NUMBER 12 SPRING/SUMMER 1985

Los Alamos Science

LOS ALAMOS NATIONAL LABORATORY

In this issue we focus on some mainline interests of the Laboratory. The laser fusion program at Los Alamos is undergoing a change in direction—from work with long-wavelength $C0_2$ lasers at the Helios and Antares facilities—to work with short-wavelength lasers. Los Alamos originally chose the $C0_2$ laser because of its long-held status as the world's most efficient and economical laser. The philosophy was that it could be easier to design a small fusion target to accommodate the CO_2 laser than to invent a new laser.

A ten-year effort was made to uncover the mechanisms by which the energy of $C0_2$ laser light is deposited in a fusion target and to design targets that would direct those mechanisms toward the goal of efficient implosion. The story told by Dave Forslund and Phil Goldstone covers a mind-boggling array of new phenomena and new ideas—all discovered in the heat of programmatic promises and deadlines. We follow their path back and forth between hope and disappointment as they advance through the uncharted territory of laser-plasma interactions.

For years the project was banking on the creation of a steep density gradient in the absorbing plasma as the way to exploit long-wavelength laser light. The steep gradient enhances the collective process called resonant absorption. Unknown was how this process scales with energy. Does the absorption efficiency rise or fall as the laser intensity is increased by factors of 10 or more? The discovery that the steep density gradient is a short-lived phenomenon—one that decays through new types of plasma instabilities—came as a surprise and a devastating blow to the plans for $C0_2$-laser-driven fusion.

The blow to programmatic goals was tempered by the discovery of truly exotic processes. High-intensity laser-plasma interactions create enormous magnetic fields and produce rapid particle acceleration in ways that have never before been seen on earth. These processes offer possibilities for novel particle accelerators, for an approach to fusion that combines magnetic and laser fusion concepts, for studying the effects of new particle beam weapons, and for exploring dynamical mechanisms that might produce extragalactic radio jets.

Will we be able to use Antares to advance this new realm of plasma physics, or will this fabulous, space-age facility be abandoned like an old Hollywood movie set—standing idle because fusion monies won't be available to support it? How do we allow a major state-of-the-art facility to be planned and built with only a limited vision of what might be discovered there? We expect new findings, but we face them unprepared, lacking the flexibility to change directions and follow the new avenues of research. One could imagine such changes evolving naturally; instead they meet with endless reviews, endless requirements for justification, and the threat of imminent shutdown.

But people in this business are used to all that. Antares is just one among many continuing examples of the agonizing inefficiency brought about by the short funding cycles and the lack of long-range vision in American science.

What will happen next in laser fusion research at Los Alamos? First a concerted effort is being made to find adequate funding to continue work at Antares in high-energy plasma physics. Meanwhile, the hope for laser fusion is being placed on the development of a new short-wavelength laser driver—the KrF laser. It was always known that short-wavelength laser light would be more readily absorbed by a target than long-wavelength light. The problem was to develop a short-wavelength laser that was both powerful and efficient. Livermore invested in a high-intensity glass laser, but this laser has a low efficiency and places severe demands on the optical system. The KrF laser, discussed in this issue by program manager Reed Jensen, is a gas laser, which, though not inherently efficient, has been made so through a novel

multiplexing system. At this moment spirits are high because the new design is producing more power than anticipated and is ahead of schedule. If all goes well, it will be ready for short-wavelength fusion experiments in just eighteen months.

The article by Mike Norman and Karl-Heinz Winkler on supersonic jets presents through numerical simulation the first self-consistent model for extragalactic radio jets. Exciting to the astrophysics community because it provides a means for understanding the structure and stability of those enormous cosmic eruptions, this work also provides a beautiful example of how to do basic science with a supercomputer. The computer is used as a numerical laboratory to investigate the phenomenology of a nonlinear system, and the essence of the physics is extracted through graphic representations of the solution. Mike and Karl-Heinz "clobbered the problem" with an unprecedented amount of computer power, orders of magnitude greater than that applied to any other problem in astrophysics and comparable to the amount usually devoted to studies of controlled fusion or nuclear weapon designs. They pushed their research tools to the limit, using all the memory of a Cray-1 computer and as much computer time as they could beg, borrow, or steal. The example they have set is sure to be followed by many as supercomputers become available to a wider and wider segment of the research community.

Reversing their usual competitive stance, Los Alamos and Livermore have joined forces in a most unusual project that is already advancing the rate of progress in medical genetics. Together they have become suppliers of human genes. That's right. You can order a vial of DNA fragments from the human chromosome of your choice. What can be done with such a vial? In "Genes by Mail," Livermore project leader Marv Van Dilla, Los Alamos project leader Larry Deaven, and other members of the Los Alamos team discuss how these so-called gene libraries will be used to aid basic research on gene structure and expression and to develop probes for diagnosing genetic diseases, for locating genes on chromosomes, and for detecting mutations and studying their role in adaptation and evolution. They also discuss gene therapy, an exciting prospect of research in medical genetics.

This issue ends with a philosopher's challenge—a challenge we may be tempted to ignore, especially those of us who believe that who we are and how we perceive is explainable in terms of the causal objectivity usually applied to the physical world. Upon closer examination, however, the strict dichotomy between objectivity and subjectivity breaks down. The role of intention is seen to underlie not just our actions and thoughts, but every act of perception as well. And intentionality implies the element of choice, another complication we might like to ignore. But ignore it we can't, because in our attempts to understand ourselves in scientific terms and to build machines that imitate human intelligence and human behavior, we are brought face to face with our very limited understanding of the nature of conscious acts. Gian-Carlo Rota and David Sharp, in their dialogue on mathematics, philosophy, and artificial intelligence, point to this confrontation as the central drama of our time, one that will bring about a major revolution in our understanding of ourselves.

This issue contains some warnings and some prophetic visions of the future. One is the revolution in research that is now at our doorstep, brought about by the enormous power of computers to unravel complex phenomena. Another is the revolution in medical genetics that promises a radically new approach to diagnosis and care. Finally, and most profound of all, there is the intellectual revolution, the revolution in awareness that will come as we attempt to use computers to mimic human function. ■

Necia Grant Cooper

Los Alamos Science
NUMBER 12 SPRING/SUMMER 1985

CONTENTS

RESEARCH AND REVIEW

Photon Impact—High-Energy Plasma Physics with CO_2 Lasers 2
by David W. Forslund and Phillip D. Goldstone

The Tools of High-Energy Laser-Plasma Physics 6

KrF Laser—The Advance toward Shorter Wavelengths 28
by Reed J. Jensen

Supersonic Jets 38
by Michael L. Norman and Karl-Heinz A.Winkler

Sidebar 1: Shock Waves versus Sound Waves 42

Sidebar 2: Shock Waves, Rarefactions, and Contact Discontinuities 44

Sidebar 3: Steady-State Jets 46

Sidebar 4: Pinch Instabilities and the Bernoulli Effect 61

Genes by Mail 72
An interview with L. Scott Cram, Larry L. Deaven, Carl E. Hildebrand, Robert K. Moyzis, and Marvin Van Dilla

What Is a Chromosome? 74

Making the Libraries
I. Supplying the Chromosomes 80
II. Sorting the Chromosomes 83
III. Cloning the DNA 86

PEOPLE

Mathematics, Philosophy, and Artificial Intelligence 92
A dialogue with Gian-Carlo Rota and David Sharp

On the cover (and overleaf).
The ion jet (white column) is a dramatic example of the exotic physics generated when high-intensity CO_2 laser light (yellow column) impacts the surface of a laser fusion target (sphere). The colors of the target represent the extreme magnetic fields that encircle the spot and help drive the highly collimated plume of ions from the surface.

Editor
Necia Grant Cooper

Art Director
Gloria Sharp

Science Writer
Roger Eckhardt

Science Editor
Nancy Shera

Assistant Editors
Maria Hughes
Dixie McDonald
Mick Scheib

Production and Layout
Sue von der Linden

Illustrators
Jim Cruz, Don DeGasperi,
Lenny Martinez, Mary Stovall,
Sue von der Linden

Photography
LeRoy N. Sanchez

Photo Laboratory Work
Ernie Burciaga, Debbie Fisher
Jerry Leyba, Chris Lindberg, Ken Lujan
Mark Martinez, Dan Morse,
Brian O'Hare

Phototypesetting
Samia Davis, Joyce Martinez,
Kris Mathieson, Chris West

Circulation
Dixie McDonald

Printing
Jim E. Lovato

Los Alamos Science. is published by 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.

Address Mail to
Los Alamos Science
Mail Stop M708
Los Alamos National Laboratory
Los Alamos, New Mexico 87545

PHOTON IMPACT

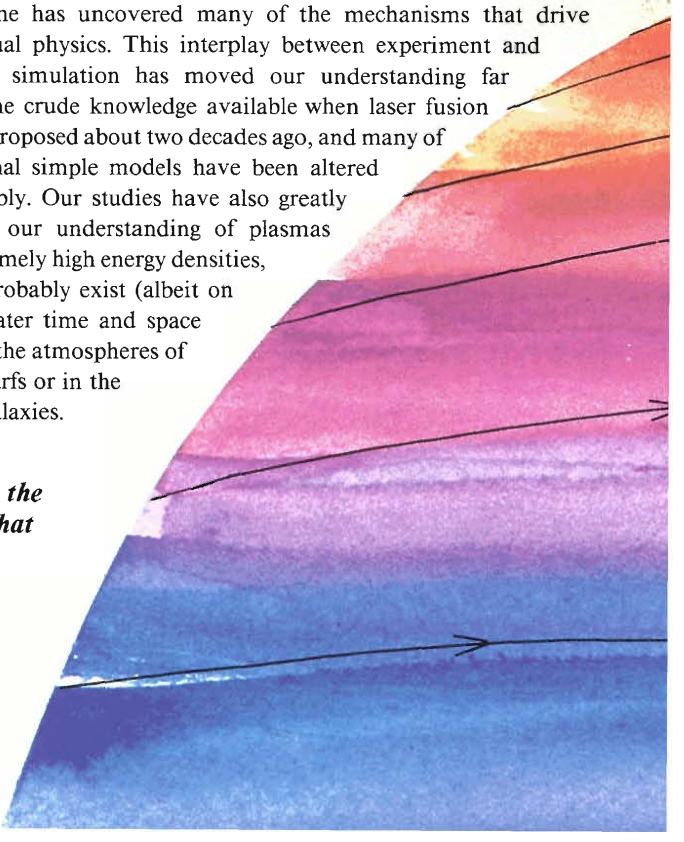
high-energy plasma physics with CO_2 lasers

by David W. Forslund and Philip D. Goldstone

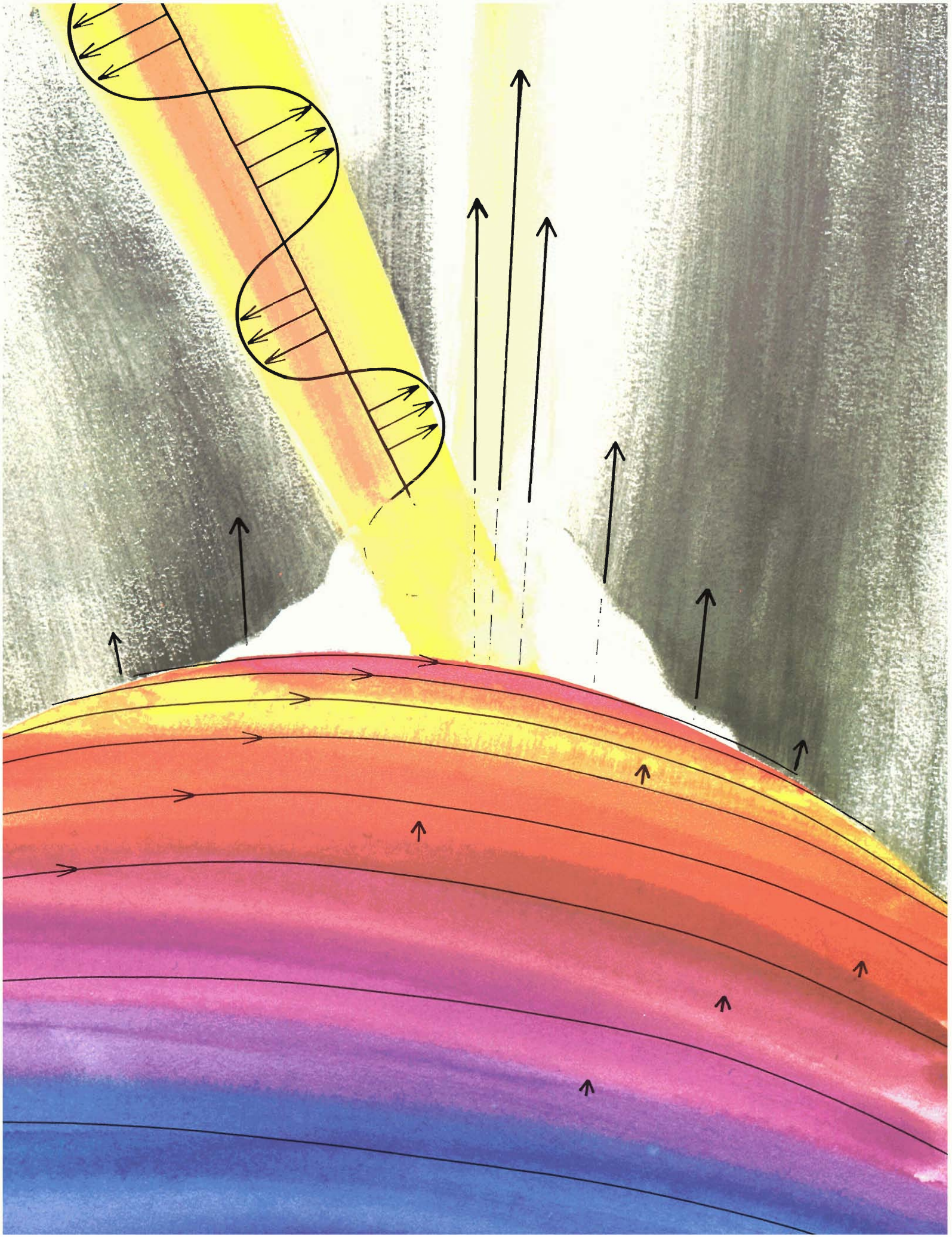
Research on controlled fusion attempts to mimic in the laboratory the thermonuclear burn of stars. At Los Alamos, Antares, a giant CO_2 laser, has been dedicated to the hope that eventually small amounts of matter can be imploded to densities and pressures exceeding those in the center of a star by using the power of high-intensity light. Because the physics of laser fusion involves an enormous range of uncharted phenomena, our experiments and theoretical simulations have produced many surprises. For example, in some experiments highly collimated ion jets are seen streaking from the laser target, reminiscent of the huge jets that sometimes streak for millions of parsecs from the centers of distant galaxies. Perhaps we have inadvertently imitated exotic features of the heavens other than thermonuclear burn.

We have also learned that laser fusion will not likely be achieved with CO_2 lasers. Antares and its lower-energy predecessor, Helios, were built to take advantage of the reliability, high repetition rate, and high efficiency of the CO_2 laser—features that will be needed for large-scale energy production. With the 10-micrometer (μm) light from these lasers, we have been able to explore pertinent physics carefully, learning, for example, how that physics depends on the intensity and wavelength of the light. Among other things, these experiments have demonstrated emphatically that if laser fusion is to be achieved, wavelengths much shorter than 10 μm are needed. Currently, the laser fusion program at Los Alamos is exploring possibilities for new types of lasers, including the KrF laser, a promising, highly efficient gas laser with 0.25-μm light (see "KrF Laser").

The theoretical effort at Los Alamos that parallels the experimental one has uncovered many of the mechanisms that drive the unusual physics. This interplay between experiment and computer simulation has moved our understanding far beyond the crude knowledge available when laser fusion was first proposed about two decades ago, and many of our original simple models have been altered considerably. Our studies have also greatly enhanced our understanding of plasmas with extremely high energy densities, such as probably exist (albeit on much greater time and space scales) in the atmospheres of white dwarfs or in the cores of galaxies.



the main actors in the drama of high-energy plasma physics are the intense electromagnetic field of the CO_2 laser, the strong magnetic fields that may encircle the laser spot, and the ions driven from the target surface, sometimes, as depicted here, in the form of a highly collimated jet.



This article, although noting the difficulties and successes of CO_2 laser fusion, concentrates rather on the journey of exploration into a unique state of energy and matter. Within the hot atmospheres of our miniature almost-suns, we have generated phenomena that cannot be easily observed in the laboratory by any other means. Being able to produce such exotica has provided an opportunity to increase our knowledge dramatically about the fourth state of matter. We describe here the current advances achieved because of this opportunity.

Why is the problem of laser fusion so complex? Using the energy of a laser beam to implode microscopic bits of matter requires precise control of a great variety of phenomena. The initial absorption of laser light at the target generates a hot plasma of electrons and ions, which can be likened to the corona surrounding a star. Additional laser light must penetrate this corona and be absorbed efficiently. The absorbed energy must then be transported to the inner layers of fuel in a manner that will produce the desired implosion. Energy that penetrates the target too rapidly preheats the fuel, a highly undesirable effect because it makes the implosion more difficult by greatly increasing the amount of energy needed to compress the fuel.

Vast changes in scale immensely complicate the research. We must deal with at least eight orders of magnitude in density, understanding the physics of matter from 10^{-5} gram per cubic centimeter (g/cm^3) in the tenuous corona where the laser light is absorbed to 10^3 g/cm^3 at the center of the imploding pellet. Distances range from 10^{-5} cm for charge separation in the plasma to 10^0 cm for the distance that hot ions may be ejected from the target during the laser pulse. In times as short as 10^{-15} second, the electromagnetic fields of the laser light and the plasma change appreciably, yet 10^{-8} second is needed to compress the pellet. Figure 1 shows the range of these three parameters for one computer simulation.

Vast changes extend to other parameters as well. Electrons in the cold pellet with initial energies of less than 0.1 electron volt (eV) may be accelerated in the corona to energies as high as 10^6 eV. If the target is illuminated uniformly and no instabilities develop in the plasma, there will be no strong magnetic fields. However, some of the strongest magnetic fields yet generated in the laboratory—as high as 10^4 teslas—may surround isolated laser spots.

While the physics of laser fusion discussed here emphasizes the nearly collisionless physics of very hot plasmas, high-density collisional physics at the core of the target must also be understood. We must include the effects of spontaneously generated magnetic fields, the atomic physics of weakly to highly ionized materials, and the effects of these ions and fields on the transport of particles and photons of all energies. Frequently the processes are far from equilibrium, requiring elaborate rate equations for a reasonably accurate description.

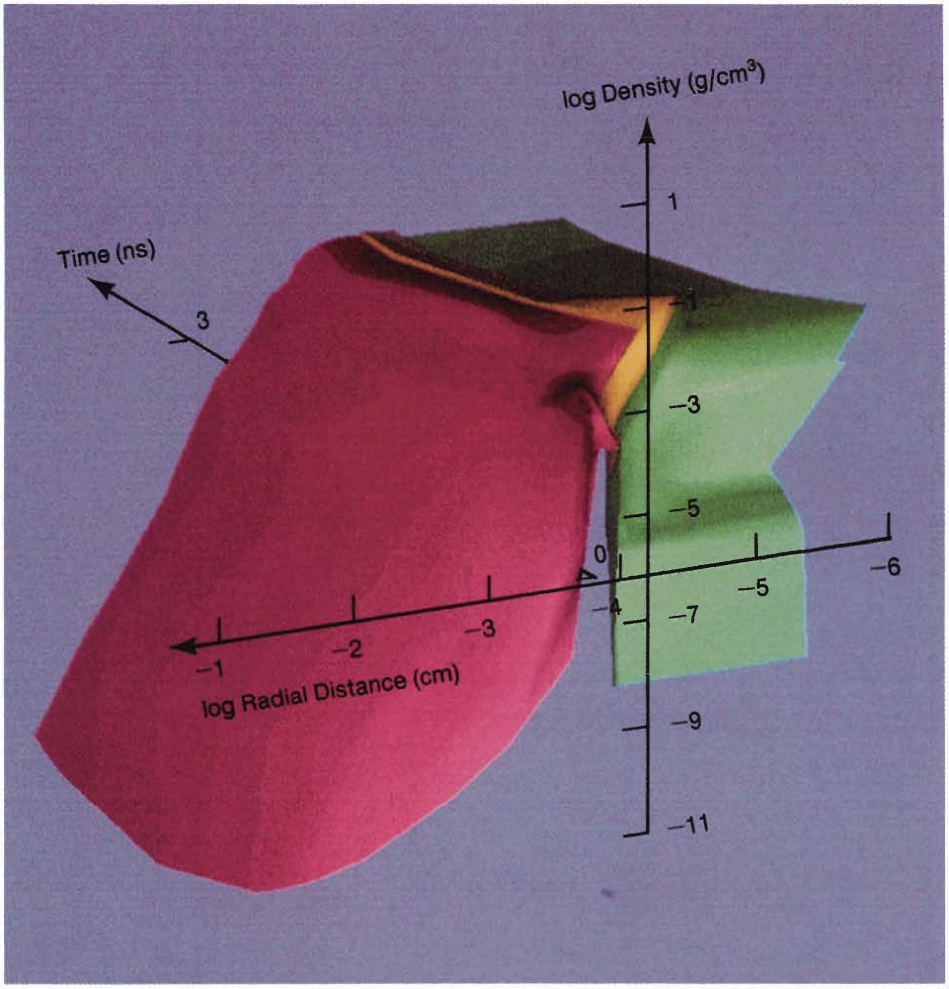
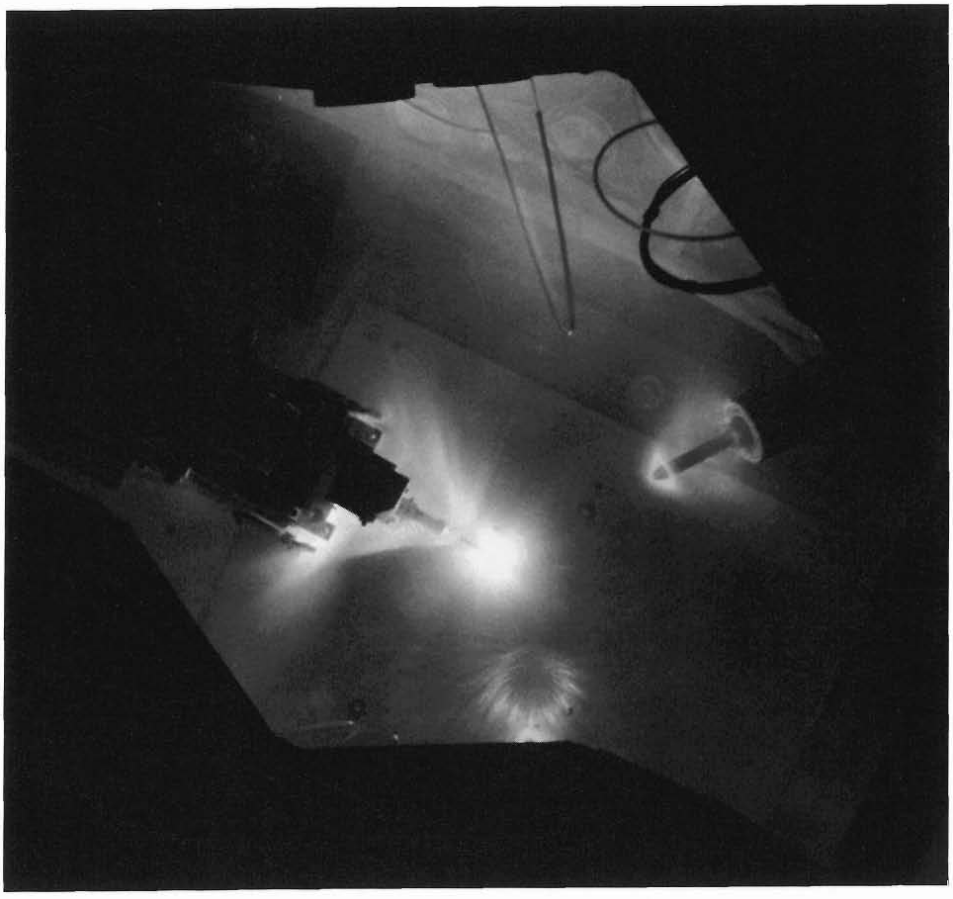


Fig. 1. Results of a computer simulation of laser fusion with the LASNEX code. The colored surfaces represent the three layers of an imploding fuel pellet as a function of density, radial distance from the pellet center, and time. The red outer layer is plastic; the yellow middle layer is a thin coating of glass that provides structural integrity; and the green inner layer is the fuel. The scales on the axes indicate the vast parameter ranges needed to adequately describe laser fusion. Both the density scale and the radial-distance scale are logarithmic; only the time scale is linear. The surfaces show the outer layer of plastic quickly blowing outward and toward lower densities to form the surrounding corona of plasma, while the imploding fuel and glass shell, at first, become more dense. The kink at high densities in the yellow and green surfaces shows the plastic shell and some of the fuel bouncing outward after the initial implosion. *(This figure was drawn with the Grafic program developed by Mel Prueitt.)*

Collective Effects. One of the biggest problems for laser fusion has been how to deposit light energy in the target efficiently. The very

properties of lasers that allow production of the extreme high power and intensity needed for fusion can work against a desirable form of energy deposition. In particular, the wave nature, the high coherence, and the narrow bandwidth of the light all contribute to peculiar collective effects in the deposition that reduce the efficiency. It is the highly organized properties of the laser light that tend to drive the hot plasma in the target far from thermodynamic equilibrium, a state with potentially serious consequences. Much of the history of laser fusion is a record of attempts to predict, measure, and avoid the deleterious effects of the coherent properties of the light.

