branch provides a home within the University for types of research that in the past, at least, have had trouble fitting into existing programs. Special needs were met by each branch. At UCLA, NASA-supported extraterrestrial research and space physics became an important part of the Institute. San Diego's program was founded to support the growth of physical oceanography and marine geophysics, and at the Riverside branch, cosmic ray astronomy was the original thrust.

At Los Alamos, the purpose of the new branch is to provide a strong link between the Laboratory and the University and to foster the joint use of facilities by scientists involved in geophysics and interplanetary physics. The Laboratory hosted its first IGPP Workshop in Space Plasma Physics last fall, to allow researchers to meet and exchange information.

Initially the Los Alamos program will involve faculty exchange, with actual development of research programs evolving with time. The range of research that may be supported by the IGPP will include geothermal energy, tectonophysics, seismology, planetary exploration and planetary geology, and related fields including cosmic-ray study, solar physics, space plasma physics, solar system dynamics, atmospheric science and climatology, solid-state physics, high-pressure physics, and geochemistry. All research sponsored by the IGPP will be unclassified, and one of the duties of the Assistant Director of the Institute will be to advocate opening research facilities at the Laboratory to University scholars. The Assistant Director will report to the systemwide IGPP Director and to the Los Alamos Director regarding Laboratory policy, staff, and programs, and will have both line and program responsibility within the Laboratory's matrix organization.

The Institute's initial budget includes a $250,000 allocation for Regents' Fellowships that provide salary, transportation, and per diem expenses for both scholars and Los Alamos staff members involved in IGPP programs. The terms of Fellowships will be flexible and geared to individual circumstances. Support for a graduate student working here under the direction of a professor at his home university will be limited to $15,000 for a nine-month term. Support for a nine-month visit by a *full* professor will amount to $40,000. It is anticipated that the Fellowship fund will support seven or eight scholars each year. All Fellowships will support visits of more than three months, and thus will not be used for summer study.

Proposals for Fellowships should be submitted to the Los Alamos IGPP Assistant Director, who will endorse those candidates whose proposals are to be forwarded to the applicants' University Department Chairmen for approval. Successful applications will be submitted to the IGPP systemwide Director, who will make the final decision on awards ■

spatial effects and determine velocities and thicknesses of magnetospheric boundaries. Magnetic fields were measured by magnetometers developed by the University of California at Los Angeles and ion velocity distributions were measured by fast plasma analyzers developed jointly by Los Alamos and the Max Planck Institute.

First analyses of data showed that there exist two distinct populations of backward-streaming energetic ions: (1) "reflected" protons with sharply peaked beam-like velocity distributions along $\vec{B}$ and energies seldom extending much above 10 keV (Fig. 3a) and (2) "diffuse" protons with relatively broad velocity distributions extending to considerably higher energies (Fig. 3c). Large-amplitude magnetic field and ion-density fluctuations (with periods ≤ 1 min) are associated with the diffuse protons, but not with the reflected protons.

At the Workshop, more recent ion velocity distribution data were presented

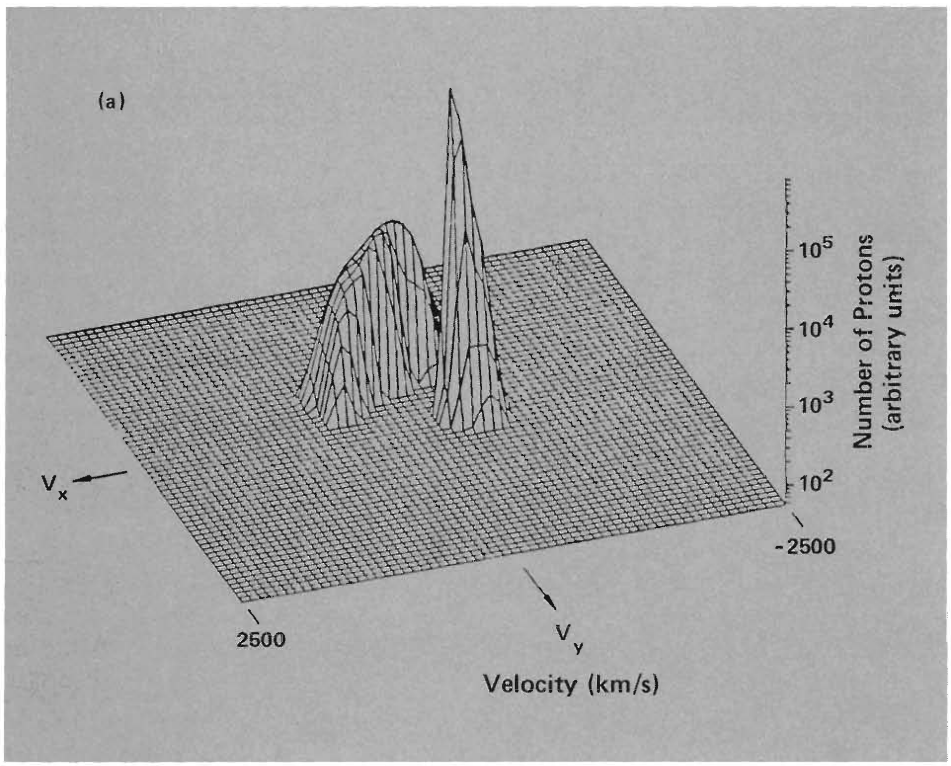
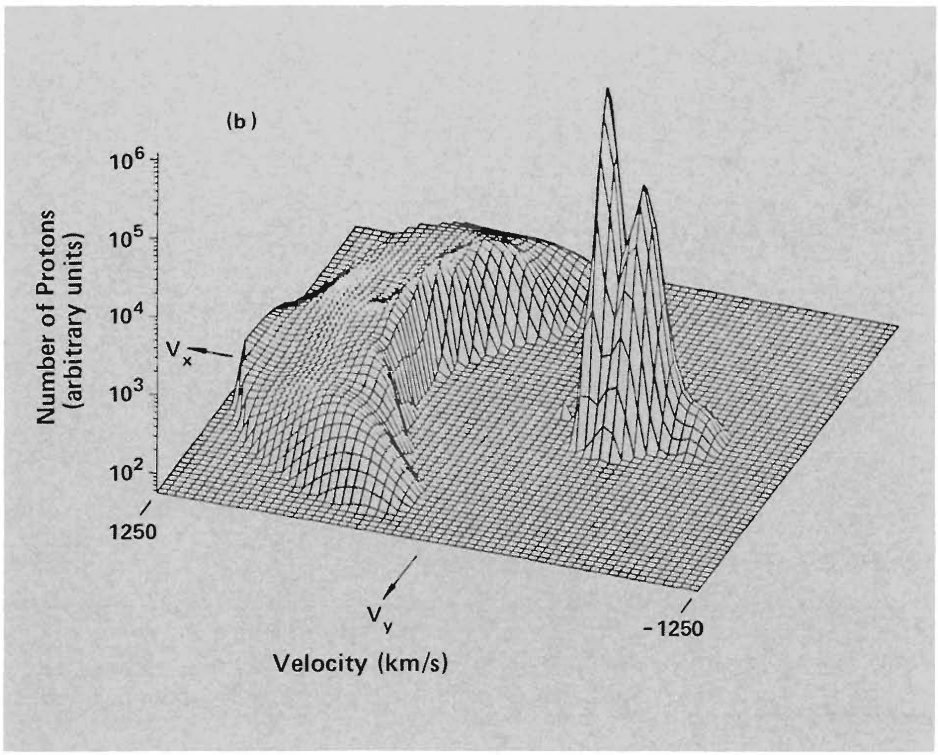
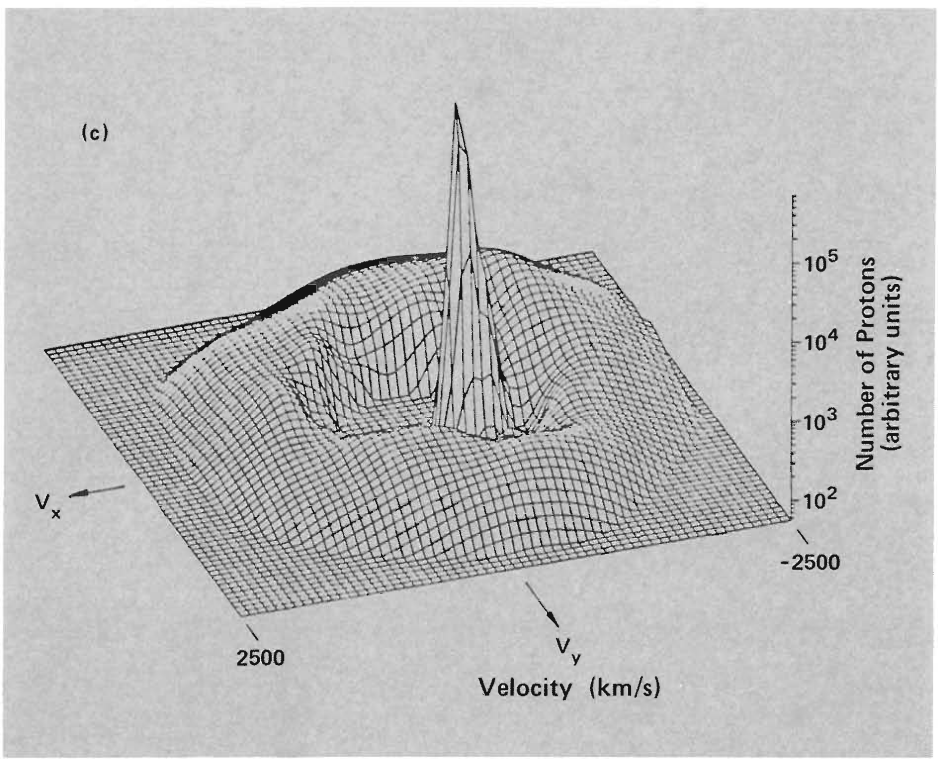