Collective effects are especially dominant at the CO_2-laser wavelength of 10 μm, whereas, at shorter wavelengths, plasma-light interactions tend to be more collisional, and collective effects are less important. Specifically, the collective effects of CO_2 lasers produce *hot electrons* in the corona surrounding the target. These electrons have very high kinetic energies and penetrate the target material quickly. As already pointed out, energy in this form is not favorable for laser fusion.



A fusion target bathed in light and plasma during a shot with about 20 kilojoules of energy from Antares. The target is mounted at the end of an insertion mechanism that projects from the left; the object to the right is an x-ray pinhole camera; the circular pattern toward the bottom is a reflection. *(Photo by Fred Rick.)*

In the early years of research on CO_2 laser fusion, it was thought that the absorption necessary for implosion could only be achieved at high laser intensities if collective absorption processes played a significant role. Simple estimates implied that the temperature, or average kinetic energy, of the hot electrons would be high enough to be potentially harmful. When experiments showed a lower hot-electron temperature than estimated, optimism for long-wavelength lasers increased. Computer simulation models confirmed and explained the experiments. Unfortunately, when more intense lasers and more sophisticated experimental instruments became available, measurements at the higher intensities showed a new, dramatic increase in the hot-electron temperature. Computer simulations at the new intensitites, using more powerful computers not available earlier, again agreed with the experiments.

These latest results have greatly diminished our confidence in the possibility of *long-wavelength* laser fusion. However, because hot electrons represent an extreme non-Maxwellian distribution of particles, these experiments are able to generate a variety of interesting and complex plasma phenomena. Here we discuss these phenomena and outline the mechanisms that appear to be important in the interaction of CO_2 laser light with plasmas.

Laser Light Absorption and Transport

When laser light impinges on a target, the initial effects are absorption of the light and production of hot electrons. Then energy flows into and through a variety of energy-flow channels, including magnetic fields, acceleration of energetic ions, bremsstrahlung emission, microwave emission, and heating of a dense "thermal" plasma. Ultimately, it is the heating of the dense plasma and the resulting ablation of the target surface that produces the implosion by a rocket-like reaction.

Although our understanding of the later energy-flow processes has been largely empirical, much of our understanding of the initial absorption processes has been theoretical (see "The Tools of High-Energy Laser-Plasma Physics"). This is so because, at the 10-μm wavelength, there are only a limited number of unambiguous experimental signatures that define *specific* absorption processes rather than reflecting an overall effect. As a result, most of the information on absorption comes from computer simulations and analytic theory that have been iterated to reproduce macroscopic experimental data, such as the hot-electron energy spectrum or total absorption. As a result, the analysis has, in fact, been a difficult problem. Fortunately, we have been very successful in putting together a detailed picture of specific absorption mechanisms. What are the main elements of this picture?

The Critical Density. Initially, the laser light arrives at the target in a near vacuum with the intensity adjusted to achieve nearly uniform illumination. Such uniformity is difficult to maintain, however, because as the target heats up, hot plasma is blown off, and the propagation of light is altered by the changing density of the plasma. In fact, at a certain density called the *critical density* n_c, light does not penetrate further.

In more mathematical terms the index of refraction n of a plasma is given by

$$n^2 = 1 - \frac{\omega_p^2}{\omega^2}, \qquad (1)$$

where ω is the laser frequency and ω_p is the natural frequency of oscillation of electrons in the plasma. This latter frequency, called the local plasma frequency, is given by

$$\omega_p^2 = 4\pi \frac{n_e e^2}{m}, \qquad (2)$$

where n_e, e, and m are the electron density, charge, and mass, respectively.

continued on page 10

The Tools of High-Energy Laser-Plasma Physics

The laser fusion program has considerable strength in both theory and experiment. The complexity of the physics has kept the program at the forefront in developing new theoretical tools, experimental techniques, and diagnostics. For example, it is experimentally taxing to measure phenomena on a time scale of picoseconds when the entire experiment occurs in about a nanosecond and to resolve phenomena spatially at the 10-μm wavelength of CO_2 light when targets may be as large as millimeters.

In addition, many of the phenomena on the submicrometer, subpicosecond scale—the truly microscopic—strongly influence the macroscopic behavior but can't be readily measured. Computer simulation is then heavily relied on to couple microscopic phenomena to macroscopic observables.

Theoretical Tools

Three codes—WAVE, VENUS, and LASNEX—are used to simulate laser fusion phenomena. As depicted in the figure, each code simulates phenomena occurring within different ranges of density, time, and space.

WAVE. The code that deals with the most microscopic aspects of energy absorption and transport in the target is WAVE, a 2-dimensional particle simulation code. WAVE is an explicit code that does a first-principles calculation; that is, it solves, in a self-consistent manner, Maxwell's equations and the relativistic Newton's laws for particles in 3-component electric and magnetic fields. WAVE typically advances 10^6 particles on a grid of 10^5 cells for 10^4 time steps.

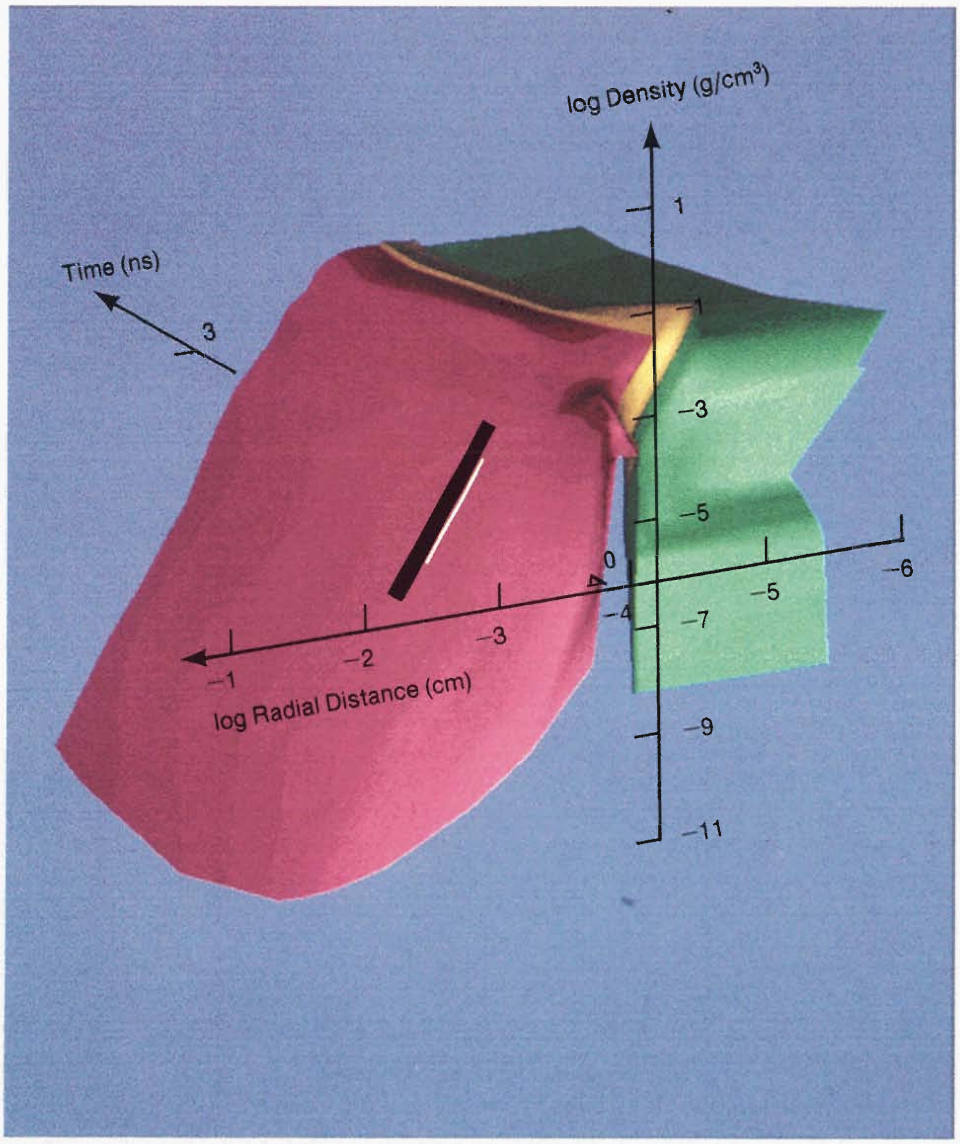
Only a portion of the absorption problem can be modeled by WAVE because the time step must be limited to a small fraction of the period of oscillation of the laser light (that is, a small fraction of about 0.03 picosecond). Thus WAVE can cover only a few picoseconds in a given problem, even with the immense power of the Los Alamos computing facilities. In turn, the grid size of one cell is limited to the distance light travels in one time step. As a result, the code is used primarily to study the details of absorption in the region between the underdense plasma of the corona and the overdense plasma close to the surface of the imploding target. At the very low densities of the underdense plasma,

absorption lengths become too long; the code cannot both have the required cells per wavelength of light and still span the absorption length. In the overdense plasma the time and space scales are so short that it is impractical to resolve them with WAVE.

Because of such limits to the time and space scales, the initial conditions and the boundary conditions for the fields and particles are unknown and must be put into the calculation separately. Much of the skill in using the code involves estimating these conditions correctly; they should be consistent with themselves, with hydrodynamic calculations, and with experiment.

Why not expand the time scale by increasing the time step? WAVE calculations are performed by alternately advancing the particles and then the fields. If the time step becomes much larger than a natural frequency in the problem, the calculations can become unstable. The problem is avoided by estimating what the particles *are going to be doing* and then plugging these estimates into the step that advances the fields. This type of code is called *implicit* and forms the basis of our second theoretical tool.

VENUS. Use of the implicit code, VENUS, allows the time and grid size to be increased by about a factor of fifty. Although the increase is obtained by neglecting some high-frequency phenomena, it allows us to handle the spatial scales needed for the study of collective phenomena. In particular, VENUS has been able to verify a model in which very strong self-generated magnetic fields play an important role in electron transport. These magnetic fields also provide the mechanism for copious fast-ion emission, that is, for the ion jets that have been observed streaming from targets. Because of their success with the ICF program, a new generation of implicit codes is rapidly being adopted by the magnetic fusion and space physics communities.



This figure depicts the approximate range of parameters dealt with by each of the computer codes WAVE, VENUS, and LASNEX in typical simulations of laser fusion. The entire surface is a LASNEX calculation, whereas the thin black line on the red surface represents the range of a VENUS calculation, and the even thinner white line represents the range of a WAVE calculation. For example, while the LASNEX calculation can span a range of time from 0 to several nanoseconds, VENUS is confined to a range of only 0.1 nanosecond and WAVE to a range of 0.01 nanosecond.

THE TOOLS

Although quantum mechanics and atomic physics are not included in the implicit codes, the fully developed strong turbulence that can occur in laser-plasma interactions is still described accurately. Such descriptions are limited only by the computer resources. In the last decade the scale of accessible problems has increased by a couple of orders of magnitude because of improvements in both computer hardware speed and the numerical algorithms employed in the implicit codes. This increase has greatly broadened our understanding of the physical processes.

LASNEX. The code that deals with phenomena on the largest time and space scales is called LASNEX. This code is a 2-dimensional radiation flow and hydrodynamics code. LASNEX is used to model electron transport, radiation flow, hydrodynamics of implosion, and possible thermonuclear burn in the target. However, the models in this code are phenomenological, and the simulation is of relatively macroscopic phenomena.

We can see how these three codes mesh by summarizing the various steps for modeling the production of hot electrons. WAVE generates the initial models for such things as the distribution and energies of the hot electrons, but only in the limited region between the overdense and the underdense plasma. To predict large-scale behavior of the plasma, a model generated by WAVE is used in LASNEX and adjusted to fit macroscopic data. However, the time scale of typical LASNEX calculations is too coarse to resolve electron flow in the underdense plasma with its accompanying generation of magnetic fields. As a result, the model is incorporated into VENUS to study these important collective phenomena.

The insight gained from using WAVE and VENUS to simulate and verify models has led to two important improvements for LASNEX. The various physics packages used in the code have been improved, and a better choice of input conditions to the code is possible.

Experimental Tools

A wide variety of experimental techniques have provided the basic data against which the theoretical models must be compared.

Absorption and Energy Flow. We obtain the overall energy delivered to a target in one of two ways. The scattered 10-μm light not absorbed by the target can be measured directly with infrared-sensitive detectors and subtracted from the incident energy. Or all of the ion, electron, and x-ray energy emitted from the target can be measured using calorimeters that are gold-coated to reflect, and therefore reject, scattered CO_2 radiation.

To determine the hot-electron temperature and the amount of hot-electron energy, we can measure either electron or x-ray spectra. Although hot electrons from the target can be detected with magnetic electron spectrometers, only a few electrons actually manage to escape. Therefore, the bulk of the electron energy is studied by measuring the bremsstrahlung radiation resulting from collisions of hot electrons with atoms in the solid parts of the target. We measure these hard x rays with a multichannel broadband x-ray spectrometer that uses an array of filtered scintillators coupled to visible light photodiodes. The measurement covers the spectrum from about 30 kiloelectron volts (keV) to large fractions of an MeV with nanosecond resolution. One detector that responds to x rays of energies greater than 100 keV is used to study the time history of the hot-electron energy with subnanosecond resolution.

The target, heated by hot electrons, emits soft x rays (below 1 keV). This radiation is detected with separate multichannel spectrometers that have filtered vacuum photodiodes sensitive to soft x-ray illumination. The diodes are electrically designed for fast response and are coupled to ultrafast oscilloscopes (developed in the nuclear weapons program), providing a time resolution of better than 300 picoseconds.

Spatial details of the intensity or energy of the x rays from the surface of the target are resolved with multichannel x-ray collimators that restrict the field of view of the diodes. Although such collimators are simple in principle, the small targets used in the laser fusion program require that the collimating pinholes be machined and positioned precisely. To achieve the required 25-μm accuracy, the pinholes, only 150 μm in diameter, are aligned with respect to the target and to each other with optical techniques.

The total absorbed energy measured with ion calorimeters, the fast-ion energy obtained with filtered calorimeters, the energy deposited in the target as determined from the hard x-ray bremsstrahlung emission, and the target surface temperature obtained from soft x-ray measurements, all help reveal the overall energy flow in the experiment. The comparison of these related measurements has provided considerable information about the basic physics of plasmas created by CO_2 lasers. However, considerably more detailed information is required if the physics is to be understood on a more fundamental, microscopic basis.

High-Resolution Imaging and Spectroscopy. Optical and x-ray emission can be imaged, with simple optical or pinhole cameras, to map the distribution either of hot electrons in the target or of the heating of the corona. These techniques can identify nonuniformities in energy deposition due to imperfections in the laser beam or to plasma effects. By using film and filter combinations

to compare images of x rays at fractions of a keV with images of x rays in excess of tens of keV, a picture of the hot-electron flow can be built.

Crystal or grating spectrographs are used to obtain high-resolution x-ray and vacuum ultraviolet spectra that offer a wealth of information on the details of the atomic physics of the plasma, as well as on the plasma conditions themselves. For example, we can identify plasma parameters, such as temperature and density, by selecting specific x-ray transitions from elements in the target and then examining the relative intensities of x-ray lines from different ionization states, as well as for lines from a single ionization state. Spectral line broadening and detailed lineshapes relate to the density of the plasma surrounding the emitting atom. Particularly in the case of extreme conditions far from thermodynamic equilibrium, line intensities and lineshapes must be compared to detailed atomic physics codes to yield accurate temperature and density information. Such information is very useful for evaluating the conditions of the imploding fuel.

Specific characteristic x rays can give us further important information. For example, hot electrons have enough energy to create inner electron shell vacancies, giving rise to K-line x rays, but thermal background electrons do not have enough energy. Therefore, we can image the target in K-line radiation and track the flow of hot-electron energy. We can do this by using two pinholes in a single pinhole camera with specially matched filters: one passes the light of the K line and one does not. By subtracting one image from the other, we can build, in effect, a single-wavelength image of the target. A more attractive, state-of-the-art solution is to use the so-called layered synthetic microstructure (LSM). The LSM acts like a crystal in that it reflects, in accordance with its designed lattice spacing and Bragg's law, only selected frequencies. Placed between the pinhole and the film, the appropriate LSM can be used to reflect only the transition of interest.

Microfabrication techniques can be used to place materials at strategic locations in the target, for example, in a spot on the target surface or as a thin layer on or beneath the surface. The x-ray emission from these materials then serves as a tracer, allowing us to determine the amount or the time of energy flow to these locations by observing the spectral lines characteristic of the tracer material. This technique also allows us to examine in detail the plasma conditions of a well-defined region; we don't have to be content with averages over all densities and temperatures in the plasma.

In addition, both images and spectra can be time resolved to less than ten picoseconds with optical or x-ray streak cameras, enabling us to follow the evolution of the plasma as the experiment proceeds. For example, optical streak cameras, filtered to record only one of the high harmonic frequencies of the CO_2 light, can determine if the conditions producing such harmonics are present throughout the nanosecond experiment. On the other hand, infrared optical emission (for example second harmonic emission from the steep density gradient, light reflected from the critical-density surface in the plasma, light that is Brillouin scattered near 10 μm, or light that is Raman scattered at longer wavelengths) must be measured with fast infrared sensors and spectrometers. Because there is nothing equivalent to a streak camera sensitive to these low-energy photons, it is difficult to examine, in the detail we would like, the evolution of these signatures of the absorption process.

Many of these instruments—particularly the x-ray pinhole cameras, spectrographs, and collimators—must be close to the target and must survive the blast of particle debris and intense x-rays from each shot. In a typical Antares shot hundreds of gigawatts of hard x rays and gigawatts of microwaves are emitted. Heroic measures are required to shield detectors, film, and electronics sufficiently to make the necessary measurements.

The wide range of measurement techniques we have used in this hostile environment, combined with our ability to modify target design in a microscopic manner, has allowed us to develop a detailed empirical base against which the theoretical picture can be compared.

Comparison with Experiment

The code that provides the most direct link with experimental data is, of course, LASNEX. With its post processing packages, LASNEX can simulate directly the outputs of such diagnostic records as pinhole photographs and streak camera images, as well as bremsstrahlung, ion emission, and soft x-ray emission spectra.

The code also calculates the source spectra and so can be used to help deconvolute experimental data. For example, it can extract the electron temperature from the hard x-ray bremsstrahlung spectrum or extract the compressed fuel properties from the x-ray pinhole images and x-ray spectra. Using the code to model a variety of phenomena simultaneously on a given target shot greatly enhances our confidence that we understand the target behavior.