Fig. 3. Typical proton velocity distributions in the foreshock region. The component with the higher peak (greater density) corresponds to the solar wind protons. The component with the lower peak represents the (a) reflected, (b) intermediate, and (c) diffuse energetic proton populations. The magnetic field is approximately in the direction of the average velocity of the energetic protons; the v_x axis points back toward the sun.





showing the intermittent presence of a third proton population. This "intermediate" population (Fig. 3b) is spread out in velocity space and appears to be a transition between the reflected and diffuse proton populations.

Also reported at the Workshop were complementary data on magnetic field fluctuations. Relatively small-amplitude, relatively high-frequency (~1 Hz) magnetic fluctuations accompany the reflected protons. These waves as well as larger-amplitude, lower-frequency (~0.03 Hz) fluctuations are present in association with the intermediate population. In the presence of the fully diffuse proton population, the high-frequency magnetic waves disappear, and the low-frequency fluctuations steepen into shock-like features that often break into whistler-mode packets. Both high- and low-frequency fluctuations have compressive components (that is, plasma density fluctuations accompany the magnetic field fluctuations) and in general propagate at angles oblique to $\vec{B}$.

Los Alamos and Max Planck Institute

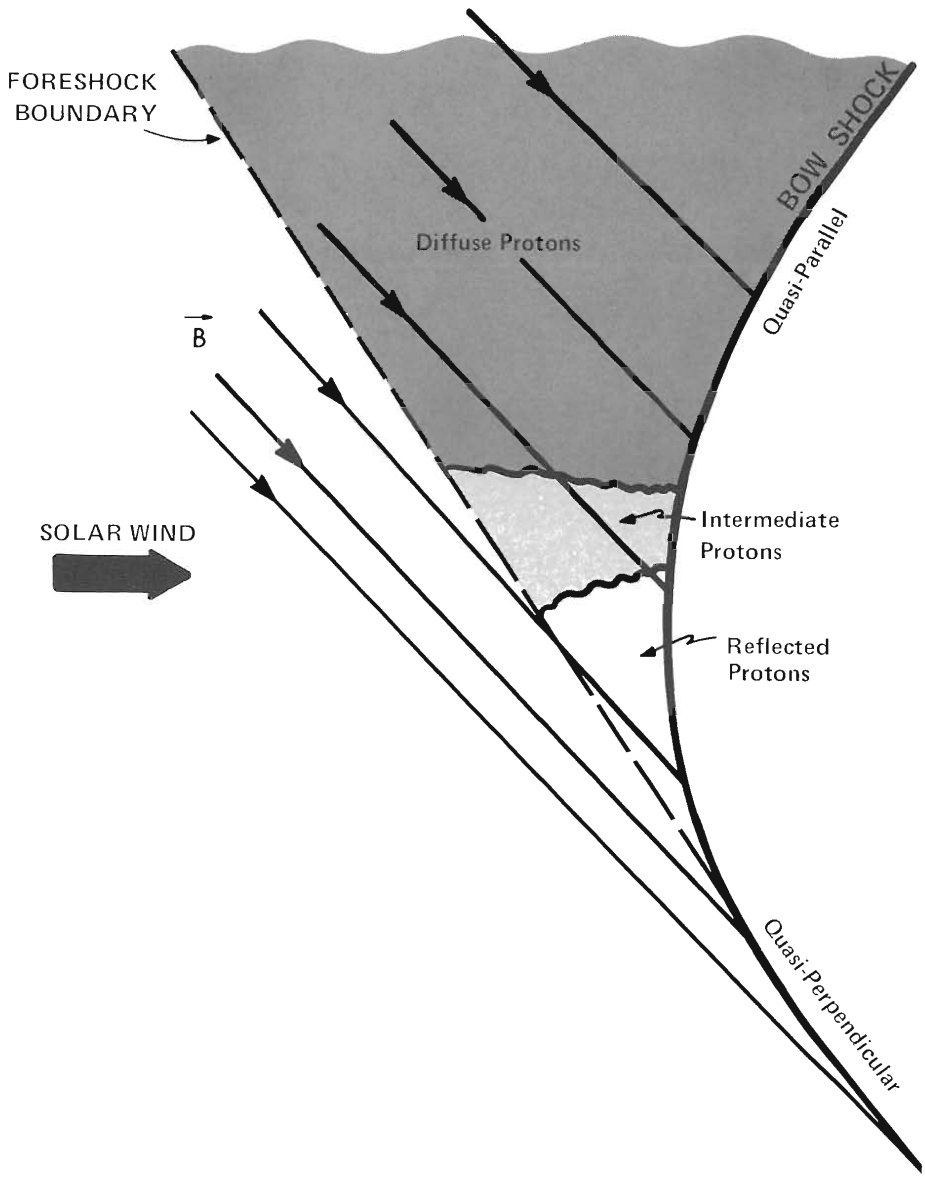


Fig. 4. Spatial distribution of the various proton populations predicted by the Los Alamos-Max Planck Institute model.

researchers postulate a model in which a beam of protons reflected from the shock transfers its backward momentum to the solar wind through wave-particle scattering, thereby decelerating the solar wind and creating the intermediate and diffuse proton populations. In this model the reflected proton population is formed by ion reflection at the quasi-perpendicular bow shock of a small fraction (~1%) of the solar wind protons. (Ion reflection is a characteristic of high Mach number collisionless shocks.) The resulting non-Maxwellian proton velocity distribution is unstable to one or more plasma instabilities, which grow into large-amplitude magnetic fluctuations. These magnetic waves pitch-angle scatter the beam (that is, change the direction of a proton's velocity vector without significantly altering its magnitude). Such wave-particle scattering leads to the intermediate and eventually to the nearly isotropic diffuse distributions.

This model explains the observed deceleration of the solar wind as it enters the upstream region populated by diffuse protons and long-period waves. In addition, it predicts the location of the three proton populations in the foreshock region (Fig. 4). Detailed data analysis is now underway to test this prediction. Early results indicate correlations between reflected protons and quasi-perpendicular shocks and between diffuse protons and quasi-parallel shocks as suggested by the model.

An important limitation on refinement of this model is theoretical. Although the magnetic fluctuations are observed to be compressive, present theories are limited to noncompressive instabilities that cannot produce the high-energy component of a typical diffuse proton population.

The other model of the upstream region, presented by Greenstadt *et al.*, proposes that, rather than a causal link between the two proton distributions, both elements are part of the overall picture of the bow shock. Although Greenstadt's model is at present lacking in details, some such model may eventually provide a better explanation of the observed association of diffuse protons with quasi-parallel shocks, of the high-energy component of the diffuse population, and of the numerous cases of wave appearance immediately after the solar wind magnetic field undergoes a rapid local change from a quasi-perpendicular to a quasi-parallel geometry.

Many other subjects were discussed at the Workshop, including magnetospheric dynamics, the ionosphere, magnetic reconnection, and numerical simulations of space plasmas. There was general agreement that, in spite of the increased sophistication of recent spacecraft and an increased understanding of some small-scale processes in space, scientists are still a long way from a comprehensive theory of plasma dynamics in the earth's environment.

Workshop participants were enthusiastic about the opportunity to exchange information, and a second Workshop, to be hosted by the University of California at Los Angeles, is planned for later this year.

Scientific presentations of the Los Alamos Workshop are abstracted in "Workshop on Space Plasma Physics," edited by M. Ashour-Abdalla and S. P. Gary; this publication is available from S. P. Gary, Group P-4, Los Alamos National Laboratory ■

FURTHER READING

G. Paschmann, N. Sckopke, J. R. Asbridge, S. J. Bame, and J. T. Gosling, "Energization of Solar Wind Ions by Reflection From the Earth's Bow Shock," J. Geophys. Res. **85**, 4689 (1980).

S. J. Bame, J. R. Asbridge, W. C. Feldman, J. T. Gosling, G. Paschmann, and N. Sckopke, "Deceleration of the Solar Wind Upstream From the Earth's Bow Shock and the Origin of Diffuse Upstream Ions," J. Geophys. Res. **85**, 2981 (1980).

E. W. Greenstadt, C. T. Russell, and M. Hoppe, "Magnetic Field Orientation and Suprathermal Ion Streams in the Earth's Foreshock," J. Geophys. Res. **85**, 3473 (1980).

Interested readers are directed to a forthcoming volume of the Journal of Geophysical Research, which will be devoted to research related to the earth's foreshock.

An Interview with Peter Carruthers

Physics, Philosophy, Leadership, & Policy

by Leonard M. Simmons, Jr. and Geoffrey B. West

Peter Carruthers recently stepped down — or stepped up, as he puts it — from a seven-year tenure as Leader of the Los Alamos Theoretical Division to return to the main work of his professional life — research in pure physics.

During these seven years we have seen a new side of Pete — a tough leader with vision, foresight, and an instinct for making things happen. He has changed the image of the Laboratory in the eyes of the scientific community, and has fought hard and successfully in Washington for support of basic research in physics.

The metamorphosis from a scholarly professor to tough Division Leader was indeed a shock to us. We had known him at Cornell University as something of a boy wonder, dedicated to his work and surrounded by graduate students not much younger than he. His breadth of experience in both solid-state and particle physics was rare among his contemporaries. His openness, encouragement, and enthusiasm for new ideas — his sardonic wit, good judgment, and appreciation for real talent attracted many students to him. With his horn-rimmed glasses, mild exterior, and office overflowing with books and papers, he appeared more like an old-fashioned scholar of classical manuscripts than a hard-driving physicist on his way to the top. He had and still has serious interests in music, bird watching, and trout fishing and, of course, an intense deep love of physics. Thus we were somewhat surprised when Harold Agnew, then Director of Los Alamos, invited Pete to become Theoretical Division Leader; even more surprising, however, was that Pete accepted!

Although Pete entered the Laboratory as a novice in administration, he used what leverage he had to accomplish a great deal. He restructured the Theoretical Division and established new groups in many areas (high-energy physics, theoretical biology, statistical physics and materials theory, theoretical molecular physics, applied mathematics, and detonation theory). He stimulated intellectual excitement and a strong sense of exploration; he hired talented people and left them free to work. He brought eminent scientists from outside to participate in the life of the Laboratory. It was an uphill battle and for those who had known Pete in his previous life, quite something to see! To us, Pete was Clark Kent stepping into a telephone booth before important meetings, to emerge as Superman ready to fight for what he believed in.

Pete had not been part of the scientific establishment before coming to Los Alamos, but his new position gave him entrée into the corridors of power and he quickly took full advantage of the opportunity. He became involved in national science policy, participating as Chairman of the National Science Foundation's Physics Advisory Panel, member of the National Academy of Sciences' Committee on U.S.-USSR Cooperation in Physics, member of the Department of Energy's High Energy Physics Advisory Panel, and Chairman of the Board of Trustees of the Aspen Center for Physics. He joined JASON, a group of U.S. scientists who work three months each year on scientific and policy aspects of the country's defense and energy problems. He still holds many of these positions today, and recently he has been appointed a Senior Fellow of the Laboratory.

We now know Peter as someone who cares deeply on a grand scale, but still remains a champion of the little guy, the bright young scientist who needs to find a job. He can be ruthless and uncompromising and act harshly when necessary (but not without the side effect of sleepless nights). As one of the few Division Leaders at Los Alamos to have been appointed directly from an academic position at a great university, he has fared well and served well. He has become a force in the world of science policy and will undoubtedly continue in that role.

Physics, Philosophy, Leadership, & Policy

"Leadership is the king waving the flag in front of the army saying, 'Let's go get the bastards.'"