Matching the experimental data frequently requires an iteration in the input conditions to LASNEX, and often the search for a specific signature of a phenomenon predicted by the code leads to suggestions for modifying existing diagnostic techniques or developing new ones. From this interplay between theory and experiment new target concepts are developed that might better use the energy flow. ■

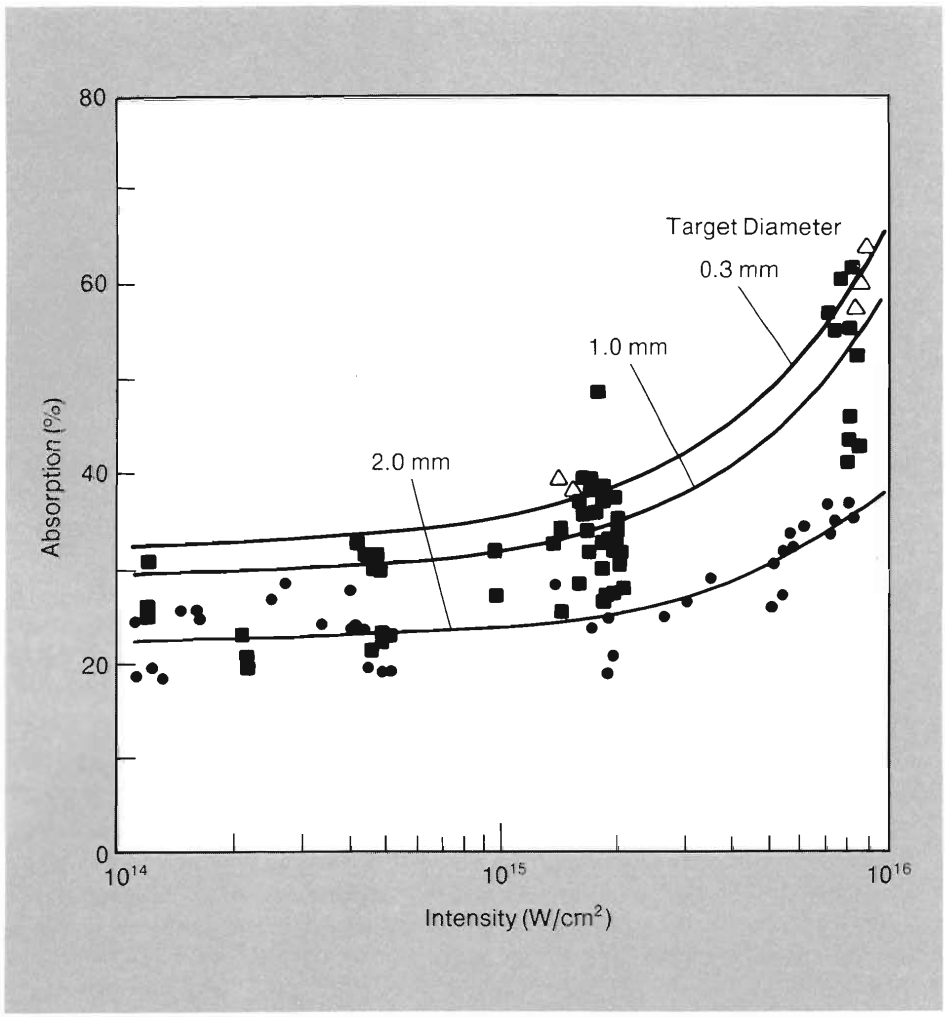


Fig. 2. Absorption of Helios CO_2 laser light on spherical gold targets of various diameters as a function of intensity. The targets were illuminated symmetrically by eight laser beams, and the absorption was measured with fourteen calorimeters deployed around the target chamber, each sensitive to particle and x-ray fluxes but not to scattered 10-μm laser light. The decrease in apparent absorbed energy with larger target diameter is attributed to ion jets that are more directional, returning toward the focusing mirrors and missing the calorimeters. For small targets ion emission is more isotropic.

We see from Eqs. 1 and 2 that as the plasma density n_e increases, so does the plasma frequency ω_p until eventually ω_p becomes greater than the laser frequency ω and the index of refraction becomes imaginary. At this point, light will penetrate no further and is reflected, making this the critical density surface. The region of the plasma with density very close to n_c plays an important role in the absorption mechanisms to be discussed because it is here that the laser and plasma frequencies essentially match ($\omega = \omega_p$).

How much of the incident light is absorbed? Fig. 2 shows the absorption on spheres of various sizes as a function of Helios CO_2 laser intensity. Typically, at lower intensities the absorption is about 30 per cent, whereas it can exceed 60 per cent at higher intensities.

Our description of the absorption processes will start with the simplest form of absorption, inverse bremsstrahlung. This is the dominant absorption mechanism only at low intensities. For higher intensities we have to invoke more complex mechanisms—in particular, those resulting from collective effects. However, the main features of inverse bremsstrahlung form a necessary introduction to the other mechanisms.

Inverse Bremsstrahlung. Inverse bremsstrahlung is a result of collisions between electrons and background ions. Electrons oscillating in the laser electric field scatter randomly off the ions, absorbing a photon of laser light in the process. Thus, the coherent oscillation energy of the laser is converted into the random kinetic energy of electrons.

The rate at which energy is lost from the laser field by inverse bremsstrahlung is approximately equal to the rate at which the electron plasma is heated. Equating these two rates leads to an expression for the dependence of the damping rate ν_D of the laser field on the electron-ion collision frequency ν_{ei}:

$$\nu_D = \nu_{ei} \frac{n_e}{n_c} . \qquad (3)$$

Because ν_{ei} is the frequency for classical Coulomb collisions, it is proportional to $T_e^{-3/2}$, where T_e is the electron temperature. Thus, the efficiency of inverse bremsstrahlung will *increase* as electron temperature *decreases*. In other words, inverse bremsstrahlung preferentially heats low-energy electrons, keeping the plasma close to thermodynamic equilibrium.

How does absorption by inverse bremsstrahlung depend on the wavelength of the laser light? At a specific density, long-wavelength light is absorbed more efficiently than short wavelengths. However, the sharp increase in plasma density encountered by the light as it moves through the corona toward the target reverses this dependence.

First, as suggested by Eq. 3, most of the incident laser light is absorbed near the critical density surface where $n_e = n_c$. But Eqs. 1 and 2 show that n_c, and thus the position of n_c in a density gradient, will be different for different wavelengths. In fact,

$$n_c \propto \lambda^{-2} , \qquad (4)$$

which means that shorter wavelength light penetrates further, encountering electron plasma that is both denser and cooler. These features increase the absorption efficiency for short wavelengths.

A second consideration is the thickness of the region of the plasma that is absorbing significant amounts of light. The sharper the density gradient, the thinner the region between outer, underdense corona and the reflecting surface beneath. A measure of this sharpness is the *density scaleheight L*, the distance over which the density drops by a factor of $1/e$. Also, the absorption length for light, which is proportional to the wavelength λ, must be taken into account. If these parameters are incorporated with the energy-balance considerations that led to Eq. 3, we find that significant absorption occurs only for intensities, in watts per square centimeter (W/cm^2), of

$$I < 5 \times 10^{14}\, ZL/\lambda^4\,, \qquad (5)$$

where Z is the ion charge state, L is in cm, and λ is in μm.

For example, in CO_2-laser-fusion experiments with gold targets, Z is typically less than 79, λ is 10 μm, and the density scaleheight L might be as large as 0.1 cm, which corresponds roughly to the distance the underdense plasma expands in a nanosecond. Such numbers show that absorption of CO_2 laser light by inverse bremsstrahlung is negligible at intensities above 10^{12} W/cm^2, an intensity lower than that at which any of the data of Fig. 2 were obtained. Essentially, the electron-ion collision rate is too slow to account for the amount of energy actually absorbed at the intensities of interest.

The dependence of bremsstrahlung absorption on wavelength (Eq. 5) was well known before any experiments or simulations had been performed. Although research has revealed effects that modify the λ^{-4} relation, the basic conclusions remain the same. To be able to use the CO_2 laser for laser fusion, one must depend upon collective effects. In contrast, for the shorter wavelengths of visible and ultraviolet lasers, such as the KrF laser, inverse bremsstrahlung dominates the absorption process. Forms of collective absorption are present, but only to a modest extent.

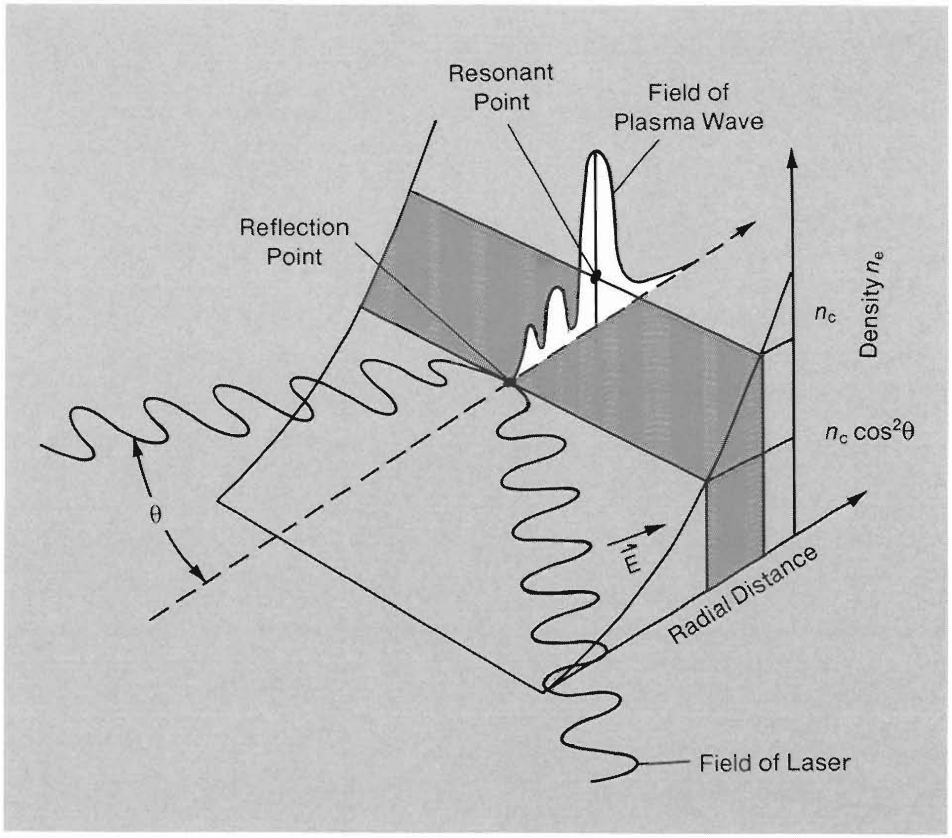


Fig. 3. Resonant matching absorption. The sinusoidal line represents the electric field* E *of laser light oscillating at frequency* ω. *The sloping surface represents a density gradient in the plasma that increases toward the target. Because the light is oblique (at angle* θ*), it is reflected when the density of the plasma reaches* $n_c \cos^2\theta$. *Oblique light also has a component of its field perpendicular to the surface, and that part of the field can tunnel inward to couple with the longitudinal electrostatic field of the plasma wave. This coupling occurs most efficiently at the resonant matching point, that is, the point where the plasma density equals* n_c *and the plasma frequency* ω_p *equals* ω. *The steeper the density gradient, the smaller the gap between the reflection and resonant matching points and the more efficient the absorption.

Two basic absorption mechanisms involving collective effects have been identified as important: resonant absorption and parametric instabilities. Both mechanisms are at work in all laser plasma interactions but are particularly prominent at the wavelength of the CO_2 laser. Also, both are a result of the pressure of the incident light wave, the so-called *ponderomotive force.*

This important low-frequency force is proportional to the light intensity and the density ratio n_e/n_c. For example, at the intensity of 10^{16} W/cm^2 achieved at Antares, the pressure of the light at its reflection point is about 5 megabars. Such a large force is able to distort the flow of expanding plasma at low densities. It is thus responsible for most of the instabilities induced by the incident radiation in the outer layers of underdense plasma and plays a critical role in the collective absorption mechanisms.

A Mechanism for Moderate Intensities: Resonant Absorption

The first collective absorption mechanism, resonant absorption, is a *linear* coupling of the electric field of the laser beam to *longitudinal electron-plasma waves.* This coupling will not occur unless the electric field of the light has a component parallel to the density gradient. Thus, the light must be oblique to the surface of the plasma (Fig. 3).

An electromagnetic wave with incident angle θ is reflected *not* at the critical density n_c but at the lower density $n_e = n_c \cos^2\theta$. If

the absorption is to be efficient, the light must reach the resonant-matching point at n_c where $\omega = \omega_p$. In plasma with a steep density profile, these two points are close together, and a significant part of the wave may tunnel quantum mechanically to the resonant-matching point. Once there, the component of the electric field vector that lies along the density gradient induces density fluctuations at the local plasma frequency ω_p. This coupling acts as a source of plasma waves, extracting energy from the incident electromagnetic wave.

The efficiency with which resonant absorption converts light energy to plasma-wave energy depends sensitively on the incident angle and the length of the gap between the reflection point and the resonant-matching point. For example, simulations show that at a typical angle of incidence of 20 degrees, this length must be *less than a wavelength of light* to obtain an absorption greater than 20 percent. If the ponderomotive pressure is large, the reflecting light and the locally generated plasma wave together produce a sharp density gradient in the region. In other words, the pressure of the light creates a sudden wall in the plasma that keeps the length short. This allows resonant absorption to be an effective process.

Figure 4, a simulation using our WAVE code, illustrates the enormous density gradients that can be produced. At the peak of the gradient, the density has increased to sixty times critical density with most of this increase happening over a distance of only 1 or 2 μm!

Although resonant absorption is a linear coupling between electric field and plasma wave, the gradients involved are so extreme that nonlinear phenomena abound. In fact, resonant absorption at high laser intensities is identified experimentally by a nonlinear phenomenon: the emission of second harmonic light (Fig. 5). How does this occur?

If laser energy is coupled resonantly into the plasma wave, density fluctuations begin growing that cause the index of refraction of the plasma to also change with frequency ω.

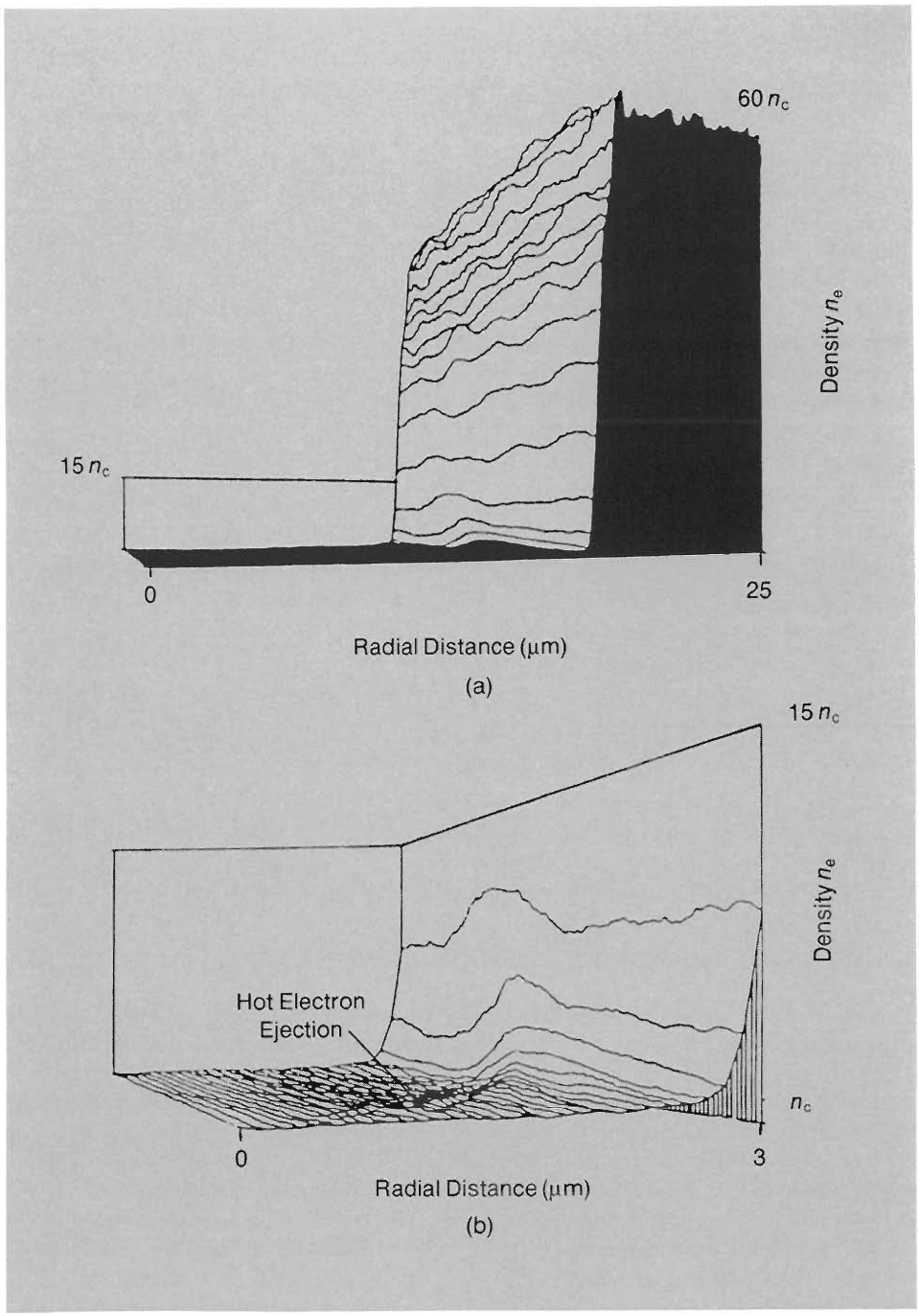


Fig. 4. A calculation by WAVE of the density n_e of electrons in the plasma close to the critical density surface n_c. (a) The density gradient rises to a peak value sixty times n_c over a distance of only a few micrometers. (b) This detail shows a stream of hot electrons being ejected from the surface of the density front.

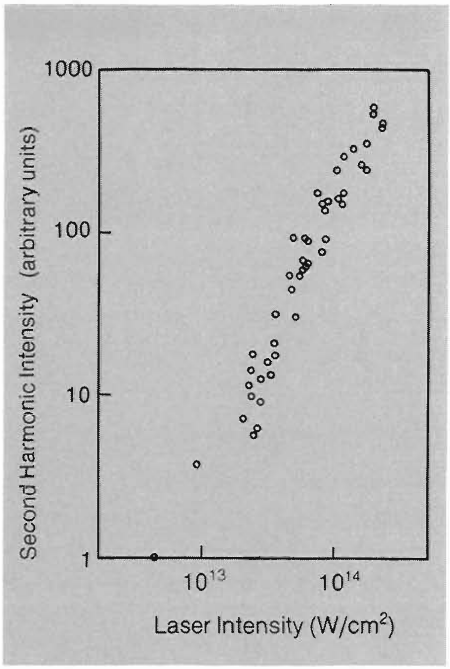


Fig. 5. Second harmonic emission as a function of the intensity of incident laser light at the target. This emission is experimental evidence for resonant absorption and a sharp density gradient in the plasma. (Adapted from H. A. Baldis, N. H. Burnett, G. D. Enright, and M. C. Richardson,* Applied Physics Letters *34(1979): 327-9.)

The incident light interacts with this modulated index in a nonlinear or anharmonic manner, generating light at the second harmonic frequency 2ω. The intensity of the second harmonic is proportional to both the incident light intensity and the plasma wave intensity and is thus a direct measure of the amount of energy in the plasma wave.

Wave Breaking. Once energy is placed into the plasma wave by resonant absorption, a number of mechanisms quickly dissipate the energy throughout the plasma. In particular, hot electrons are generated that are shot outward from the sharp wall formed by the density gradient (Fig. 4(b)). What produces these electrons and how energetic will they be?

The primary mechanism is a process known as wave breaking that is analogous to a surfer riding a wave, only here it is an electron riding a plasma wave. In wave breaking only a few electrons are accelerated by the plasma wave and only over distances of less than a micrometer. However, these few electrons gain energies of hundreds of keV in less than one laser cycle (about 10^{-14} second), becoming the hot electrons.

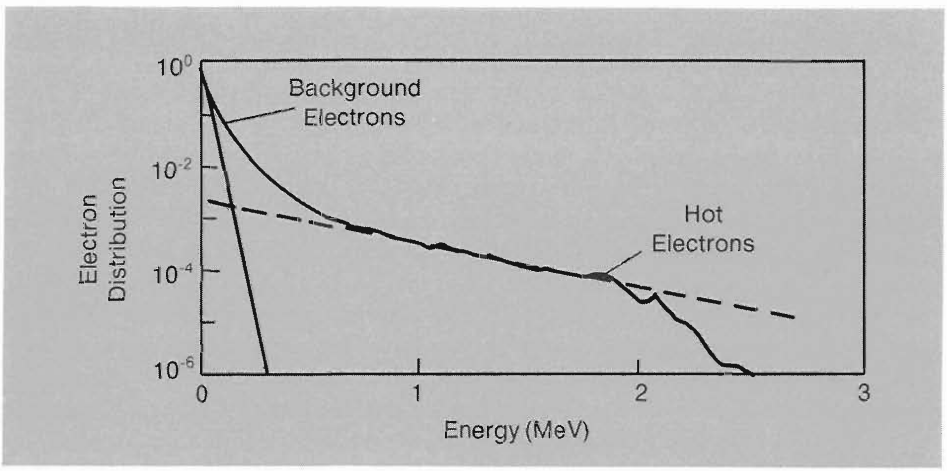


Fig. 6. This calculation with WAVE of the energies of electrons clearly shows two components: large numbers of cooler background electrons and small numbers of hot electrons distributed over a wide range of energies. Even though the hot electrons are few in number, they carry essentially all the energy deposited in the plasma by resonant absorption.

The energy gained by electrons via wave breaking is proportional both to the electric field in the plasma wave and to the *width* of the plasma wave. Although a steepened density profile increases the efficiency of resonant absorption, it also causes that absorption to occur in a physically narrower region of the plasma, generating a narrower width for the plasma wave and reducing the energy of the accelerated electrons. In other words, this effect should make the hot electrons generated by resonant absorption somewhat cooler than originally thought.

Initial experimental evidence that showed the scaling of hot-electron temperature with intensity to be much weaker than expected spurred a detailed exploration of the effect with simulations and experiments. The theoretical results can be summarized by looking at the scaling predictions with respect to intensity I and wavelength λ. Before the studies, it was thought that hot-electron temperature would scale as $I\lambda^2$. Afterwards, it was thought that the temperature would scale as $(I\lambda^2)^{1/3}$; that is, the production of highly energetic electrons would be much less dependent on laser wavelength. This result implied that resonant absorption could help achieve the needed conditions for CO_2 laser fusion.

What else did computer modeling predict? Although wave breaking is a coherent acceleration process, there is a small random change in the acceleration from cycle to cycle. This stochastic element produces a near Maxwellian distribution of hot electrons, but the region in the plasma where significant wave breaking occurs is very thin. In the larger region of lower density plasma the fraction of electrons that get accelerated is small. The result, overall, is plasma that typically has two components: one component consists of a large number of cooler background electrons, and the other consists of a small number of energetic electrons spread over a broad energy spectrum. Despite their small numbers, it is the hot electrons that carry most of the energy.