INTERVIEW

"There has to be a feeling of freedom and reward. You can't get good science out of people who recognize that they are being managed."

SIMMONS: When you came to Los Alamos your experience in administration consisted of running an undergraduate course of maybe 50 people. What made you think you could run a big theoretical division?

CARRUTHERS: It never occurred to me that I couldn't do a thing like that. I just thought it took good judgment. Of course, it takes mostly stamina.

SIMMONS: Do you think administrative experience is important for most administrative jobs in science?

CARRUTHERS: A little bit will make life easier, but I don't think that's as important as having the right instincts for finding good people and letting them do what you want done. The key issue is that the people who work for you respect you and respect your judgment and fairness. Of course an illustrious scientific reputation can get you respect, but you may be a wretched administrator, nevertheless. On the other hand, a good administrator *per se* may not have sharp scientific judgment. And neither may be able to judge the quality of people. Even if you're a competent scientist, after a while you may lose the freshness of that competence and, since there's no reason for a scientist to be an administrator in a scientific establishment unless he maintains scientific judgment and competence, we might be better off with good administrators. It's a very complicated business.

One of the principal evils of the federal science establishment, both in Washington and outside, is the emergence to power of a permanent ruling class of rotating bureaucrats who don't command the respect of the people that they control. There is an entropy death facing American science with its present trend to expansion of titles, and functions, and memos, and the ever-present Xerox and Kodak machines that simply produce communications and the need for more communications.

Ever since I've been here there's been an increasing trend, both externally and internally, towards the illusion that you can manage science, whereas all you can really do is to get good people who are interested in the subject you want to develop. This increasing accountability at all levels of the federal establishment exudes a cold air that drives out the kind of neurotic and creative people that you need to make a breakthrough. There has to be a feeling of freedom and reward. You can't get good science out of people who recognize that they are being managed.

Of course you do have to deliver, but there is a way of getting to the answer by leadership, which I distinguish from management. Leadership is the king waving the flag in front of the army, saying "Let's go get the bastards." Then the captains will race along enthusiastically and fight the battle. With management the king sends a telegram from 50 miles behind the lines saying "Why aren't you to latitude 42.54? According to our long-range plan, you were supposed to have been there last week. Kindly fly to Washington and explain why you are not yet at your milestone." One thing that can be done to improve things is to encourage the few people who have both first-rate scientific talent and some aptitude or tolerance for leadership and responsibility to get in there and take their turns. That should mean that it's not a lifelong sentence that will destroy their research careers. Another thing is to make sure that there aren't so

many rewards for mere politicians and entrepreneurs. That requires the best leadership at the highest level.

WEST: Do you have any regrets, having left Cornell?

CARRUTHERS: I often feel quite nostalgic for Ithaca itself. But I've found that people in Los Alamos are really more fun. On the other hand, my responsibilities here have kept me from doing much serious research. I was at Cornell for 17 years altogether, as a graduate student and faculty member. It was extremely peaceful, and I never realized what an idyllic, quiet place it was and how conducive it was to doing flat-out scientific research with a minimum of distraction. Of course the long, gray, wet winter encourages work.

WEST: Why did you take the job here?

CARRUTHERS: I don't know what Faustian tendencies were growing in me but I began to realize that the Laboratory was an enormous resource, and that it might be possible to do something with it. In the area of physics I knew about, which was essentially pure academic-type physics, the Theoretical Division did not have very much to offer. Its areas of excellence were in the more applied areas that I would have to learn about later. But finally I decided I needed to change my life, and I became very eager to get the job.

One of the attractions was the sheer physical beauty of the Rocky Mountains. The other thing is that Los Alamos is a very dynamic place, however confused, in which all kinds of things are going on. There is much more traffic through Los Alamos of significant scientists involved in national affairs than there ever was at Cornell. I wanted to have a try at living in a different kind of environment.

At Cornell, each department had a moat around it. There was not a community of scholarship in which humanists talked with scientists about significant issues. If there was, then I wasn't part of it. I think Los Alamos is much more integrated intellectually.

WEST: Had there been a disenchantment for you in university life?

CARRUTHERS: Yes. University life is entrenched and rigid, and it's very hard to transform the way a place is. I felt frustrated that there was no chance to influence the future. I felt this was an opportunity to change the world in some way beside writing yet another paper.

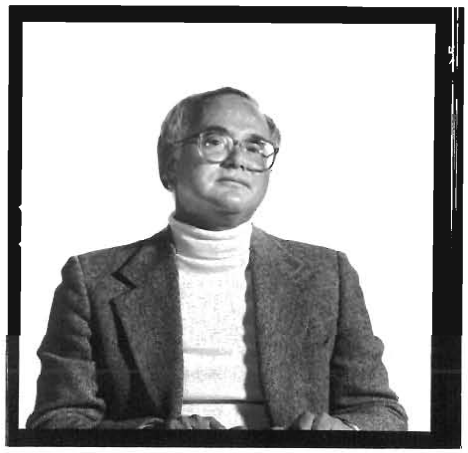
SIMMONS: Universities have the reputation of being bastions of liberalism. Don't you believe that's true?

CARRUTHERS: I certainly don't believe it's true. Especially after I testified for the students after a police riot.

University life has many virtues, and often I miss them. Especially I miss the students and the general mix of cultural opportunities. But it certainly has its limitations, and it's not clear that universities will always have the predominant role in scientific research that they have had in the past. In a way, the entire sociology of science may be undergoing a change—the same kind of change which makes it possible to have really good science at a national laboratory, presumably dedicated to giant projects and technology. It may mean something for the whole future of science in the country.

WEST: Do you see that as positive?

CARRUTHERS: If it creates excellence, then I would say it's positive; whether it's better than what went before, I don't know. It's just different,



and I'm not sure what it will be. It might be the case that in 10 or 20 years there will be more exciting, first-rate science done at institutes and laboratories than at universities. Universities are in a terrible financial situation. They are over-tenured. The age of the faculty keeps growing. The number of students is decreasing. There are very few job opportunities for young people. It's not clear that the sociology of the traditional university is going to allow the nurturing of science in the vigorous way that it was when we were coming through the system.

There are, of course, new pressures on our country which never existed before, and those pressures require that scientists pay attention to new issues. The institutes or laboratories may be the proper vehicle for that. Not to say that I don't have many criticisms of the way federally supported science is managed in this country.

WEST: Thinking about coming to Los Alamos, did you see that somehow you could play a role in national issues?

CARRUTHERS: Yes, that was a very conscious part of it. After World War II many scientists became involved in de-

Physics, Philosophy, Leadership, & Policy

"The key issue is that the people who work for you respect you and respect your judgment and fairness."

INTERVIEW

"To me one of the greatest hopes for science in this country is the informal collection of scientists who are sensititive to policy issues and willing to be advocates of science, not just for their own institutions but on the national scene."

fense issues. They became plugged in to the right power circles and were turned to for advice. Now there is a new generation coming in, and I view myself as part of that new generation.

SIMMONS: Did you think you would be able to continue to do a significant amount of science after you came here?

CARRUTHERS: I foolishly felt I could spend half my time on research. In Ithaca, if I could spend 4 hours a day working on physics that would be a good day, by the time teaching, students, etc., were taken care of. When I talked to Harold Agnew he said sure, take as much of your time as you want. But gradually it wears you down.

WEST: Did you enjoy the taste of power that you had as Division Leader—the fact that the corridors of power in Washington opened up?

CARRUTHERS: Power comes from various sources, and some of it, much of it, comes *ex officio*, if the job is sufficiently high. For example, just being a Group Leader or Division Leader has intrinsically a certain amount of fiscal power and people immediately notice that. There are people who had never paid any attention to me until I had the fiscal power of being a Division Leader. Suddenly they noticed me and were extremely friendly. Now, I'm going to find out which of them are my real friends.

In Washington, unless you have either an enormous scientific reputation or some large responsibility, the doors of power are not available to you. Of course, once you get sufficiently known to all the people who move in this sphere of influence, the titles are not so necessary. I certainly found that being a Division Leader opened up many opportunities for me, *ex officio*, that wouldn't have existed had I been just a professor at Cornell.

I enjoyed being in the center of the action very often. I suffer from the schizophrenia of wanting to be simultaneously a quiet scholar working in a corner and also making the right things happen—and making sure that the bad guys don't get in there and make the wrong things happen. There's no doubt that those things strongly motivate me.

WEST: Do you think you can return to being the quiet scholar sitting in the corner?

CARRUTHERS: That's the big question. It's not so much that I'm corrupted, but whether I'll be left alone. In this general area (on the national scene) in which genuine talent is so scarce, if you open your mouth and say a few sensible words people may overestimate your intrinsic merit in this regard. Soon you're serving on every committee that's available, and spending all your time on airplanes.

To me one of the greatest hopes for science in this country is the informal collection of scientists who are sensitive to policy issues and willing to be advocates of science, not just for their own institutions but on the national scene. These people are always in touch with each other. They prevent lots of bad things from happening, and even occasionally cause a good thing to happen.

SIMMONS: As far as Los Alamos is concerned, there seems to be constant warfare between advocates of basic science and advocates of purely programmatic work. Is that healthy?

CARRUTHERS: The people who work on practical and applied things often feel that the basic-science people are para-

sites on the body of the parent organization, and that those people even have contempt for the people who are screwing screws into hardware. There is a natural suspicion in the two camps.

In the Laboratory at large, the problem is very real. In T-Division, we have integrated the spectrum of interests, so that the people doing basic work become aware of some of the long-range, applied problems and contribute to solving those problems, and the people doing nuts and bolts work may be helped by being close to people working at the frontiers of science. Both extremes profit from the existence of the other, if they will only agree to be friends.

I don't see how you can have a first-class multidisciplinary national facility like Los Alamos unless you have a strong team of people who are at the frontier of fundamental science. Without that you're not plugged in to the real intellectual life of the scientific community. At the same time, you cannot justify a large multihundred-million-dollar facility of the sort that we have without addressing some of the genuine technological issues of the country. This organism is not necessarily a freak invented by happenstance. It could be one of the strongest scientific organizations as we move out of this century. The universities may never recover from the demographics and the sociology of our culture, and our culture is not going to be able to subsidize the research activities of the professors on the basis of their teaching activities. I expect that there will be an increase in institutions, but not to the extent that the Soviets have institutionalized their whole scientific establishment, with very little first-class research done in universities. I think that the national laboratories, or variants thereof, may be on the rise.

SIMMONS: Why did you choose a career in physics?

CARRUTHERS: When I was young I became interested in birds and fishing, all of the outdoors and the creatures in it. I became locally famous as a promising young biologist. It was claimed—though falsely—that I could identify a bird by listening to it walk on a branch. I became a Boy Scout nature counsellor, and I was the court of appeal for many merit badges on biological subjects. After a while I began to realize that biology was very hard. I was looking at the stars, became interested in physics, and read books, some of the old-fashioned inspiring books on relativity and quantum mechanics. I also got books which were way beyond me. As a high school junior I couldn't understand Riemannian calculus and its implications for relativity, but I stared lovingly at the equations. I was very excited by these books, and at the same time I felt that biology was farther from the fundamental essence of exact science, although I wouldn't agree with that point of view any more. I had decided by the time I was 16 that I would become a physicist. In the meantime, I was playing a lot of music, but I was not in a serious musical community, and there weren't any influences on me which would have led me into that as a career.

WEST: You were pretty much self-motivated?

CARRUTHERS: In high school I was regarded as a freak. I was the smart, fat kid with the violin, the honorary Jew of Middletown, Ohio. When I first went to college I came under the spell of Bertrand Russell. I read all of his atheistic works and joined his cult. When you're 19 that can be just right. I even founded a Philosophy Club at Carnegie Tech. We met and seriously discussed all the major philosophical questions, no doubt in some naive way.

SIMMONS: Did you read Whitehead?

CARRUTHERS: Oh, yes, he was a very dull fellow, but I felt obliged to read him because he was considered so important—just like lots of contemporary theories.

WEST: What kind of an education did you have at Carnegie?

CARRUTHERS: During the first two years I found physics not very interesting, but I knew that the good stuff was just beyond. I spent a lot of time reading broadly. I soon became intoxicated with just learning. In my sophomore year, at one time I was taking 11 courses. I tried to learn everything I could.

At the end of my sophomore year I had reached the right level to really be interested in physics. I remember being inspired by Max Born's book on atomic physics. There is a chapter on wave-corpuscles or wave-particle duality, and the blinding insight of that remains with me. It was much more excit-

"Bethe is a hero, and a father figure to me, an intellectual father figure."

INTERVIEW

ing than the austere beauties of relativity, which seem to me essentially kinematical, but don't have the awe-inspiring dynamical content of quantum mechanics.

When I first went to Cornell as a graduate student I was taking quantum field theory without having ever learned scattering theory—a slightly perverse way to do it. The instructor and my classmates seemed to have no interest in the pathological diseases which make that theory not a theory. I assumed that I was slow, because they all nodded their heads simultaneously in phase whenever some obscure mystery went past at the speed of light.

I was very much interested in statistical mechanics and the many-body problem and was turned off initially by the field theory problems. So I started doing research, and as soon as I learned Fermi's "Golden Rule," I was writing papers in solid-state physics. However, it happened that Feynman was taking a sabbatical at Cornell, and I found to my amazement that in his lectures he laughed at all the absurdities of field theory. It came as a liberating influence to find out that the things that bothered me were exactly the things that this great man considered absurd and which had to be removed from the correct theory eventually.

WEST: What are you referring to?

CARRUTHERS: The entire structure of infinities and mathematical sicknesses in an otherwise beautiful structure. I began thinking a bit about particle physics problems. I was so far advanced in writing solid-state research papers that I could have graduated after maybe a year and a half of graduate school, but I did not want to be identified as some particular kind of theorist. I wanted to be a general theorist, and so I decided I would do a thesis in particle physics. I screwed up my courage and went to visit Bethe, which is a very scary thing for a student. He seemed to have heard of me and took me on. Then, of course, he went back to Geneva for some indefinite time for disarmament talks.

I came in very early in this period and saw him studying a *Physical Review Letters* article having to do with the one-pion exchange mechanism of producing an extra pion off a nucleon. It was written by Charlie Goeble. He said, "Well, Charlie Goeble is a very smart fellow, and I'm not sure that this is exactly right, but there is something very interesting here and why don't you look at it." So, I looked at it and decided that what was missing was the interaction of one of the final pions with the nucleon. I wrote a thesis on this subject.