A WAVE calculation of the distribution of electrons at high intensity is shown in Fig. 6.

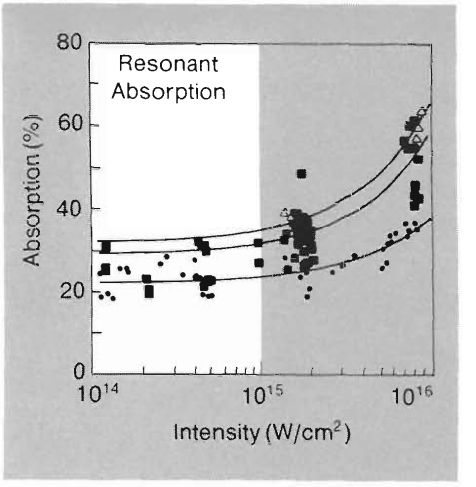


Fig. 7. Resonant absorption explains the almost constant absorption observed for intensities from 10^{14} to 10^{15} W/cm^2.

A major uncertainty in such a calculation is the value of the cool background temperature. We cannot use WAVE—our most rigorous simulation code—to determine this temperature self-consistently, because the temperature depends on the radiative cooling processes, the expansion of the solid material, the ionization rates, and so forth. Nevertheless, wave breaking can be modeled in a relatively simple way with our LASNEX code, and such modeling provides a reasonable match to experiment.

What *is* the match to experiment? First, theory predicts the nearly constant absorption of 20 to 30 per cent observed at intensities from 10^{14} to 10^{15} W/cm^2 (Fig. 7). Further, *at these intensities* scaling with wavelength is weak; that is, a large number of measurements of hot-electron temperature from x-ray emission and ion emission are consonant with the predicted $(I\lambda^2)^{1/3}$ dependence (Fig. 8). At lower intensities, steepening of the density profile is weak, and the original strong scaling holds.

The observation of high harmonics of the incident laser light is our most spectacular confirmation of wave breaking and the steepness of the density gradient. Put simply, the incident laser energy strips electrons from the density gradient wall in the plasma (Fig. 4(b)), generating many harmonics because of the highly nonlinear nature of the process. One might say that incident light is pressing at the density gradient in a manner analogous to fingernails scraping across a blackboard. More than thirty-five harmonics of the incident light have been detected experimentally in the scattered light (Fig. 9), confirming the presence of the steepened density gradient.

Figure 10 illustrates the anharmonicity that gives rise to these high harmonics. An electron, attempting to oscillate at the plasma frequency within the thin region of the sharp density gradient, sees a highly asymmetric environment. It moves from low to high density, from a region of low restoring force to one of high restoring force, experiencing extremely anharmonic acceleration. In effect, the electrons of the plasma wave see a restoring force that is essentially proportional to ω_p^2.

How many harmonics should we observe? The highest plasma frequency equals the frequency of the maximum harmonic that gets scattered. Because the highest plasma fre-

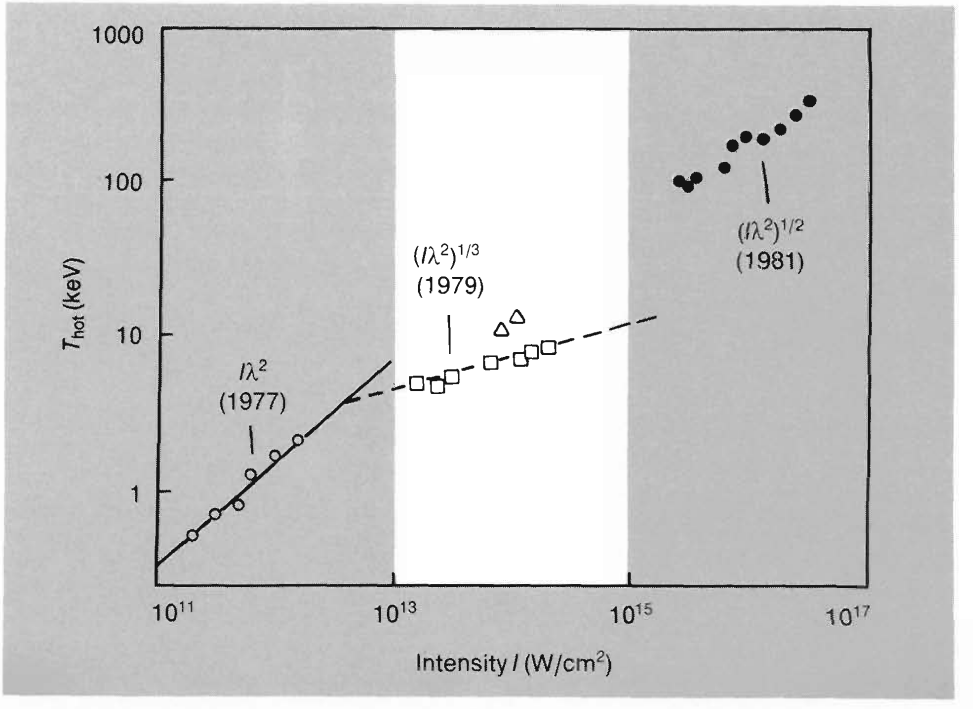


Fig. 8. As the intensity I of the light increases, measurements of hot-electron temperature show different $I\lambda^2$ dependencies, where λ is the wavelength. In 1977 the low-intensity hot-electron temperature scaled as $I\lambda^2$. About that time fast-ion data indicated a weaker scaling at higher intensities, motivating calculations that showed a weaker scaling for T_{hot} in a steepened density profile. By 1979 x-ray data at higher laser intensities was published that revealed the scaling to be proportional to $(I\lambda^2)^{1/3}$, that is, consonant with resonant absorption in a density profile steepened by the ponderomotive force. These results encouraged researchers to think the hot-electron temperature might not be too serious an impediment to laser fusion. The most recent data, at even higher intensities, show a scaling that is proportional to $(I\lambda^2)^{1/2}$. Besides increasing the difficulty of CO_2 laser fusion, these last data imply the onset of another absorption mechanism. (The data in the figure are taken from: C. Stenz, C. Popovics, E. Fabre, J. Virmont, A. Poquerusse, and C. Garban,* Le Journal de Physique *38 (1977):761; D. W. Forslund, J. M. Kindel, and K. Lee,* Physical Review Letters *39 (1977):284 (which cites further original data sources); G. D. Enright, M. C. Richardson, and N. H. Burnett,* Journal of Applied Physics *50 (1979):3909; and W. Priedhorsky, D. Lier, R. Day, and D. Gerke,* Physical Review Letters *47 (1981):1661.)

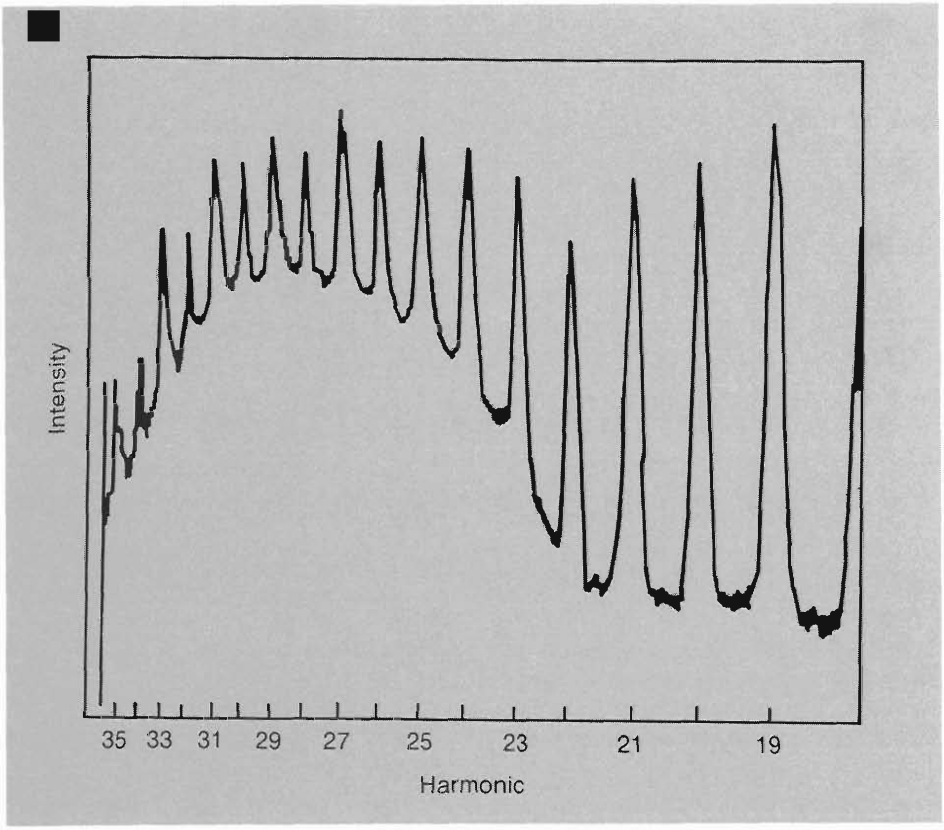
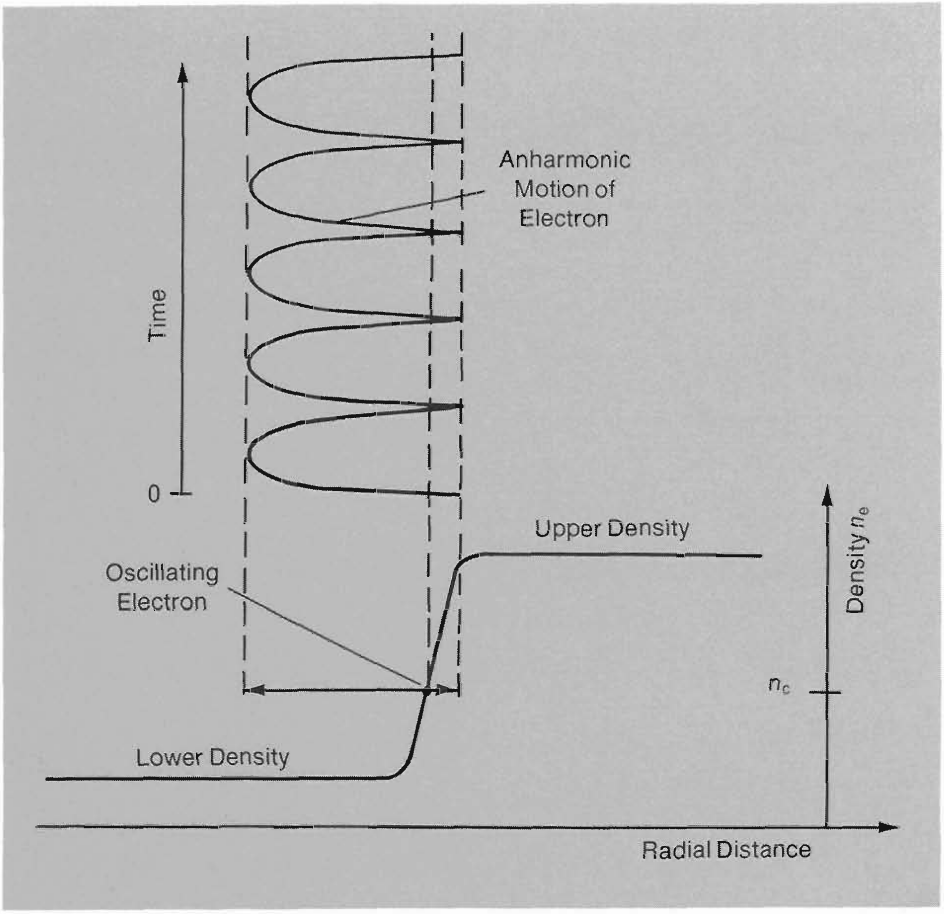


Fig. 9. Harmonics of the laser light frequency ω are experimental evidence for a steepened density gradient and acceleration of electrons by wave breaking. The highest harmonic, here 35ω, is a measure of the height of the upper density shelf (see Fig. 10).



quency occurs at the highest density, the maximum harmonic should equal ω_p in the upper density shelf. Since $\omega_p^2 \propto n_e$ (Eq. 2), the square of the highest harmonic is a measure of the density in the upper shelf. We have observed close to 40 harmonics, implying that the plasma wave is seeing densities up to 1600 times critical density, that is, essentially solid density.

In recent years both experiments and more extensive calculations have indicated that the very steep density gradient may dissipate quickly. Ultrafast optical streak cameras show that the high harmonics last only during the rise time of the laser pulse, about 200 picoseconds. Such transient behavior is due, in part, to the outward expansion of high-density plasma, which eliminates the anharmonic environment that the plasma wave sees.

In addition, experiments above 10^{15} W/cm^2 showed that absorption increased dramatically (Fig. 11), contrary to the prediction of a nearly constant coefficient for resonant absorption, and that hot-electron tem-

Fig. 10. An electron oscillating within the thin region of the sharp density gradient undergoes extreme anharmonic motion, experiencing a restoring force that is essentially proportional to ω_p^2. The result of this anharmonicity is the generation of scattered light containing the higher harmonics nω of the laser frequency (Fig. 9). The highest plasma frequency—ω_p in the upper density shelf—determines the highest possible harmonic frequency. Because the density of the plasma is proportional to ω_p^2 (Eq. 2), the density of the upper shelf is proportional to the square of the highest observed harmonic. Thus, the observation of the 40th harmonic implies a density for the upper shelf that is 1600 times n_c.

perature scaled as $(I\lambda^2)^{1/2}$ rather than $(I\lambda^2)^{1/3}$ (Fig. 8). These facts and the short lifetime of the density gradient suggests that an additional absorption process—another type of collective mechanism—occurs at these high intensities.

High-Intensity Absorption and Parametric Instabilities

Resonant absorption involves a *linear* coupling of light to the plasma wave. Parametric instabilities, on the other hand, result from a *nonlinear* coupling with plasma waves. As before, the coupling is through the ponderomotive force, but now light couples to plasma wave frequencies *different* from the resonant frequency.

In general, the conditions for a parametric instability are met when the frequency of a scattered wave ω_s, the laser frequency ω, and the plasma frequency ω_p satisfy conservation of energy,

$$\omega_s = \omega - \omega_p , \qquad (6)$$

and the wavevectors k_s, k, and k_p satisfy conservation of momentum,

$$k_s = k - k_p . \qquad (7)$$

In essence, part of the energy of the laser light appears in the plasma as an oscillating motion of particles—either in electron-plasma waves or in ion sound waves. The remaining energy may appear as scattered light or as another plasma wave, but, in either case, it will be at the lower frequency ω_s.

One concern about this type of mechanism is that a significant fraction of the laser energy may be backscattered away from the target as light at frequency ω_s. However, as we will explain in a moment, a more serious concern is the generation of very energetic hot electrons.

In general, parametric instabilities occur in the underdense plasma farther out than the critical density surface (Fig. 12). That this should be the case is obvious from the fact that the plasma-coupling frequency ω_p is necessarily lower than ω, so the matching must occur in plasma of lower density than n_c. The effect then is to reduce the energy of the laser light before it can reach the resonant matching point. Depending on the instability, overall absorption may be either increased or decreased.

Scattering Processes. The most important parametric modes appear to involve coupling of light waves to *electron-plasma waves*. When the scattered wave is light, the mode is known as stimulated Raman scattering.

In Raman scattering the conservation relations require the light to interact with a range of plasma waves whose frequencies are less than or equal to half the light frequency. This means the interaction occurs in the plasma at

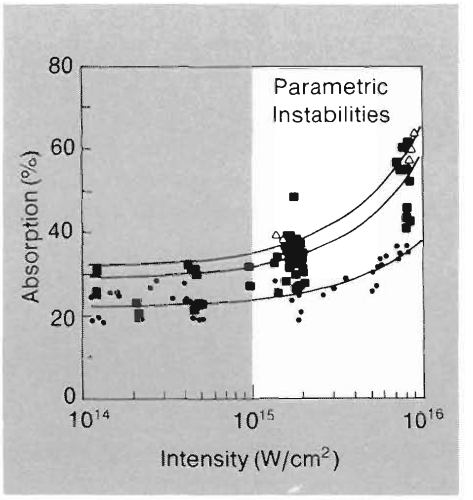


Fig. 11. The dramatic rise in absorption above 10^{15} W/cm^2 requires the introduction of other collective absorption mechanisms: parametric instabilities.

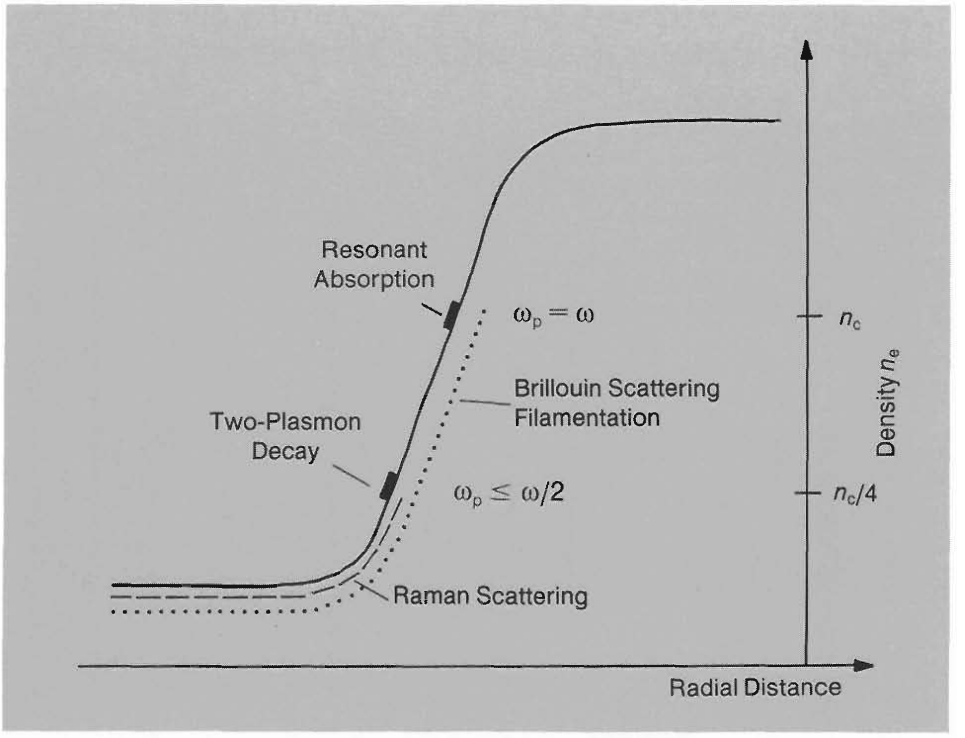


Fig. 12. Resonant absorption occurs in the plasma close to the critical density surface where $\omega_p = \omega$. Parametric instabilities couple to plasma waves with frequencies equal to or less than ω_p, with most occurring in the underdense plasma below $\omega/2$. Parametric instabilities include stimulated Raman scattering, two-plasmon decay, stimulated Brillouin scattering, and filamentation instabilities.

densities less than one-fourth n_c where the gradient is not as steep and the interaction region is longer. Thus a broad spectrum of electron-plasma waves are excited that, again through wave breaking, accelerate electrons to high energy. However, in contrast with the short acceleration distances of less than a micrometer for resonant absorption, acceleration distances for Raman scattering can be as long as a millimeter.

A process, called *beat wave acceleration*, is being developed that will take advantage of these relatively long acceleration distances to produce extremely energetic particles. Two lasers with a frequency difference equal to the appropriate plasma frequency are used to enhance the plasma-wave amplitude over that generated by Raman scattering itself. The resulting high-velocity and large-amplitude plasma waves may be capable of accelerating particles to the energies envisioned for the next generation of high-energy physics experiments. Although numerous practical problems must be solved, this scheme could produce a relatively compact high-energy accelerator.

When the plasma frequency is *exactly* half the laser frequency, a parametric interaction called two-plasmon decay occurs. In this case, the incident light decays *directly* into two electron-plasma waves, both of frequency $\omega/2$, and no light is scattered. Because two-plasmon decay depends even more strongly on the density scaleheight and the temperature of the plasma than does Raman scattering, it is not considered to be a significant absorption mechanism.

The longer interaction regions for Raman scattering imply that the hot-electron temperature should be much higher than that predicted for resonant absorption. Our simulations (Fig. 13) confirm this prediction, with hot-electron temperatures climbing to over 10^3 keV at intensities of 10^{16} W/cm^2. This, of course, is very detrimental for laser fusion.

On the other hand, it should also be noted from Fig. 13 that absorption due to these mechanisms drops to less than 1 per cent of the laser light at 10^{16} W/cm^2. Apparently, the extreme energies of the hot electrons cause a strong damping of the electron-plasma waves, which, in turn, reduces the absorption by these mechanisms. Thus, stimulated Raman scattering must *not* be responsible for the dramatic increase in the observed overall absorption.

There are several mechanisms that, potentially, could explain the increased absorption. The first is Brillouin scattering, which is identical to Raman scattering except the light couples to *ion sound waves* rather than electron-plasma waves. Our calculations indicate, however, that the heating rate from such an ion-based mechanism is significantly lower than for Raman scattering and may be of importance only for the shorter wavelengths.

Extreme intensities also produce self-focusing and filamentation instabilities of the incident light, which, at least in the general sense, are viewed as parametric instabilities. At high intensities, the ponderomotive force of the light literally pushes plasma aside to form channels of less dense plasma but of increased light intensity (Fig. 14). Because the index of refraction is directly related to n_e, such action causes the index to decrease and curves the beam even further into the high-intensity region. Self-focusing refers to the case in which most of the beam is pulled together into a single, intense region; filamentation refers to the case in which intensity peaks are independently reinforced so that the beam breaks into several individual beams. Both of these effects have been observed in numerous experiments at both long and short wavelengths. However, our calculations show that these instabilities build only about as rapidly as Brillouin scattering, and thus they cannot account for the increased absorption.

Theoretically, a number of other parametric instabilities at the critical surface have been identified. Our original simulations, however, showed that the steep density gradients induced by the ponderomotive force severely reduced the instabilities, making them negligible. What then was the mechanism responsible for the dramatic increase in absorption above 10^{15} W/cm^2? Perhaps the key to the puzzle was the behavior of the density gradient for longer times. If, as suggested earlier, the steep gradient dissipated, absorption due to parametric instabilities could become important.