After three years at Cornell, I was finished, got an NSF postdoctoral fellowship, and settled in to Cornell. Then I was summoned to the Chairman's office and told that I was going to be an Assistant Professor, teaching quantum mechanics to the first-year graduate students. I said I wasn't ready to teach quantum mechanics, that I wasn't even ready to be a professor. I asked them to leave me alone and let me have my postdoc and do my research. (Of course, this situation is inconceivable nowadays.) But I joined the Cornell Physics Department. I piled my desk high with learned books on quantum mechanics, studying all of the old puzzles that quantum mechanicians like to ponder.

SIMMONS: This was about 1960?

CARRUTHERS: January of 1961. For several years I tried to keep up in both solid-state and high-energy physics but with teaching and graduate students and so on, I didn't have the stamina to do the research in two major fields. I chose to stay in particle physics and gradually stopped the other. However, that way of thinking about many-body systems has always influenced my particle physics. It's the way of thinking that has now taken over almost all of field theory and modern particle physics. So I've never regretted that experience.

WEST: Did you enjoy teaching?

CARRUTHERS: Oh, yes. That's the main thing I miss, being at Los Alamos, the students and their frisky ways. I remember being frustrated giving exams because the students always averaged 37%. I would decide to make the test easier the next time. It didn't work. I was learning at the same rate that I was making the tests easier. At one time I got a book on how to teach. I never read it, of course, and the students continued to get 37% on all exams.

WEST: Were you a good teacher?

CARRUTHERS: Well, you'll have to ask Mike Simmons, who only listened to the smutty remarks, I think. I thought I was well organized; I don't know if I was a good teacher.

SIMMONS: Did you do much preparation for your classes?

CARRUTHERS: I felt naked and defenseless if I didn't come with a complete army of notes with all derivations intact, though as you well know, there were times when I might improvise after an all-night graduate party. There were some subjects in which I felt so totally in control that I didn't need any preparation, but that was based on some early mastery of the subject.

WEST: Did you enjoy having graduate students?

CARRUTHERS: I especially enjoy *working* with graduate students. They keep you moving. I prefer *teaching* undergraduates. Graduate students are so much in awe of the professor that they don't really give him a hard enough time, whereas the undergraduate sits there and says, "Who are you? Why should I believe that? That doesn't sound very convincing to me." What are you going to do when a student tells you that?

WEST: What did you do?

CARRUTHERS: Well, I brought beer to the final exam. Along with Bob Dylan quotations to entitle each question.

SIMMONS: What was it like being Bethe's student?

CARRUTHERS: It was quite peaceful because he was spending a lot of his time in Geneva—disarming. These were the first serious disarmament talks, as I recall.

WEST: Do you see that, in any way, as a connection between what has become an interest of yours here at Los Alamos?

CARRUTHERS: It might be, but at that time it must have been very well suppressed because I was completely uninterested in politics. I couldn't imagine why he spent all his time going around the world dealing with what I considered insoluble problems. I valued him for his insight into physics, and I was sorry he wasn't around more.

SIMMONS: Did he have any particular influence on your style of research?

CARRUTHERS: He had a very strong influence on my general standards. Bethe is a hero, and a father figure to me, an intellectual father figure. After you're around him awhile you realize that it's got to be right, and there isn't any other option. It's not half right, or almost right; it really has to be right. That's a lesson that is not very much in vogue these days. Also he was usually uninterested in very abstruse theories and I think that, for a while, had an impact on my own orientation in theory, though I've drifted to becoming more abstract as the years go on.

WEST: You said that at the very beginning of your graduate career you chose particle physics because you wanted not to be identified as a specialist. Was that in any way related to Bethe?

CARRUTHERS: No, that was pre-Bethe. I saw that people got categorized and put in boxes and I didn't want to be one of those people.

SIMMONS: Let me ask you about a different kind of thing, something that I call the "Feigenbaum effect." There is a typical progression for a talented young person in science: from his Ph.D. to the postdoctoral appointment, during the first year of which he publishes a half a dozen papers, thereby acquiring his second appointment; during the first year of that, another half a dozen, thereby earning a permanent position. Some people violate all of that. Are there lots of them or only a few? Is there something wrong with this standard way of handling young scientists?

CARRUTHERS: Yes. There are quite a few sensitive people who can't stand the strain of competition with people equally good, or who have bad luck, don't have stamina, or have personal problems. Some of them are lost. This happens in every field. You lose people everywhere. I think you lose fewer people in physics than you do in most highly intellectual activities. But the "Feigenbaum effect" is real. Mitchell Feigenbaum did little to preserve his career by publication. One of the best things I have done here was to pull somebody off the street whose career was endangered, but whose talent was unmistakably superior to almost anybody else that I had ever met of that age group. This is high-risk investment, and it rarely happens at a university.

Five years later, Feigenbaum has produced major advances* in our understanding of turbulence, one of the outstanding scientific puzzles of our time. In the usual institutional context, if you were looking for somebody to study turbulence, you would advertise for an expert in fluid mechanics, preferably with computer skills. However, decisive advances usually come when talented people take a fresh look at a subject. The practice (at Los Alamos and elsewhere) of hiring "already qualified" people for specific tasks, often under the pressure of programmatic deadlines, is harmful to the long-term quality of science.

But the quality survives miraculously, despite all the human foibles that are translated into the way science is done. That's largely due to the experimentalists, I suppose. Somehow science is self-correcting. Even though credit often is assigned unfairly, the actual evolution goes on, you sort out the better ideas from the junk, and occasionally there are major insights.

SIMMONS: Let's talk about something different. The importance of music to you dates back to a very early age.

CARRUTHERS: Nine years old. I suppressed it pretty much during college and immediately after. The desire to do my own research overwhelmed every-

*See "Universal Behavior in Nonlinear Systems," Los Alamos Science, No. 1 (Summer 1980).

"There's nothing which quite combines the sensuous with the intellectual as really first-rate music. It's a daily part of my existence..."

INTERVIEW

thing else in my life. Then I began to come back to it.

When I was about eight, I heard Jascha Heifitz play the Beethoven violin concerto on the radio, and I said, "I want to do that." So my mother arranged for me to get a violin and a teacher.

There's nothing which quite combines the sensuous with the intellectual as really first-rate music. It's a daily part of my existence as you can tell by the dirty, gnarled fingertips.

SIMMONS: Did you ever have second thoughts about not following music as a career?

CARRUTHERS: There's no doubt about that. What I had secretly wanted to do for a long time, and thought was the finest creative act, was to write first-rate classical music. Of course, you may not know that you've written first-rate music until much too late, but that seems to me to be one of the finest things.

WEST: What do you see as the relationship between the experience of physics and the creative process in physics and that of music?

CARRUTHERS: I'm just old enough now that everything seems to be merging. If you're working on a Bach sonata, you may have to play a hundred times through a few lines before it really begins to work. If you're doing physics you may have to bang your head on the wall for a couple of weeks before you finally see the way through to doing it right. There is a lot of hard work involved, but when it's all in place you have great satisfaction.