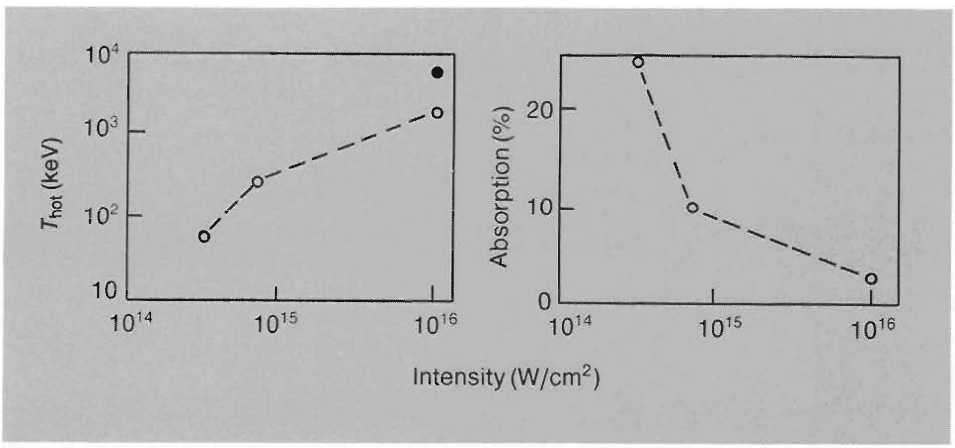


Fig. 13. WAVE simulations showing the effect of Raman scattering on hot-electron temperature absorption. The solid circle in the hot-electron temperature plot is for a somewhat lower density plasma than the open-circle points but all results show very high temperatures at laser intensities of 10^{16} W/cm^2.

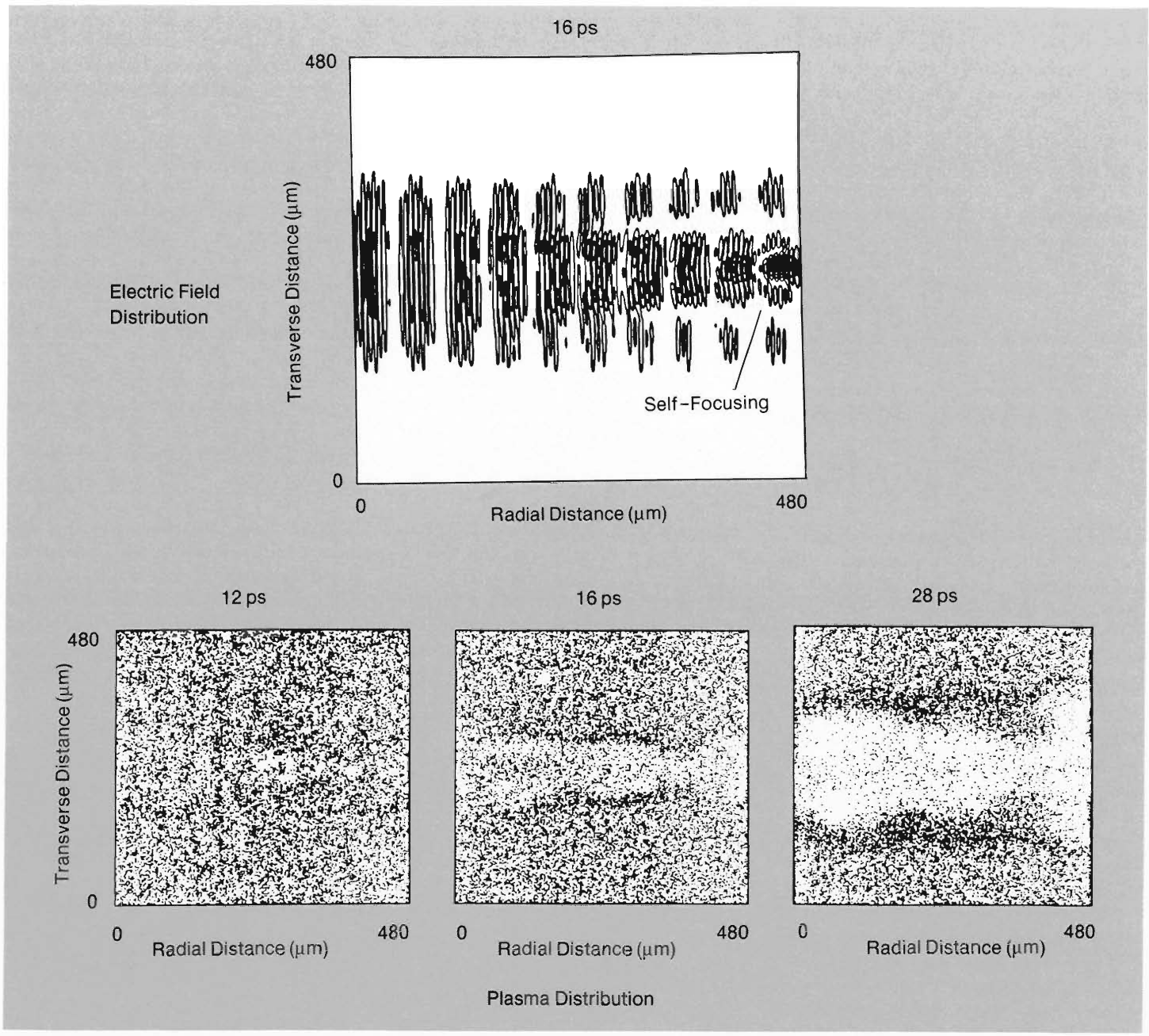


Fig. 14. Self-focusing. All four panels depict a WAVE simulation at the same region in space. The top panel shows electric field lines at 16 picoseconds. Although the "laser beam" for this simulation is actually two colinearly propagating beams, the results clearly illustrate self-focusing by the way the inner contours converge as the beam moves toward the right. The lower three panels show the distribution of the electron plasma at 12, 16, and 28 picoseconds. Here we see plasma pushed aside as time progresses, resulting in a channel in which the light is focused to higher intensity.

A Mechanism for High Intensities

The recent acquisition of more Cray computers at Los Alamos has made possible simulations that run for tens of picoseconds rather than only a few picoseconds. Figure 15 shows contour plots of the density surface from such a simulation at 1.5×10^{15} W/cm^2. Note that the density gradient, which was originally smooth, has broken up and become rough.

Apparently, a bootstrap effect occurs that starts with a small amount of energy being absorbed through parametric instabilities and deposited in plasma waves. The varied motions of these waves roughens the surface of the steep gradient, the roughness causes the gradient to have less effect on reducing the instabilities, and more energy is absorbed. The result is a substantial increase in absorption from about 25 to 60 per cent. At the same time the hot-electron temperature increases by a factor of 2 or 3 over that calculated for resonant absorption on the initially smooth surface. Also, with a rough surface the angle of acceleration for the hot electrons becomes much broader, and these particles are shot out in all directions, including inward toward the target.

At 10^{16} W/cm^2 the effect takes place much more rapdily, and the surface becomes even more turbulent. The parametric instabilities at the critical density that were suppressed by the sharp gradient now, at high intensity, appear to become dominant. This process best explains both the increased hot-electron temperatures and the rise in absorption observed at high intensities.

What, then, is our overall picture? We see that the large amplitude of the laser radiation drives the plasma into a highly nonequilibrium state. It is so far from equilibrium that classical shock waves are altered. The region of resonant absorption corresponds essentially to a *phase transition* between hot plasma with laser light and colder plasma without laser light. In this region rarefaction shocks are generated that have very different properties from conventional shock waves. Matter is thus put into an extremely unusual state that is probably only reproduced in exotic astrophysical environments.

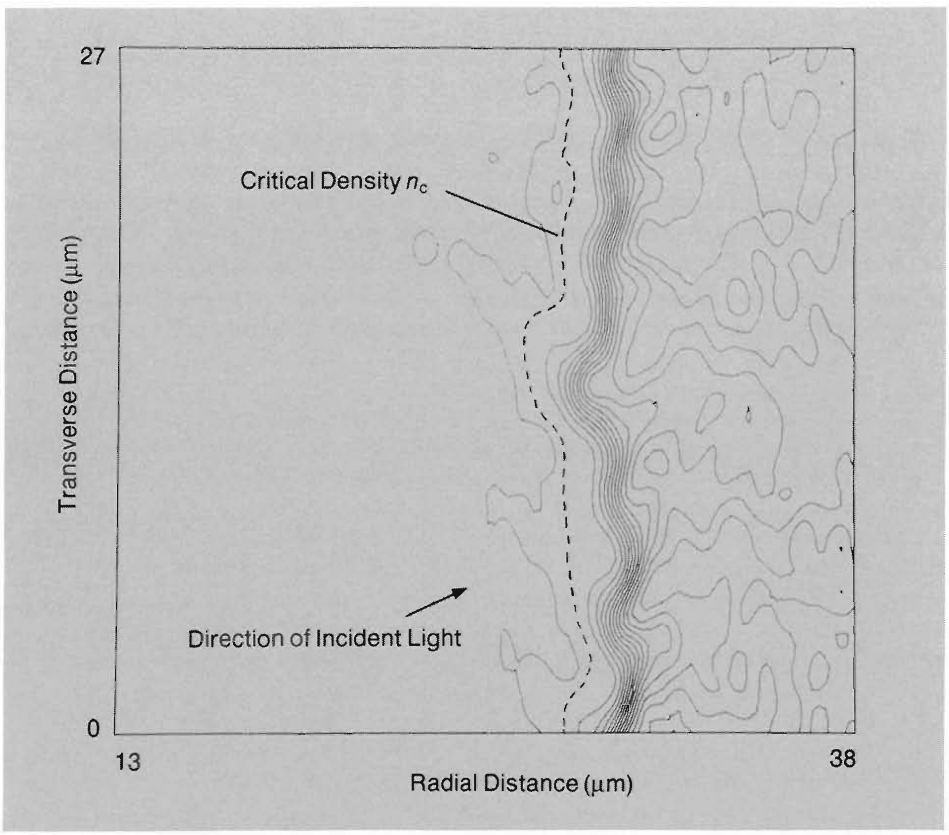


Fig. 15. A WAVE simulation that illustrates with ion-density contours the roughness that builds up on the surface of the density gradient after twenty picoseconds. The laser light here has an intensity of 1.5×10^{15} W/cm^2, is incident at an angle of 20 degrees to the surface normal, and is polarized in the plane of the figure. At higher intensities the roughness builds up even more quickly.

Fast-Ion Generation

One of the more exciting features of this unusual state of matter is the generation of *ion jets* (depicted in the opening artwork). As it turns out, these jets are another manifestation of collective effects, because, just as collective effects from the light control the absorption of laser energy, collective effects *in the plasma* dominate the transport of electrons. Specifically, the pressure of hot electrons in the corona can collectively accelerate coronal ions to extremely high energies by collisionless processes.

In the simplest model a small number of hot electrons are stripped from the face of the density gradient and leave the target. These few electrons create an imbalance of charge, and electrostatic potentials of hundreds of kilovolts develop that are unable to bleed off through the target support stalk. Thus, an electrostatic sheath of positive coronal ions is formed that confines the remainder of the electrons. Initially, hot electrons are still projected outward, but they must work against the high potential of the sheath. The net result is a "bounce" in which electrons are redirected inward and positive ions are accelerated outward.

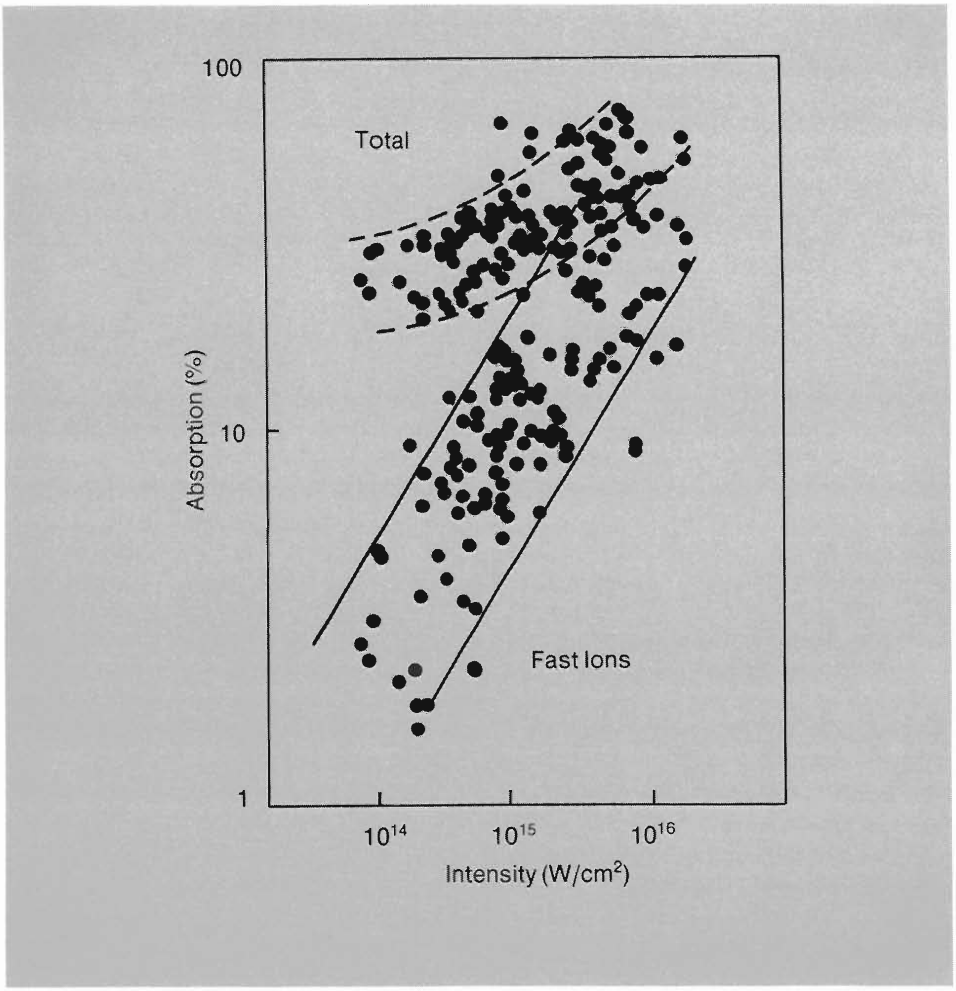


Fig. 16. These data from the Helios laser facility clearly show that at higher intensities a significant fraction of the total absorbed energy appears in the form of fast ions. Targets included spheres that ranged, in diameter, from 200 μm to several millimeters and, in material, from plastic to gold. Much of the scatter in the data is due to small-scale fluctuations in intensity across the target and to finite sampling of ions with respect to angular distribution. Within this scattering, there is little difference in fast-ion energy due to changes in the atomic number of the target.

Experimental measurement of the fast-ion energy shows that a substantial fraction of the absorbed laser light goes into fast ions, particularly at high intensity (Fig. 16). Theoretical calculations based on the fraction of energy lost by an electron when it is reflected indicate that, if an electron has only a *single chance* to bounce off the electrostatic sheath, the fraction of energy in fast ions cannot exceed 5 to 10 per cent. Thus we infer that a mechanism exists to trap the electrons in the corona, allowing them to lose more of their energy to fast ions. Known generically as *flux-limited transport*, this mechanism must somehow reduce the mean-energy penetration velocity to less than that for simple diffusion. One potential candidate for this process is the generation of intense magnetic fields by the high-energy electrons themselves.

If illumination of the target is not uniform, as is the case for a finite laser spot, the beam creates a temperature gradient ∇T tangent to the illuminated surface and a strong density gradient of electron plasma ∇n_e normal to the surface. As shown in Fig. 17 (a), these mutually perpendicular gradients generate a magnetic field **B**, with field lines encircling the spot, at a rate given by

$$\frac{\partial \mathbf{B}}{\partial t} = -\nabla \times \mathbf{E} = (\nabla n_e \times \nabla T)/n_e \,. \quad (8)$$

Typical values of the parameters in Eq. 8 yield an extraordinarily high value for $\partial \mathbf{B}/\partial t$. For example, with a density gradient of 10 inverse μm, a laser spot size of 100 μm, and a peak temperature in the plasma of 50 keV, a magnetic field equal to a full megagauss can be reached in 1 picosecond. In the time it takes an ejected hot electron to cross the laser spot the field is strong enough to reduce its gyromagnetic radius to less than the density scaleheight. As a result, hot electrons can no longer stream freely from or into the target but are confined by the magnetic field close to the surface.

Furthermore, simulations with our VENUS code show that the electric and magnetic fields can drive the trapped electrons along the target surface at velocities close to the speed of light ($v = cE/B$). This rapid drift (Fig. 17 (b)) reduces the inward transport of electrons under the laser spot and carries some of the electrons and their energy great distances across the target. Moreover, electrons drift along surfaces of constant magnetic field, and a considerable number of them end up circulating around the edge of the spot where the field is strongest.

The result of this pileup of electrons around the spot is a dramatic increase in their effective pressure P that produces a strong electric field, $\mathbf{E} = -\nabla P/n_e$, directed parallel to the density gradient over the spot (Fig. 17 (c)). This field accelerates ions, producing an ion jet, or plume, normal to the target surface.

Figure 18—despite the resemblance, noted by the Antares experimental team, to a certain extraterrestrial—is an x-ray image of a spherical target at Antares. The target is being hit with six laser beams simultaneously, and five of the impact regions can be seen in the photograph. The resulting distribution of emitted x rays demonstrates in a vivid manner the existence of the intense magnetic fields that circle the laser spots. Between the "eyes" of the extraterrestrial is a narrow region of significant x-ray emission formed where the magnetic field lines from adjacent spots meet and, because of their opposite polarity, cancel. As a result of the null magnetic field, electrons are no longer trapped, and energy can be released. This effect was first seen dramatically on spheres, flat targets, and cylindrical targets in experiments at Helios where, in reference to the Martian features, they were dubbed "canals."

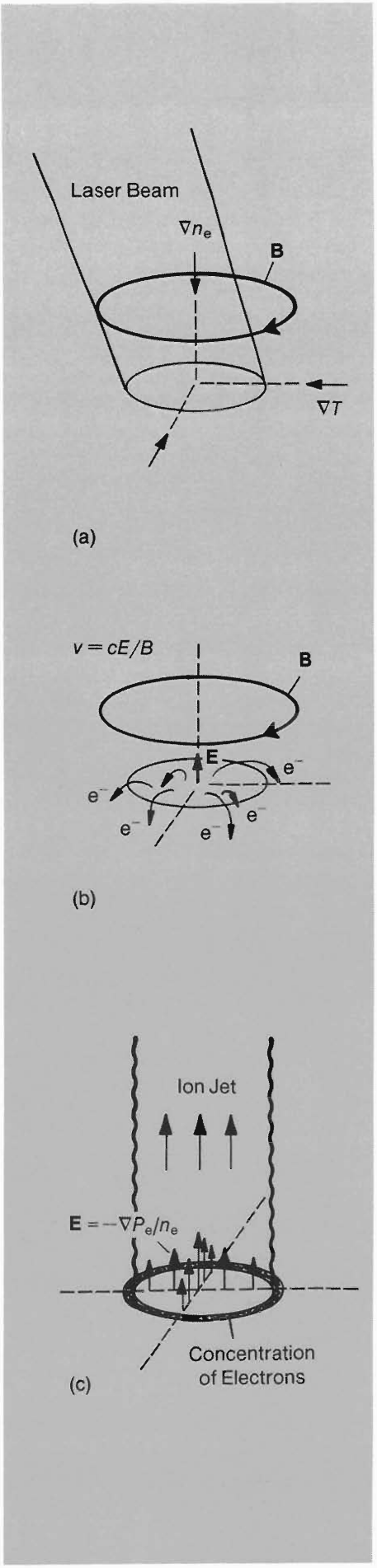


Fig. 18. An ~1-keV x-ray image of the surface of a target being hit simultaneously by the six laser beams of Antares. The narrow region of x-ray emission at the center of the photo is evidence of strong magnetic fields encircling each laser spot; here the fields from adjacent spots cancel, allowing the electrons to move freely into the target rather than being trapped above the surface.

Fig. 17. (a) At the surface of a target, a finite laser beam creates mutually perpendicular density and temperature gradients (∇n_e and ∇T, respectively) that generate an intense magnetic field **B** *encircling the spot. (b) This magnetic field confines hot electrons close to the surface of the target, but they are free to drift sideways under the influence of both magnetic and electric fields with a velocity v given by cE/B. In this manner electrons are driven away from the center of the spot at velocities close to the speed of light with many concentrating at the edge of the spot where the magnetic field is highest. (c) The resulting electron pressure P_e generates high electric fields over the region, and these fields accelerate ions upward, forming the ion jet.*

Ion Jets. To summarize, then, inhomogeneities of temperature and density around a laser spot create intense magnetic fields that trap and channel the flow of the hot electrons. The pressure from these trapped electrons generates intense electric fields normal to the surface that accelerate ions, forming a jet. A partial image of one jet has been captured (Fig. 19) with a pinhole camera that separately images x rays emitted from the target and ions accelerated out of the plume. What are the properties of such a jet?

Figure 20 shows the angular distribution of ions leaving a flat disk target that is large compared to the size of the laser spot. These data were gathered on a *single* laser shot, using three sets of calorimeters filtered to give different energy thresholds for the ions. The plume has a halfwidth at half maximum of about 10 degrees, in excellent agreement with particle simulations using VENUS.

For targets smaller than about a millimeter, including small spheres, the effect is washed out and the ion angular distribution becomes more isotropic. The effect is also washed out in thin targets where electrons can travel from the laser spot by reflexing through the shell and disrupting the current flow that creates the magnetic fields.

The acceleration of ions in a jet is equivalent to that achieved using intense pulsed-power ion diodes, only here the intensity is much higher than for any conventional diode. In fact, these accelerated ions might be directed at a second target to drive a fusion reaction, except that the collimation of the ion jet is not sufficient.

Experiments also suggest that at high intensity the ion emission may remain high even when the illumination is more uniform. At this point no theory has been devised that conclusively reproduces this effect.