I can't account for the intellectual component of music, but I feel it extremely deeply. The structure and the counterpoint, the development of the themes, are rooted in our mental structure in the same way, or related to the way, that I feel the texture of mathematical equations which express physical laws. There is a beauty to that.

WEST: Do you see or feel that as a spiritual experience?

CARRUTHERS: Both things are spiritual, but there is the question about truth. In science there is finally an experiment. You may have thought that your equations were extremely beautiful but, in fact, they are likely to be wrong. However beautiful, you have to throw them away.

If you've written a marvelous sonata, in what sense is it right or wrong? It may have internal consistency, structure, and beauty. There is always this extra constraint on scientific work, which at least superficially distinguishes science from the arts. The arts seem more free-wheeling and lacking of boundary conditions. In fact, the farther you are from the exact literal truth, the more insight you may give to the actual truth. You may value it more because it's so bizarre that it reflects onto the truth in a way which illuminates the truth, whereas with the physical theories, if they are wrong, they are out, and you are ruined. But in the actual doing of science, you don't proceed that way. You're optimistic, and you don't throw away ideas until you've had the fun of creating something.

SIMMONS: Do you see more of an analogy between art and mathematics?

CARRUTHERS: Yes, I certainly do. Mathematics can create its own structure of logic and beauty and doesn't have to face an experiment.

WEST: Changing the subject, who are some of the physicists that you most admire? The people outside physics? Whom do you see as wise men?

CARRUTHERS: I always especially admired Landau because of his universal scope and overwhelming intuition. It's easy to admire some of the historically great mathematicians. You can't imagine how they did what they did. Musicians, poets, writers, and so on too numerous to list. Montaigne's essays are splendid examples of wisdom. I often read one at bedtime. The cynical old bastard has thought of everything and is gifted with the best touch. I love Russian literature; I find wisdom in all this madness, as in Kafka.

SIMMONS: This is the second time that Russia has come up in the conversation, and I want to ask your opinion of U.S.-Soviet scientific relations. What is a proper and profitable posture for individual scientists in the U.S. and for U.S. institutions dealing with Soviet colleagues?

CARRUTHERS: I'm somewhat of a moderate on this question. As in the U.S., there are many kinds of people in the Soviet Union. There are those who have used the system to promote themselves, and there are many highly principled, brilliant people who somehow manage to create in this system. I feel that you have to encourage the latter and maintain contact with them. Simply turning your back on the scientists because the government is repressive in regard to human rights is probably a mistake. On the other hand, they're too outrageous in sending Sakharov off to live under house arrest in Gorki. Some official notice is necessary, not just for his sake but for the cause of these people in general. The cessation of official exchanges for a finite fixed time was a useful expression

of that, though the Soviet government does not seem to understand that this is felt by the majority of the American scientists. The August decision of the U.S. Academy to continue the ban on exchanges is destructive to the exchange program, however.

SIMMONS: A number of colleagues have said they cannot in good conscience go to meetings in the Soviet Union or to meetings where Soviet scientists were present. Is that dangerous?

CARRUTHERS: I think the self-righteousness is self-evident. It's very easy to be self-righteous in this world, and I don't see the point of that. It's very frustrating to realize that we have to work on a time scale which is so long that we may not live to see the outcome of it. By encouraging the moderate elements in the Soviet Union, perhaps in 50 years there will be noticeable change. I think we should do that. The fact that we are frustrated year after year and have all of these absurd and humiliating developments doesn't seem to me to be reason to give it up. I think a continuing, restrained and, above all, adult response to these problems is what we need.

SIMMONS: Acting as the representative of our Academy of Science, you've dealt with representatives of the Soviet Academy. Our academy has little impact on anything. Theirs runs almost everything in science. What's it like dealing with them in those circumstances? Do they realize the difference?

CARRUTHERS: My general impression is, they overestimate the political influence of the American academy. They find it hard to realize that it's not symmetrical. On the other hand, they don't really run everything, because there is the State Committee on Atomic Energy, which is somewhat like DOE, and there are, of course, the KGB and the Communist Party to complicate things. For example, dealing individually with Russian physicists, I found them extremely cooperative and helpful, but by the time an agreement filters through the system to an actual exchange program, so many people have put in their two-cents worth that the resemblance to the original agreement is hard to see. Their internal politics is even more complex than ours.

WEST: Do you feel there should be more institutional guidance in U. S. science and technological programs than there is at present?

CARRUTHERS: I certainly do. The National Academy of Sciences occupies the turf but doesn't do anything much except issue reports.

SIMMONS: What about the Physical Society?

CARRUTHERS: The Physical Society is an ineffective organization. In the American Physical Society you have, of course, only a part of science. The officers are donating their time. Most of them don't take their gloves off and go in there and fight. You have to be willing to go into the arena these days to advocate your cause. It's not a matter of gentlemen discussing the future of science over tea. There are too many powerful interests in the country who will use up all the resources with no attention to the long-range health of the society. You have to be willing to represent your case in the most powerful possible way. That means confrontation.

SIMMONS: Is there a way of using the Physical Society, of changing it so that it can be an advocate for physics?

"I can't account for the intellectual component of music, but I feel it extremely deeply. The structure and the counterpoint, the development of the themes are rooted in our mental structure in the same way or related to the way, that I feel the texture of mathematical equations which express physical laws. There is a beauty to that."

Physics, Philosophy, Leadership, & Policy

"The greatest virtue is to survive, living and leading the life that you consider productive and decent."

INTERVIEW

CARRUTHERS: It does advocate. The advocacy depends very much on who's President at any given moment. We have gentlemen Presidents who don't really work hard at it, and we have street fighters who get in there and stir things up.

SIMMONS: Street fighters?

CARRUTHERS: Well, you do it in hallways and corridors. It's not really street fighting—it's corridor fighting.

SIMMONS: What are the gut issues?

CARRUTHERS: Unfortunately, money is the bottom line. There is the partitioning of the money that science gets, and there is tremendous confusion about basic versus applied science. In the last few years, the Carter regime has been proudly announcing giant increases for basic science. For example, the National Science Foundation, a case with which I am familiar. Every time I look at one of those budgets closely I find that the increases are not at all in basic science. They are in engineering or something else, educational programs or things like that, and at the end of the year basic sciences don't even get inflation. It's been true for several years.

WEST: What effect will this have?

CARRUTHERS: In the case of the Science Foundation, the deterioration of the university research base has been extremely serious: cuts in the number and size of grants to university professors, cuts in the number of postdocs and graduate students supported by these funds, dilapidated equipment, the U.S. falling behind competitively with respect to other countries in certain areas. Unfortunately, a similar trend is going on in the Department of Energy, which supports the national labs. The high-energy physics budget is quite a ways down, and the Europeans are taking a very competitive run on our preeminence in that field.

My general attitude towards doing science is that if we don't try to be best we're going to be second-rate. We have to try to be best. We can address the case of basic science, and I mention high-energy physics, for example, where we've dominated in almost every sense for 20 to 30 years, partly due to the rejuvenation of American physics during the Second World War by European immigrants. Even on our native strength, it has been an extremely powerful showing. The obvious excitement of the subject has penetrated every part of the culture, and the very best young people are thereby attracted into science. First of all, it might be the most glamorous things of black holes, neutrinos and quarks and whatever—a bias toward science and technology attracts people who in earlier days might have done something different. As soon as we lose that glamorous image, it seems unlikely that the very best people will go in for these fields. It's extremely important to maintain an image of being first-class as a country in as many areas of basic science as we can. We can't maintain our position in the world unless a special effort is made to get the very best people into these areas. So we have to guide them towards working in science and technology in order to maintain the political, military and economic strength of the country in the face of a very hungry and aggressive outside world. People have to be in a high state of excitement to maintain any excellence.

SIMMONS: A lot of our colleagues see some immorality in weapons work or association with institutions that engage in national defense work. Do you have anything to say on this issue?

CARRUTHERS: The greatest virtue is to survive, living and leading the life that you consider productive and decent.

WEST: Do you feel that scientists have a responsibility to work on problems of national security and defense? Would you say, for example, that if we're to support SLAC or Fermilab, the people there should spend some of their time thinking about these problems.

CARRUTHERS: I don't think you can apply a formula to these people. They are all quite different. I do think there is a responsibility, and adopting a holy attitude towards it is naive. But if you are bored or hate that kind of work, you shouldn't be made to do it. I think scientific activities can be defended on an intrinsic basis. There is a need for intelligent people to interact with the defense community, because that is a closed society where people speak a special language and often arrive at very peculiar conclusions.

SIMMONS: Do you disagree with advocates of phased disarmament who argue that an increase in armaments would be stabilizing in many instances?

CARRUTHERS: I know something about the arms race, and it's appalling. It's increasing all the time and it's very frightening. After a summer at JASON I get very depressed. I've heard all the generals and colonels, and I've heard what the rampaging technology can do next year that it couldn't do 10 years ago.

WEST: Do you believe in a kind of nuclear stalemate theory?

CARRUTHERS: I have believed in nuclear stalemate theory, but only between the U.S. and the Soviet Union. As soon

as every chief in the world has his own bomb, you can expect nuclear war. I think there are bad times ahead.

WEST: Within that context, what do you think of the morality of someone working on defense problems? What kind of stance should that person take?

CARRUTHERS: I've been accused of lending my own—as I hesitate to say—prestige to it. I think it's also immoral to turn your back on it and say that you're too fine to be involved in even discussing it.

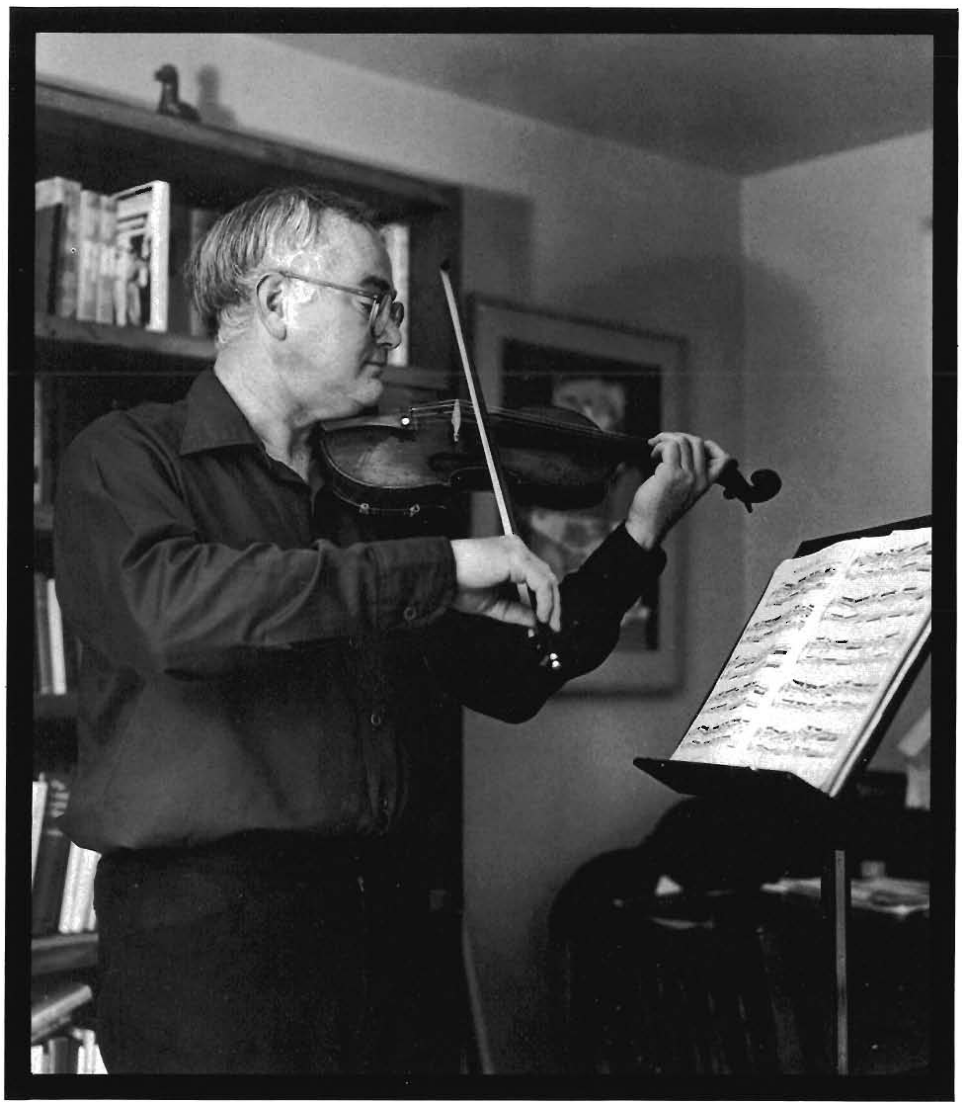
SIMMONS: Should someone whose morality is too fine to work on defense problems be working on disarmament problems?

CARRUTHERS: If they're good enough in science, let them do whatever they want to do. They don't have to be intelligent in the ways of the world to advance the cause of knowledge.

WEST: What are your reactions to people who are highly critical of nuclear energy?

CARRUTHERS: I think they haven't thought through the awful alternatives. I consider it the least of the evils. The real evil is that the planet is overpopulated, and there's no sign of change in the pressures from that direction. Nobody has ever had the guts to face that evil. I hope we don't solve it by a nuclear war. Coal is much more dangerous than nuclear power. Slavery to OPEC is much more dangerous than nuclear power. What else is there?

Of course, the problems are very tough, and science may be silly to dash forth with a quick answer. On the other hand, what's happened in the political process is a paralysis in almost every sensible proposal. We have to live with uncertainty, and we might as well live on a ten-year time span and not say, "Well, in a hundred years perhaps the salt mine will crack and there will be some horrible leak and my grandchildren will be rendered sterile or some other worse thing." There is a kind of self-righteousness in that kind of attitude which can be very counterproductive.



WEST: One of the reactions in the political sphere is that scientists have been self-righteous.

CARRUTHERS: It's almost required. You must have the self-confidence to carry you through the hostilities and criticisms. It may look like arrogance even if it's not.

WEST: Do you see that as a serious problem?

CARRUTHERS: Well, I don't know why the public trusts Walter Cronkite and not the scientists ■