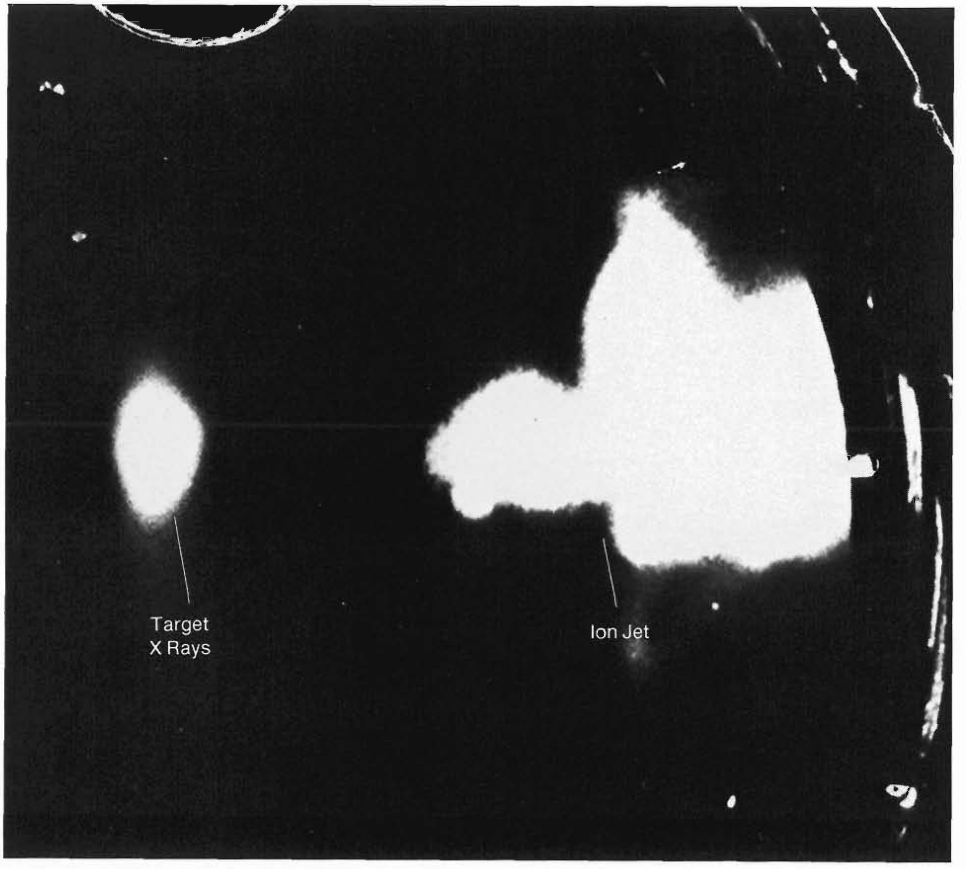


Fig. 19. A sideways image of a portion of an ion jet is captured here by superimposing two negatives: the first is a record of x rays emitted at the surface of the target during one shot, and the second is a record of ions scattered sideways out of the plume during another identical shot. The scattered ions create tracks in a plastic emulsion that is viewed with dark field imaging techniques. The early part of the jet closest to the target is not visible because essentially all the ions are accelerated in the direction of the jet; only after the plume has developed turbulence do a significant number of ions get accelerated sideways toward the emulsion. The later part of the jet is obscured by the round port through which the event was recorded.

Ion Energies. Measurements of the ion velocity distributions with magnetic analyzer spectrometers indicate that, independent of the target material, much of the ion energy is carried by hydrocarbons that are surface contaminants. Such energy partitioning is due, in part, to the favorable acceleration of ions with low atomic number. Although the energy spectrum of this multispecies ion expansion is remarkably complex—and thus difficult to calculate theoretically—its gross properties are deceptively simple.

The mean energy of the ions is not far above that required to penetrate a nickel filter with a thickness of 0.5 μm, so we can determine an "ion spectrum" in the form of transmission curves using filtered calorimeters (Fig. 21). Not only are the resulting curves well behaved, but the data can

readily be fit to an isothermal expansion model. This model works well using parameters obtained directly from bremsstrahlung measurements if we assume that Z/A, the ratio of atomic number to atomic mass, is approximately 1/2 (a value more typical of light ions such as carbon than of heavy ions). The transmission data are useful in evaluating target design concepts, for example, in determining what fraction of the ion energy will be absorbed, and thus exploited, by the various masses of a specific target design.

Very heavy ions (such as tantalum and gold) have appreciably shorter stopping range than protons or carbon and are thus more effective in driving the implosion of target material. An analysis of the multispecies ion expansion showed that if low-Z

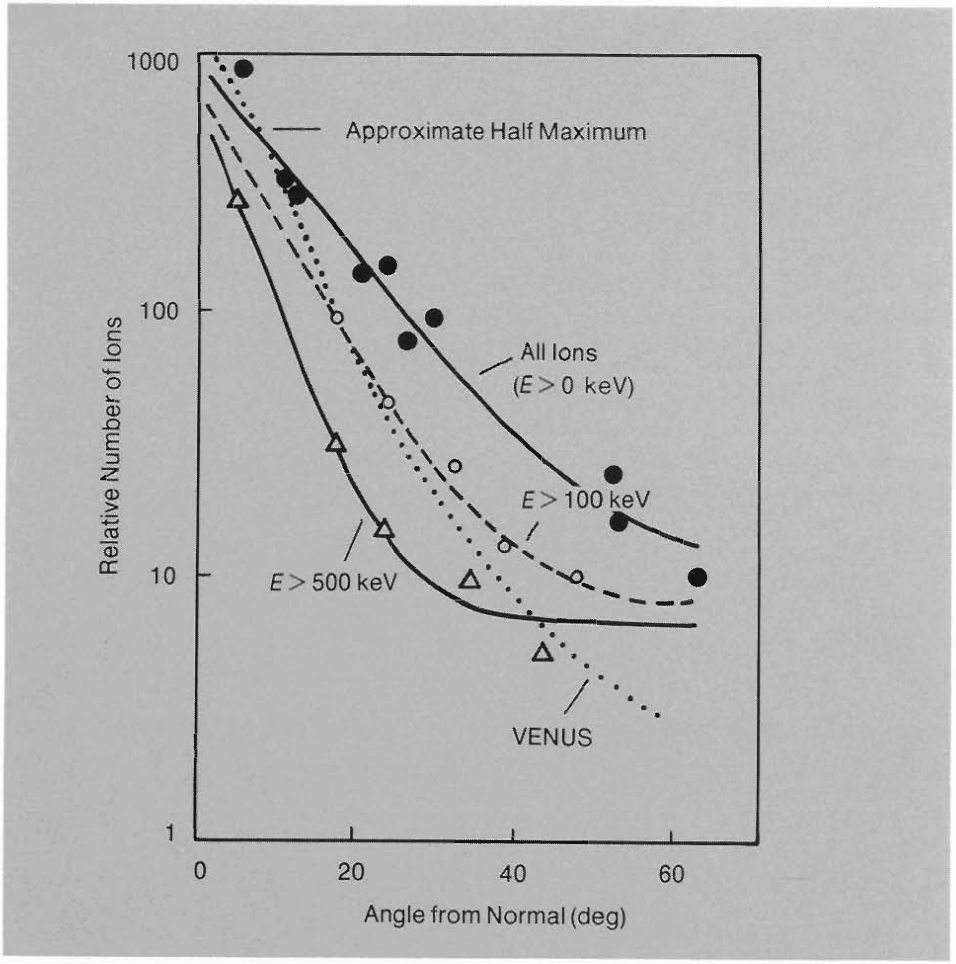


Fig. 20. These data represent the angular distribution of ions in a jet for a single shot at the Helios facility. Three sets of calorimeters were used to measure the relative numbers of ions as a function of angle, each filtered for a different energy threshold in keV per atomic mass unit. The width at half maximum of the angular distribution for all ions is about 10 degrees, and, as the energy E of the ions increases, the angular distribution narrows further. The dotted line is a prediction for the angular distribution of all ions from a VENUS simulation.

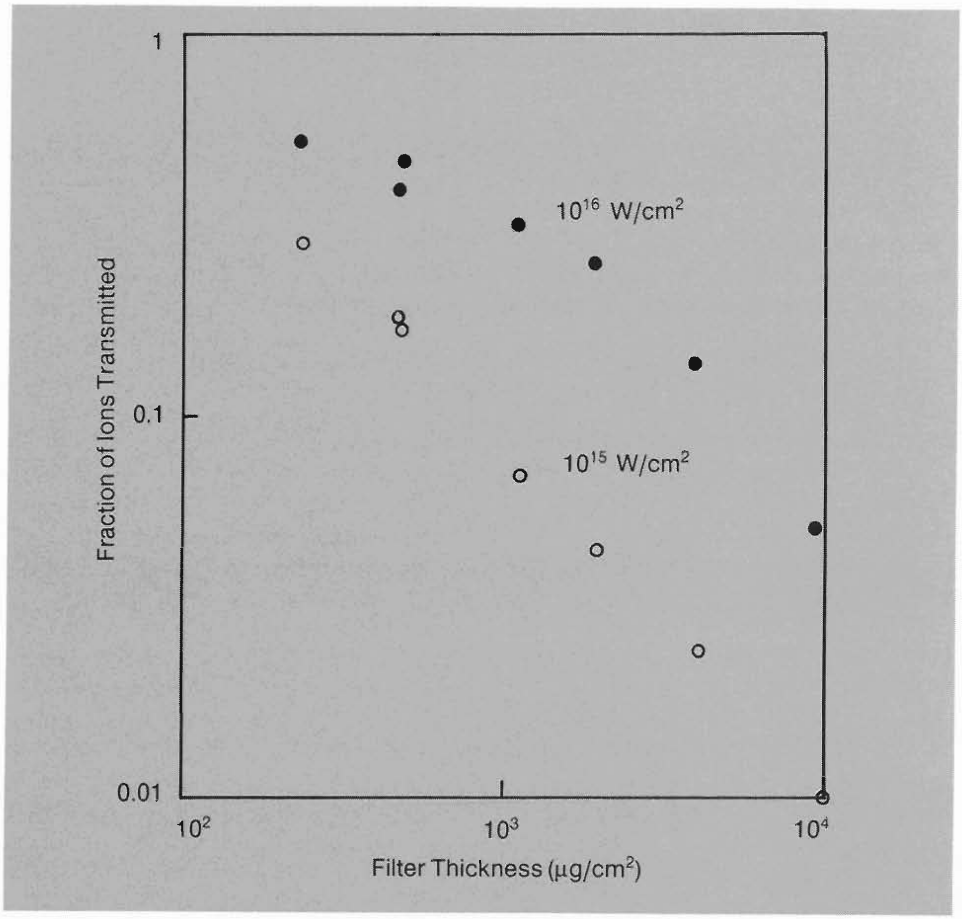


Fig. 21. Ion-range, or transmission, curves for fast ions through nickel filters. The lower intensity shot produced ions of shorter range, that is, lower sound speed, in agreement with a lower temperature in a simple isothermal expansion model.

surface contaminants could be removed, atoms of high-Z target material could be accelerated to similar velocities (the velocity of the fastest proton in a typical expansion is about 2×10^9 cm/s). In an experiment in which hydrocarbons were driven off the surface of tantalum targets by heating to white heat with electron bombardment, we observed energetic tantalum ions, the fastest of which corresponded to an energy of 500 MeV. Spectrometer data confirmed the absence of protons in the expansion and implied that the tantalum ions were as much as 60 times ionized. All this from 10-μm, or 0.1 eV, photons!

Astrophysical and Other Implications

The Antares CO_2 laser is able to generate some of the highest energy-density plasmas yet produced in the laboratory. The hot electrons themselves are equivalent to a high-current, low-impedance electron-beam source that has a current in excess of 10^8 amperes but voltages only in the hundreds of keV. This results in a source impedance of about a thousandth of an ohm, much lower than for any conventional pulsed-power source.

In part because of this low impedance, which prevents very many free electrons from escaping into free space, the intense "thermal" electron beam just described produces our familiar ion jet, which can be described as space-charge-neutralized plasma, traveling at speeds of 10^8 to 10^9 cm/s with collimation maintained over distances of at least 4 meters. In essence, the electron beam is the highest current-density particle accelerator available in any laboratory.

It has been speculated that the extreme high voltages responsible for the acceleration of ions over incredibly short distances could also be harnessed to produce a *compact*, high-voltage ion accelerator. By using a number of laser beams to strike many targets with appropriate delays between beams, a staged ion accelerator might be built in which the ions from one target are successively accelerated by subsequent targets. This process might be competititve with the beat-wave accelerator described earlier.

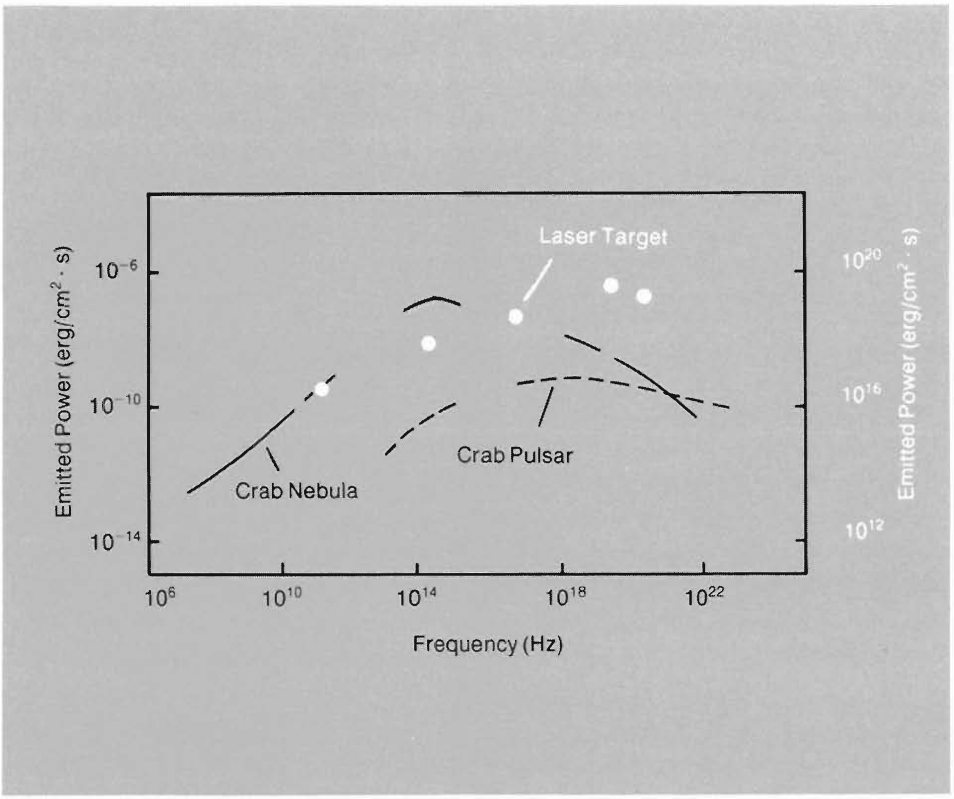


Fig. 22. A comparison of the emitted power as a function of frequency for the Crab Nebula and its pulsar and for CO_2 laser targets. The Crab Nebula spectra were adapted from "Galactic Gamma-Ray Sources" by G. F. Bignami and W. Hermsen. **Annual Review of Astronomy and Astrophysics *21(1983):67-108.)***

The ion jets themselves, although perhaps 25 orders of magnitude smaller in time and distance, are reminiscent of the galactic jets that have been so clearly identified from high-resolution radio wave photographs. In fact, an intense emission of microwaves accompanies the laboratory generation of ion jets. There is a striking similarity between the normalized electromagnetic spectrum from a typical Antares target and the equivalent spectrum from the Crab Nebula (Fig. 22). Is this similarity a coincidence or are there common features for both emissions?

Plasmas created at Helios and Antares were found to be incredibly bright in microwaves: from frequencies of 0.1 to 5 gigahertz, over a gigawatt of power was observed; from 26 to 40 gigahertz, 1 to 10 gigawatts were observed. These powers imply that about 0.01 to 0.1 per cent of the incident laser light was converted to microwaves. The source of the microwaves appears point-like, and the pulse shape, with a rise time of 0.2 nanosecond, nearly follows the laser pulse shape, so the conversion is very rapid. Such surprising emission efficiency may be due to an enhancement in the expanding corona of those plasma waves that couple efficiently to microwave emission.

A mechanism of this type is a strong candidate for explaining microwaves emitted from both galactic jets and from solar flares. Although galactic jets are detected and studied primarily from their radio or microwave emission, a wide variety of processes could generate microwaves in these extended objects. A more detailed laboratory investigation of these processes could help us discriminate between the various theories of astrophysical microwave emission.

The ability to *directly* measure in the laboratory not only the emission of electromagnetic radiation but also the *ion motion* of such objects makes study of ion jets particularly attractive and highly complementary to the elegant astrophysical measurements of galactic jets. For example, we can follow the motion of ion jets through other ambient plasmas imposed in the target chamber, learning how energetic ions propagate through ionized plasmas. Does the plume maintain its physical integrity? Is microwave emission altered as the jet strikes another plasma? Theory suggests that the emission in the submillimeter part of the spectrum will be much higher than at longer wavelengths, a fact that may help explain the match between the spectra for Antares targets and the Crab Nebula.

Also, the intense magnetic fields that help generate ion jets are known to be important for the accretion of matter in white dwarfs. Although we cannot study this process directly, we can study, on a small scale but in great detail, various acceleration mechanisms, then use our results to verify theoretical models. Such verifications would have a profound impact on astrophysical models.

And what of fusion, the initial goal of the Antares research? One very appealing idea (based on a concept that was brought to our attention by Fred Mayer of KMS Fusion, Inc.) combines features of laser and magnetic fusion. Devices that use magnetic field mirrors have difficulty maintaining stable plasmas for the necessary length of time. Theory for these devices predicts a stable plasma state at high plasma pressures. This state will have particle confinement times substantially greater than those at lower pressures. Unfortunately, conventional particle heating techniques cannot heat the plasma rapidly enough to reach this state without having the confinement destroyed by instabilities at lower pressures.

The extremely rapid heating rate of the CO_2 laser is ideally suited for creating almost instantaneously the high-pressure plasma state predicted to be stable. The direction and energy of the ions can be accurately controlled to produce just the desired conditions. Such a technique would provide both a variety of new options for achieving magnetic fusion energy and interesting combinations of phenomena important to inertial and magnetic fusion.

We have thus discovered a wealth of exotic plasma behavior at high laser intensities. These newfound phenomena allow important experiments to be performed that were not even imagined a decade ago, including bold new approaches to the intricate problem of imitating the thermonuclear burn of stars. ■

Acknowledgments

The authors wish to credit the many dedicated scientists, engineers, and technicians at Los Alamos who have developed the unique laser systems, fabricated the complex targets, developed and debugged the computer codes, devised and built the experimental instruments, and performed the creative and innovative research reflected in this article.

Further Reading

Laser Fusion:

Randolph L. Carlson, James P. Carpenter, Donald E. Casperson, R. B. Gibson, Robert P. Godwin, Richard F. Haglund, John A. Hanlon, Edward L. Jolly, and Thomas F. Stratton. "Helios: A 15 TW Carbon Dioxide Laser Fusion Facility." *IEEE Journal of Quantum Electronics* 17(1981):1662.

P. D. Goldstone, G. Allen, H. Jansen, A. Saxman, S. Singer, and M. Thuot. "The Antares Facility for Inertial Fusion Experiments—Status and Plans." In *Laser Interaction and Related Plasma Phenomena, Volume 6*, pp. 21-32, H. Hora and G. Miley, editors (New York:Plenum Publishing Corporation, 1984).

Thomas H. Johnson. "Inertial Confinement Fusion: Review and Perspective." *Proceedings of the IEEE* 72(1984):548-594.

H. Jansen. "A Review of the Antares Laser Fusion Facility." In Proceedings of IAEA Technical *Committee Meeting in ICF Research. Kobe, Japan, November 14-17, 1983*, Osaka: Institute of Engineering, Osaka University (1984) 284-298.

Plasma Physics:

D. W. Forslund, J. M. Kindel, and E. L. Lindman. "Theory of Stimulated Scattering Processes in Laser-Irradiated Plasmas." *The Physics of Fluids* 18(1975):1002-1016.

D. W. Forslund, J. M. Kindel, and E. L. Lindman. "Plasma Simulation Studies of Stimulated Scattering Processes in Laser-Irradiated Plasmas." *The Physics of Fluids* 18(1975):1017-1030.

D. W. Forslund, J. M. Kindel, and K. Lee. "Theory of Hot-Electron Spectra at High Laser Intensity." *Physical Review Letters* 39(1977):284-288.

B. Bezzerides, D. W. Forslund, and E. L. Lindman. "Existence of Rarefaction Shocks in a Laser-Plasma Corona." *The Physics of Fluids* 21(1978):2179-2185.

R. L. Carman, D. W. Forslund, and J. M. Kindel. "Visible Harmonic Emission as a Way of Measuring Profile Steepening." *Physical Review Letters* 46(1981):29-32.

W. Priedhorsky, D. Lier, R. Day, and D. Gerke. "Hard X-Ray Measurements of 10.6-μm Laser-Irradiated Targets." *Physical Review Letters* 47 (1981):1661.

D. W. Forslund and J. U. Brackbill. "Magnetic-Field-Induced Surface Transport on Laser-Irradiated Foils." *Physical Review Letters* 48(1982):1614-1617.

Fred Begay and David W. Forslund. "Acceleration of Multi-Species Ions in CO_2 Laser-Produced Plasmas: Experiments and Theory." *The Physics of Fluids* 25(1982):1675.

M. A. Yates, D. B. vanHulsteyn, H. Rutkowski, G. A. Kyrala, and J. Brackbill. "Experimental Evidence for Self-Generated Magnetic Fields and Remote Energy Deposition in Laser-Irradiated Targets." *Physical Review Letters* 49(1982):1702.

D. R. Bach, D. E. Casperson, D. W. Forslund, S. J. Gitomer, P. D. Goldstone, A. Hauer, J. F. Kephart, J. M. Kindel, R. Kristal, G. A. Kyrala, K. B. Mitchell, D. B. vanHulsteyn, and A. H. Williams. "Intensity-Dependent Absorption in 10.6-μm Laser-Illuminated Spheres." *Physical Review Letters* 50(1983):2082-5.

Allan Hauer, R. Goldman, R. Kristal, M. A. Yates, M. Mueller, F. Begay, D. vanHulsteyn, K. Mitchell, J. Kephart, H. Oona, E. Stover, J. Brackbill, and D. Forslund. "Suprathermal Electron Generation, Transport, and Deposition in CO_2 Laser Irradiated Targets." In *Laser Interaction and Related Plasma Phenomena, Volume 6*, p. 479, H. Hora and G. Miley, editors (New York:Plenum Publishing Corporation, 1984).

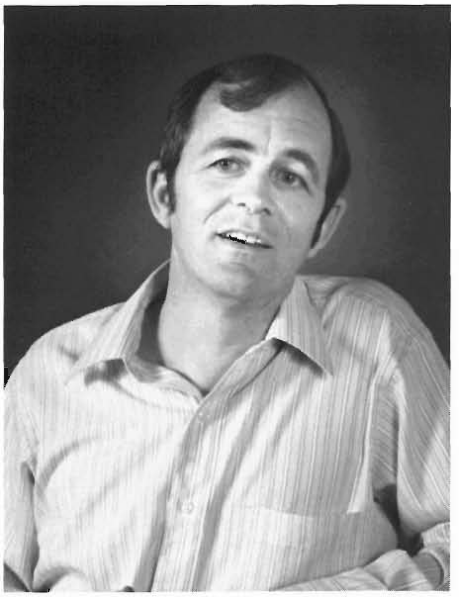
D. W. Forslund, J. M. Kindel, W. B. Mori, C. Joshi, and J. M. Dawson. "Two-Dimensional Simulations of Single-Frequency and Beat-Wave Laser-Plasma Heating." *Physical Review Letters* 54(1985):558.

Applications:

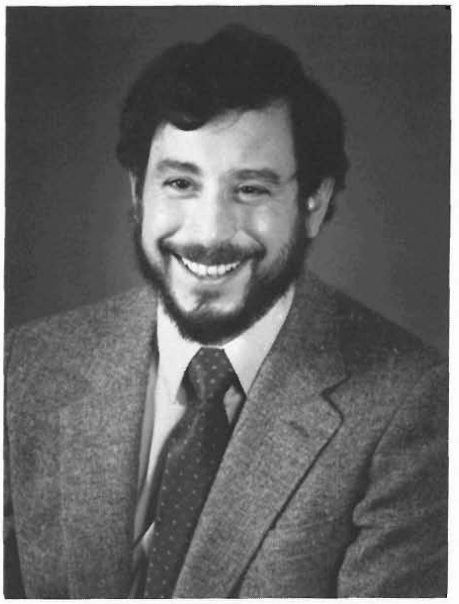
C. Joshi, W. B. Mori, T. Katsouleas, J. M. Dawson, J. M. Kindel, and D. W. Forslund. "Ultrahigh Gradient Particle Acceleration by Intense Laser-Driven Plasma Density Waves." *Nature* 311(1984):525-9.

F. J. Mayer, H. L. Berk, D. W. Forslund. "A Laser-Initiated Energetic Ion, High Beta Mirror Plasma." To be published in *Comments on Plasma Physics and Controlled Fusion.*

AUTHORS



David W. Forslund was born in Ukiah, California, on February 18, 1944. He received his Bachelor of Science from the University of Santa Clara in 1964 and his Ph.D. in Astrophysical Sciences from Princeton University in 1969. Following two years as a postdoctoral fellow in space physics at Los Alamos, he joined the Magnetic Fusion Theory group and then, in 1972, the Inertial Fusion Theory group with which he has been connected ever since. He served as associate group leader and deputy group leader before being selected as a Laboratory Fellow in 1981. He has made important contributions to space plasma physics, magnetic fusion, high-altitude nuclear effects, and laser plasma interactions as well as numerical simulation techniques for plasmas. A co-recipient of one of the Laboratory's Distinguished Performance Awards in 1982, he is also a member of the American Geophysical Union and the American Astronomical Society and is a Fellow of the American Physical Society. He is active in his local church, serving as a ruling elder, and is interested in Christian education, serving as president of the board of Covenant Christian School.



Philip D. Goldstone received his B.S. and M.S. degrees from the Polytechnic Institute of New York in 1971 and 1972, respectively, and his Ph.D. in 1975 from the State University of New York at Stony Brook. One result of his doctoral research was the discovery of a number of new resonances in the sub-barrier fission of actinide nuclei. Joining Los Alamos in 1976 as a postdoc in the Nuclear Physics Group, he studied problems in fission and heavy-ion collisions. In 1977 he joined the fusion effort in the Laser Division and shortly became the Operations Manager of the Gemini CO_2 laser. A year later, excited about the possiblities for laboratory physics research using the incredible power densities available with lasers, he began work on laser-driven shock waves, the physics of radiation-driven ICF targets, and the assessment of CO_2 lasers as fusion drivers. Since 1981 he has been Group Leader of the Fusion Experiments and Diagnostics Group in the Physics Division. His current interests include laser acceleration of particles, the physics of highly stripped ions in intense laser fields, and, of course, the dense plasma physics and hydrodynamics of ICF. He has lived in the Jemez Mountains since 1978 with his wife, Joyce, and has been active in the La Cueva volunteer fire department and other community efforts.

KrF Laser: the advance toward shorter wavelengths

by Reed J. Jensen

If energy is added to water in a swimming pool by using wavelengths comparable to the dimensions of the pool, water quickly peaks and sloshes over the edge. If, however, shorter wavelengths are used, the water, though rippling with motion, remains in the pool. There is an analogous problem when one uses laser energy to drive the implosion of a small fusion pellet. If longer wavelengths, say in the infrared, are used, then a portion of the energy "sloshes over" into undesirable modes (such as a few very hot electrons) that dissipate rather than drive the implosion. As our understanding of the physics of laser fusion has increased, awareness of the importance of fusion drivers with shorter wavelengths has likewise increased.

However, building an efficient, high-intensity laser that emits short-wavelength photons is a difficult balancing act for a number of reasons. The balancing becomes obvious when we look at the expression for laser gain. In a simple two-level laser, the gain coefficient g obeys the relationship

$$g \propto \lambda^2 / t_{\text{spont}} ,$$

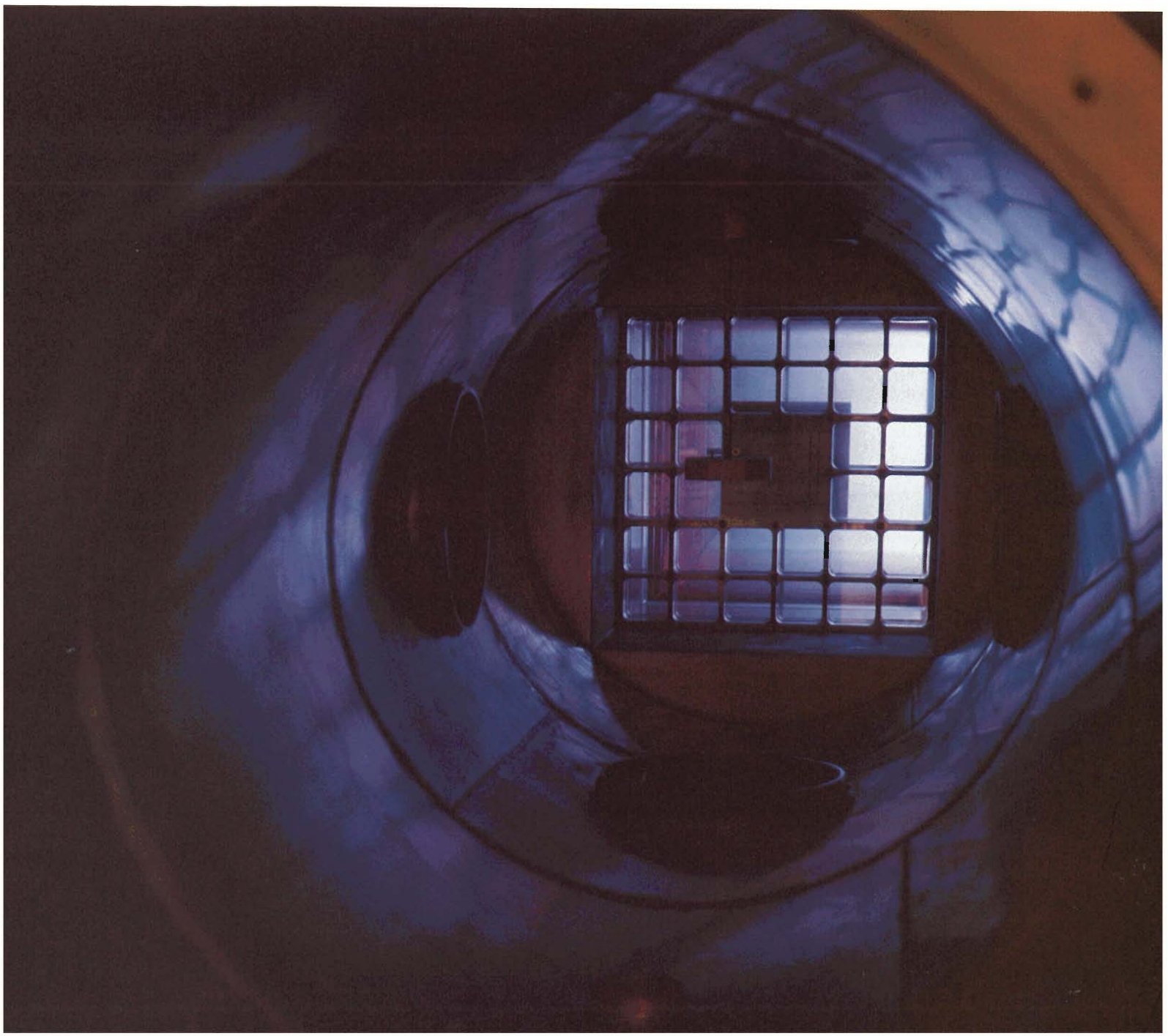
in which t_{spont} is the lifetime of the excited state against spontaneous emission and λ is the wavelength of the emitted photons. The gain itself is proportional to the factor e^g.

As we go to the high excitation levels needed for shorter wavelengths, the lifetimes of excited states tend, in general, to become short. This means the gain coefficient increases—an apparent advantage (although transition strengths can reverse this trend). But a short lifetime also means the excited atoms spontaneously emit their energy quickly. This latter fact is a disadvantage because the emitted photons stimulate further emissions, resulting in a phenomenon called *amplified spontaneous emission.* In large-volume systems, such emission is parasitic, draining energy away too quickly and reducing system efficiency. In other words, it becomes difficult to store large amounts of energy in the lasing medium.

Because of the factor λ^2, this trend can be resisted by locating a system that emits short-wavelength photons from an excited state with a moderate lifetime. For example, a system that emits at a wavelength of 0.1 micrometer (μm) can have a spontaneous emission lifetime ten thousand times shorter than that of the CO_2 laser and still have about the same gain.

Promising laser systems exist in the mid-ultraviolet, but in this region technical difficulties with optics and windows add to the difficulty of the balancing act. For example, common optical materials absorb strongly at wavelengths shorter than 0.24 μm.

A large KrF amplifier, the LAM, is discharged to generate ultraviolet power at a wavelength of one-quarter of a micrometer. *(Photo by Fred Rick.)*



We are currently developing at Los Alamos the krypton fluoride (KrF) laser—a system that balances these properties to yield a highly efficient laser able to emit intense bursts of short-wavelength photons. For example, the KrF laser operates at ¼ μm, close to the short-wavelength limit for optics but, fortunately, just on the conventional-optics side. The excited-state lifetime of the system is short—due both to spontaneous emission and to deactivation from collisions—making it impossible to store significant energy in the lasing medium. Counterbalancing this disadvantage is the laser's high gain, which yields a system able to amplify efficiently a rapid series of short pulses. As a result, considerable energy can be stored *outside* the laser during the beam's flight to the target. Such storage is accomplished by using a novel multiplexing scheme in which time-of-flight differences cause the series of pulses to meet simultaneously at the target. Thus we are developing a system that will be able to generate short, intense pulses of ¼-μm light highly desirable for the study of the physics of laser fusion at short wavelengths.

An Overview of the Laser

In early 1975 J. E. Velazco and D. W. Setser at Kansas State University suggested using a mixture of krypton and fluorine gases as a lasing medium, and in June 1975 J. J. Ewing and C. A. Brau at Avco Everett Research Laboratory reported the first laser oscillation for the highly efficient KrF system. The potential energy curves shown in Fig. 1 illustrate the major reason for the high efficiency. The ground state of the KrF system is really a trio of states ($X\ ^2\Sigma_{1/2}$ and $A\ ^2\Pi_{1/2,\ 3/2}$) that are covalent but repulsive, and the "molecule" readily dissociates into neutral krypton and fluorine atoms. As a result, there is no accumulation of molecules in the lower laser level, and a significant population inversion can be achieved. The main laser transition occurs between the $B\ ^2\Sigma_{1/2}$ excited state, made up of an attractive Kr^+F^- ion pair, and the $X\ ^2\Sigma_{1/2}$ ground state.

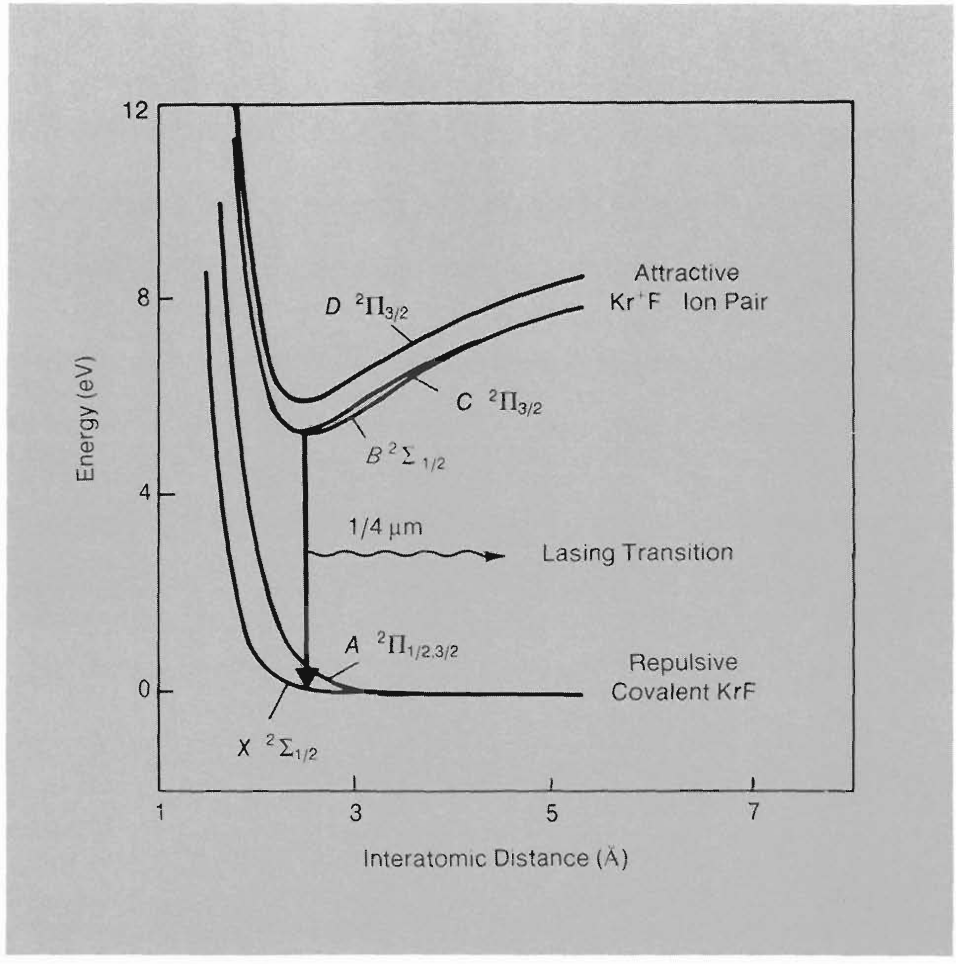


Fig. 1. The potential energy diagram for the KrF laser. The upper state of the lasing transition ($B\ ^2\Sigma_{1/2}$) is one of the states for the Kr^+F^- ion pair. The lower state ($X\ ^2\Sigma_{1/2}$) is one of two covalent states that are repulsive and so dissociate, eliminating KrF molecules from the lower level and ensuring a large population inversion. (Adapted from P. J. Hay and T. H. Dunning, Jr.,* Journal of Chemical Physics *66 (1977):1306.)

This transition can be pictured as the ion pair reverting back to the dissociating molecule by transferral of an electron from F^- to Kr^+ and emission of a photon with a wavelength of ¼ μm.

Gain. Measured gain coefficients for the KrF laser are in excess of 10 per cent per centimeter, so that a 1-meter amplifier would have a gain of e^{10}, or about 20,000, per pass through the amplifier. As pointed out above, such huge gains will not allow storage of large amounts of energy in the lasing medium as a population inversion; the excessive gains cause a loss of upper-level population by amplified spontaneous emission. By using a multiplexing scheme, however, we are able to extract the energy as a series of short

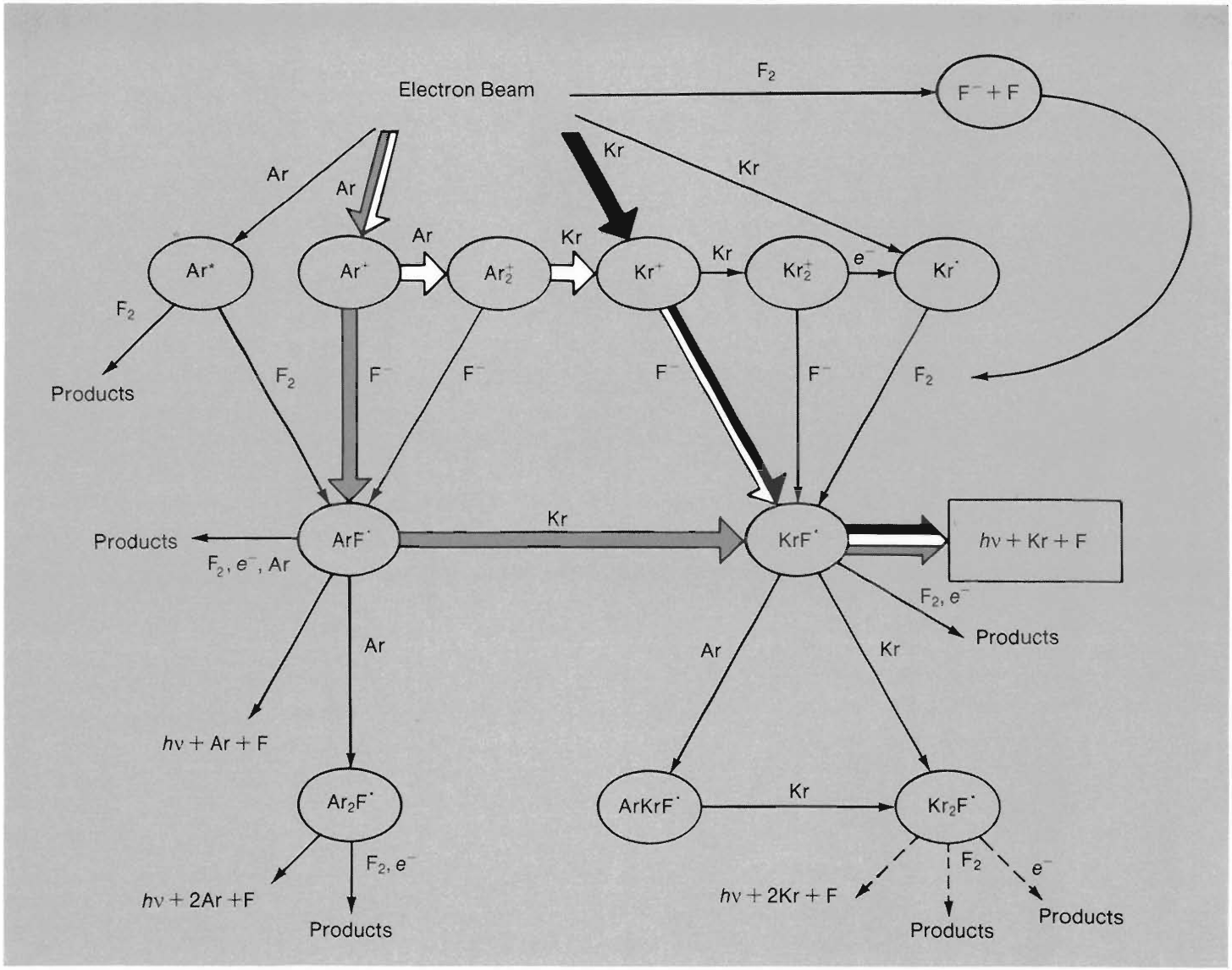


Fig. 2. Kinetic pathways for the KrF laser. Initially, all three atoms in the Ar-Kr-F_2 gas mix are ionized, but the most important path (white) to the KrF ion pair (KrF^*) passes through Ar^+ and Ar_2^+ intermediates before a krypton atom is ionized. The Kr^+ ion then combines with a previously ionized fluorine atom. The second most important path (gray) also starts with Ar^+ but then forms on ArF ion pair (ArF^*) before exchanging with a krypton atom. The path in black is energetically favorable because krypton ionizes easily but becomes important only as the amount of krypton gas in the mix is increased.

pulses *while* the laser is being pumped. This extraction suppresses the gain by reducing the gain coefficient according to the equation

$$g = \frac{g_0}{1 + \frac{I}{I_{sat}}},$$

where g_0 is the gain coefficient at low intensities (the small-signal gain), I is the laser intensity in the medium, and I_{sat} is the so-called saturation intensity (an intensity based on photon energy, the stimulated emission cross section, and the upper-state lifetime of the lasing medium).

It is apparent that if energy is extracted from the amplifiers while the laser is running at an intensity three times I_{sat}, there will be a fourfold decrease in the actual gain coefficient. This reduction will bring the system gain down from about twenty thousand to a few hundred.

Gas Kinetics. In a practical sense, the efficiency possible in the system is determined by how atoms are pumped into the upper laser level. Typically, the energy is added by ionizing a gaseous mixture that includes argon. The argon atoms play an active role by forming intermediate species, as shown in Fig. 2, a simplified flow chart for the gas kinetics of the Ar-Kr-F_2 system. Initially, all three types of atoms are ionized (here by using an electron beam), but the path (shown in white) that produces the largest population of KrF ion pairs (KrF^*) in the upper laser level involves Ar_2^+. The second most important path (shown in gray) involves the ArF ion pair (ArF^*). Once KrF^* is reached, the desired exit channel is, of course, dissociation to Kr and F and emission of a photon.

The efficiency for conversion of pump energy to upper-state population could conceivably be increased by using more direct or lower energy channels to KrF^*. In fact, im-

mediate gains in efficiency can be realized by increasing the amount of krypton in the mixture, although this gas is more costly. Such an approach would emphasize the path (shown in black) in which krypton is ionized directly—a more efficient route because of krypton's lower ionization energy.

Undesirable Absorption. Many of the species pictured in Fig. 2 absorb light of the KrF lasing frequency. What constraint does this place on the design of the KrF laser? Generally, for any absorbing species there is an intensity at which the absorption saturates—that is, the light overpowers the absorption process and starts to pass freely through the system. The fluence for this condition is given by

$$\varphi_{sat} = \frac{1}{\tau\sigma} ,$$

where τ is the relaxation time and σ is the absorption cross section for the absorbing species.

Unfortunately, some of the absorbing species in the KrF laser gas have low values of $\tau\sigma$ and do not saturate at fluences used for the design of this laser. The system, therefore, has a practical upper limit on the growth of fluence in the lasing medium. In other words, we have another balancing task; as intensity or fluence increases, the gain saturates but absorption by other species does not. It is believed that the limiting fluence for the KrF system is about 10 to 130 megawatts per square centimeter. Greater intensities can only be reached by using optical focusing and other techniques that are applied beyond the last amplifier.

Electron Beam. Ionization of the laser gas mix also has its difficulties. The gas contains fluorine, a halogen, which is an efficient electron scavenger. In fact, the formation of F^- by electron attachment plays a key role in the pumping scheme of Fig. 2. However, standard gas-discharge techniques used with many gas lasers generate electrons with energies of a few tenths of an electron volt

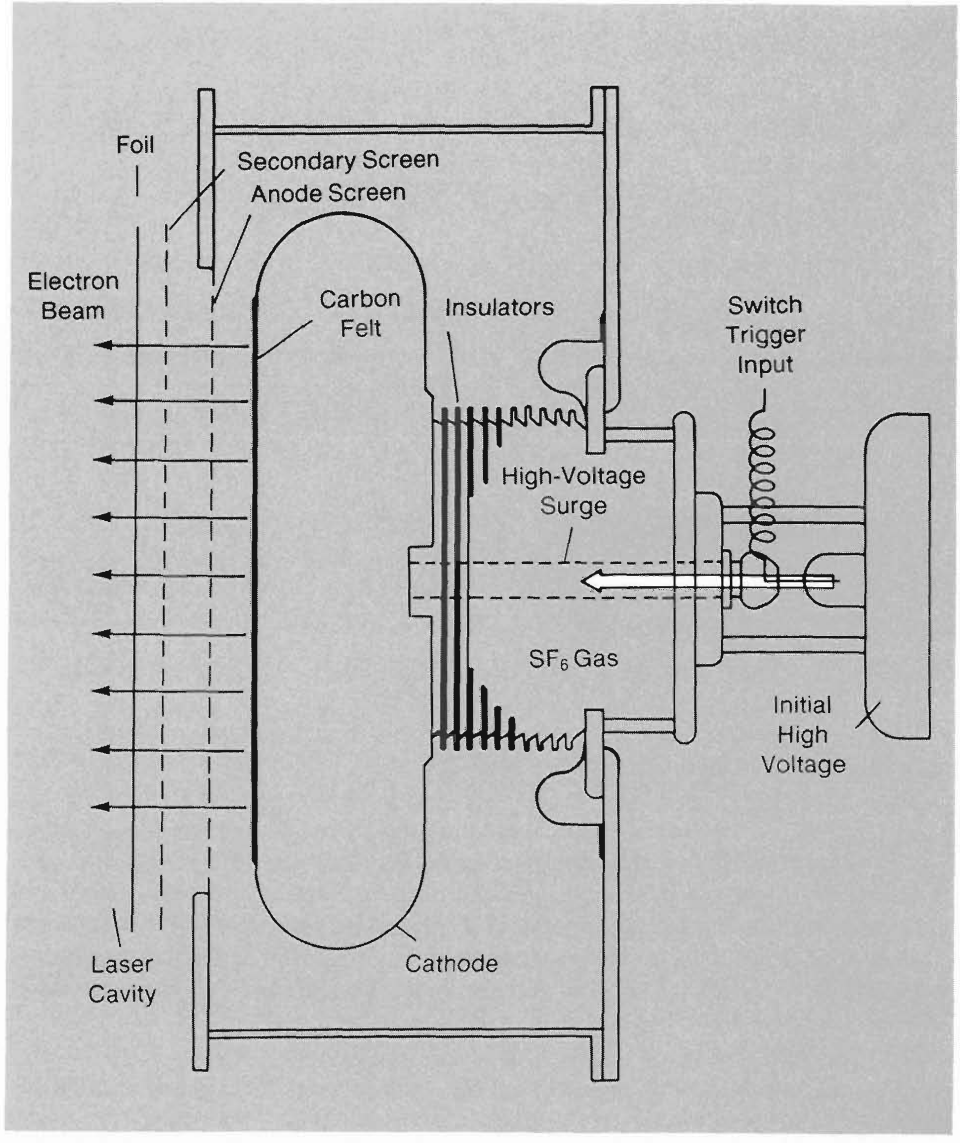


Fig. 3. A schematic of the diode for generating the electron beam. When the switch trigger is activated, high voltage surges from right to left on the cathode, where electrons are ejected toward the anode screen. These electrons also flow through a screen that eliminates stray secondary electrons and then through 2-mil titanium foil into the volume of the laser cavity. The electrons are emitted at energies of several hundred keV, and about 95 per cent are transmitted through the foil, where they uniformly excite the laser gas. Critical to the design is the cylindrical high-voltage bushing between the switch and the cathode. This bushing of insulators and SF_6 prevents undesirable arcing back toward the switch. To give an idea of the size of these electron-beam devices, the emitting surface of one of the cathodes in our recent large KrF amplifier (the Aurora system) measures 1 by 2 meters.

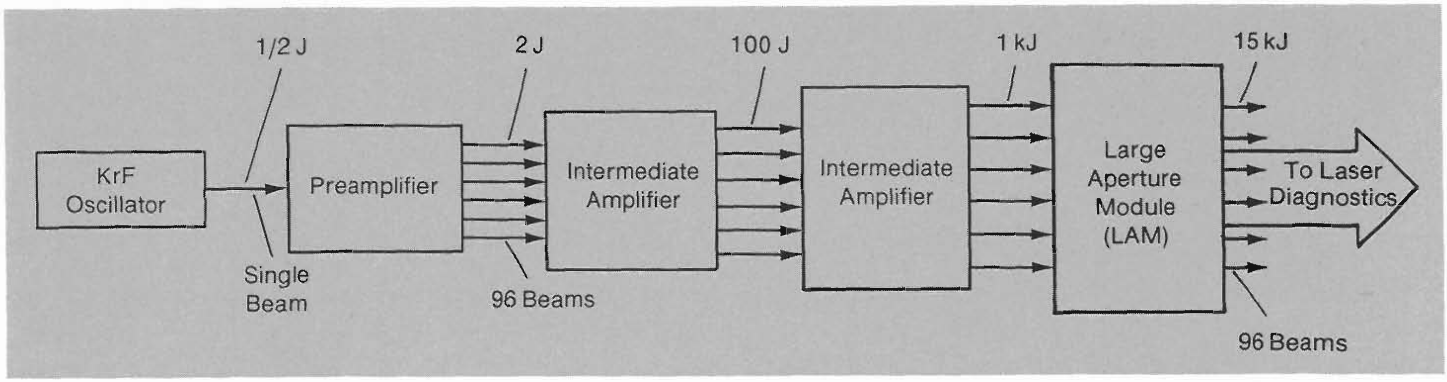


Fig. 4. The power amplifier chain for the Aurora KrF laser system starts with a single ½-joule beam and ends with about 15 kilojoules divided among 96 beams.

(eV). Because attachment energies are also of this order, such a discharge has trouble sustaining itself if electron scavengers are present.

Further, in a standard ionization-stabilized discharge system, there is a tendency for the discharge to constrict into a few very hot arcs. This phenomenon leaves most of the gas unpumped, whereas the gas in the vicinity of the arcs is far overpumped and overheated. Even though a preionization device may help generate more homogeneous current flow, the discharge still tends to constrict into hot arcs because of the large positive temperature dependence of the coefficient β in the electron replenishment equation,

$$e^- + M \rightarrow M^- + \beta e^- .$$

In other words, the number of secondary electrons generated when a primary electron collides with atoms in the medium increases rapidly with temperature, and any area of enhanced conduction quickly forms an arc.

The solution to these problems, which would be especially severe at the higher intensities required in amplifiers, is to ionize the gas directly by pumping with an electron beam. In this case, the electrons enter the laser volume with energies of several hundred keV, and any difficulties with electron attachment are simply overpowered.

In addition, an electron beam device avoids runaway arcs by injecting electrons directly into the gas through thin foil windows (Fig. 3). Electron homogeneity is established by field emission and propagation physics in the vacuum region around the cathode. The resulting even distribution of the primary electron current from the electron beam dominates all physics in the laser volume. Thus, the primary current in any particular volume of gas does not depend upon the conductivity of that gas.

Typically, electrons enter the laser volume at energies of 400 to 800 keV. In primary collisions, gas ionization generates secondary electrons with an energy loss to the primary electron of about 30 eV per ionization. Thus each primary electron requires thousands of collisions to deposit its energy in the gas. Both primary and secondary electrons participate in the various pumping and attachment processes shown in Fig. 2.

With this pumping technique we find that for Ar-Kr-F_2 mixes about 24 per cent of the energy of the electron beam appears as population of the upper laser level. With more Kr in the gas mixture and using proper pumping and extraction techniques, the efficiency should be higher. Once again, we see that the KrF laser has high intrinsic efficiency. The exact value depends, of course, on the conditions of the gas and the pumping circuit. So far, an overall efficiency from wall plug to laser beam of 4 per cent has been achieved. Such efficiency is still a little less than is achieved for the CO_2 laser, but, even so, more of the energy will be absorbed by the fusion target (about 90 per cent for KrF versus 40 per cent for CO_2) and considerably more should end up driving the implosion.

High Repetition Rate. Because KrF is a gas laser with relatively high efficiency, it can be pulsed at high repetition frequencies. To prepare for a subsequent pulse, the gas needs time for all intermediate fluorides to revert back to the elements and for electron-beam-induced shock waves to damp out. This is achieved by exchanging the gas between pulses with modestly sized pumps. Also, spurious impurities, formed by fluorine attack on surrounding materials, are removed during the exchange with filters and "getters" (metals, such as sodium or calcium, that react with the impurities). With these techniques, KrF lasers have been operated at repetition frequencies greater than a kilohertz. Such high repetition rates are the key to the eventual successful development of a laser-driven fusion power plant.

The Los Alamos Program

Los Alamos has been heavily involved in the development and use of KrF and other rare-gas-halide lasers since 1976. The work was done originally for the laser isotope-separation program, and in the first year KrF lasers were used to generate macroscopic samples of photolytic UF_5 from UF_6. At that point we were developing high beam quality, multijoule gas-discharge (rather than electron-beam) lasers. In the late 1970s Los Alamos pioneered gas cleanup schemes that paved the way to the long-lived rare-gas-halide lasers. These efforts culminated in a KrF laser that ran at 1 joule per pulse and 500 pulses per second. We also developed, jointly with Rocketdyne, another KrF laser that had double this repetition rate.

At present we are constructing a prototype KrF laser system that uses electron-beam pumping in the amplifiers and that will demonstrate the production and extraction of laser energy and the optical techniques needed for a KrF fusion laser. The system is called Aurora, and its various components—from the KrF oscillator through four stages of amplification—are shown schematically in Figs. 4 and 5. The gas-discharge oscillator will emit a single beam consisting of a 5-nanosecond, ½-joule pulse. To achieve optical multiplexing, the original beam will be split into 96 beams before being sent through the various stages of amplification. The final amplifier, called the Large Aperture Module (LAM), will have a lasing volume that is 1 meter by 1 meter by 2 meters long and an output totaling about 15 kilojoules (kJ) in the 96 beams.

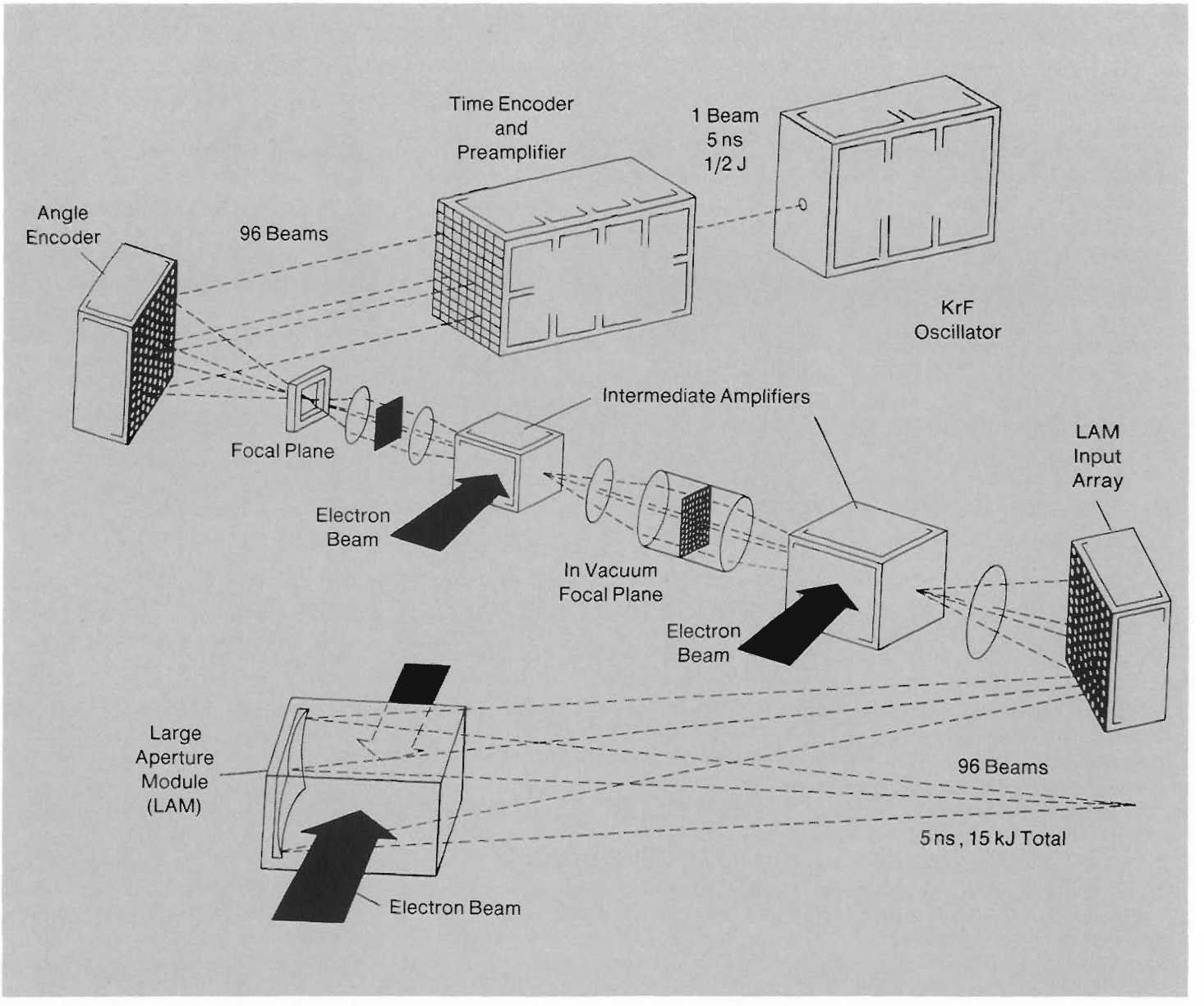


Fig. 5. A conceptual layout for the Aurora laser system. The single beam with 5-nanosecond pulsewidth from the KrF oscillator is divided by the time encoder into a train of 96 temporally separated beams. This splitting is accomplished with aperture dividing, partially reflective mirrors called beam splitters, and different path lengths for each beam. Preamplification also takes place in the same apparatus. The angle encoder aims each beam so that it will pass through the central region of both intermediate amplifiers. Final amplification to 15 kilojoules takes place in the large aperture module. The oscillator is driven by gas-discharge techniques, but the higher intensities in the amplifiers require pumping by electron beam. A demultiplexing arrangement is needed after the LAM to bring the 96 beams to the target simultaneously (see Fig. 6).

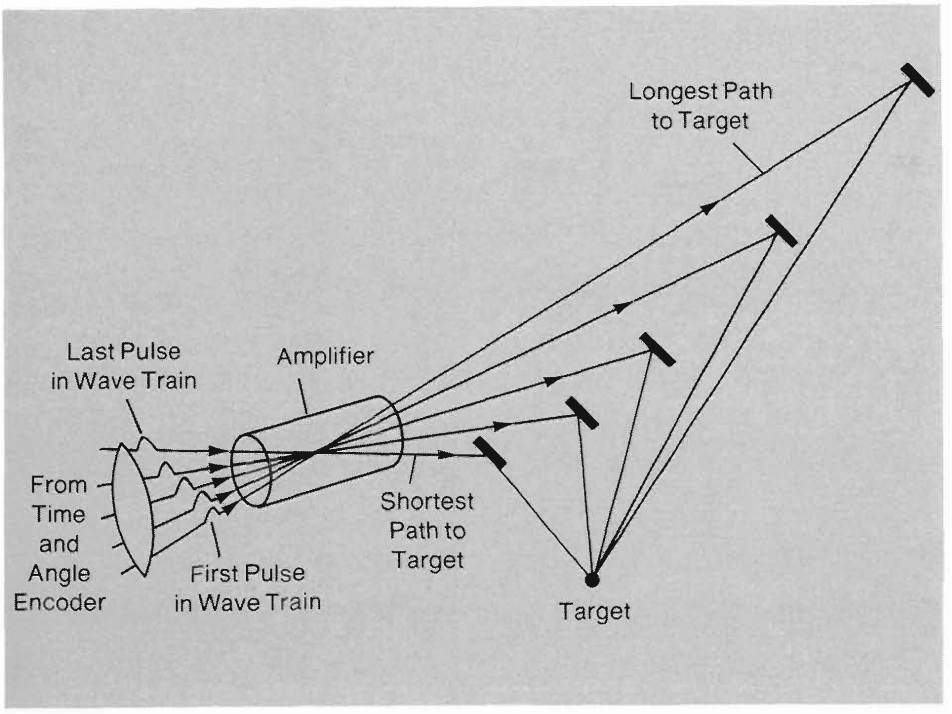


Fig. 6. A simplified optical angular multiplexing device. The five beams from the decoder represent a train of pulses that are separated in time. By adjusting path lengths so that the earliest pulse (crossing from bottom left to upper right) has the longest time-of-flight and the last pulse (crossing horizontally) has the smallest time-of-flight, the pulses can be brought together at the target simultaneously.

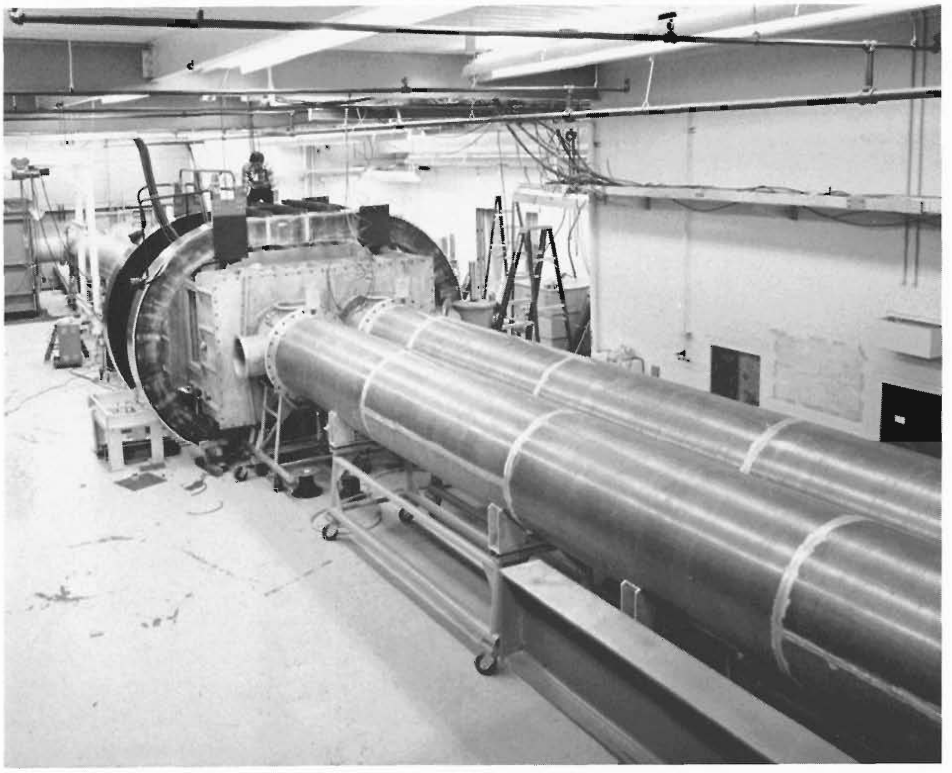


Fig. 7. The KrF laser system's final amplifier, the LAM, under construction.

Multiplexing. A prime motive for optical multiplexing is cost. It has been estimated that a cost minimum can be achieved for KrF laser fusion if there are from 50 to 100 beams, each less than 5 nanoseconds in duration, derived from a single electrical pulse. Rather than build 100 systems, we intend to use a single optically multiplexed system—a key concept to be demonstrated by the Aurora laser.

Multiplexing starts with the time and angle encoders (Fig. 5), which split the original 5-nanosecond pulse into 96 angularly separated beams. Each beam travels a different distance so that each is delayed differently; the resulting output consists of a train of 5-nanosecond pulses. Because the pulses are angularly separated, each passes through the amplifiers from a slightly different direction. The amplifiers are pumped for a relatively long time—about 600 nanoseconds—while the train of short pulses traverses their volumes.

The same time-delay concept can be applied in reverse after the final amplification to cause all the beams to arrive at the target at once. A simplified version of this part of the system is shown in Fig. 6, illustrating how the time-of-flight for each beam differs so that all beams meet simultaneously at the target.

As noted earlier, the multiplexing technique allows us to take advantage of the high gain and high efficiency of the KrF gas laser to generate a short, high-energy pulse on target. The energy of the electron beam discharge is stored in the variously delayed flights of the 96 beams. Moreover, low-cost laser energy is provided by using one system in which the amplifiers run for a relatively long time rather than by using many short-pulse systems.

Figure 7 is a photograph of the final amplifier under construction. The large oval-shaped features are magnets that provide a 3-kilogauss guide field to direct the electron beam straight into the laser chamber. The two large cylindrical tubes are water dielectric transmission lines that transmit the 1.3-megavolt electrical pulse to the cathode. One

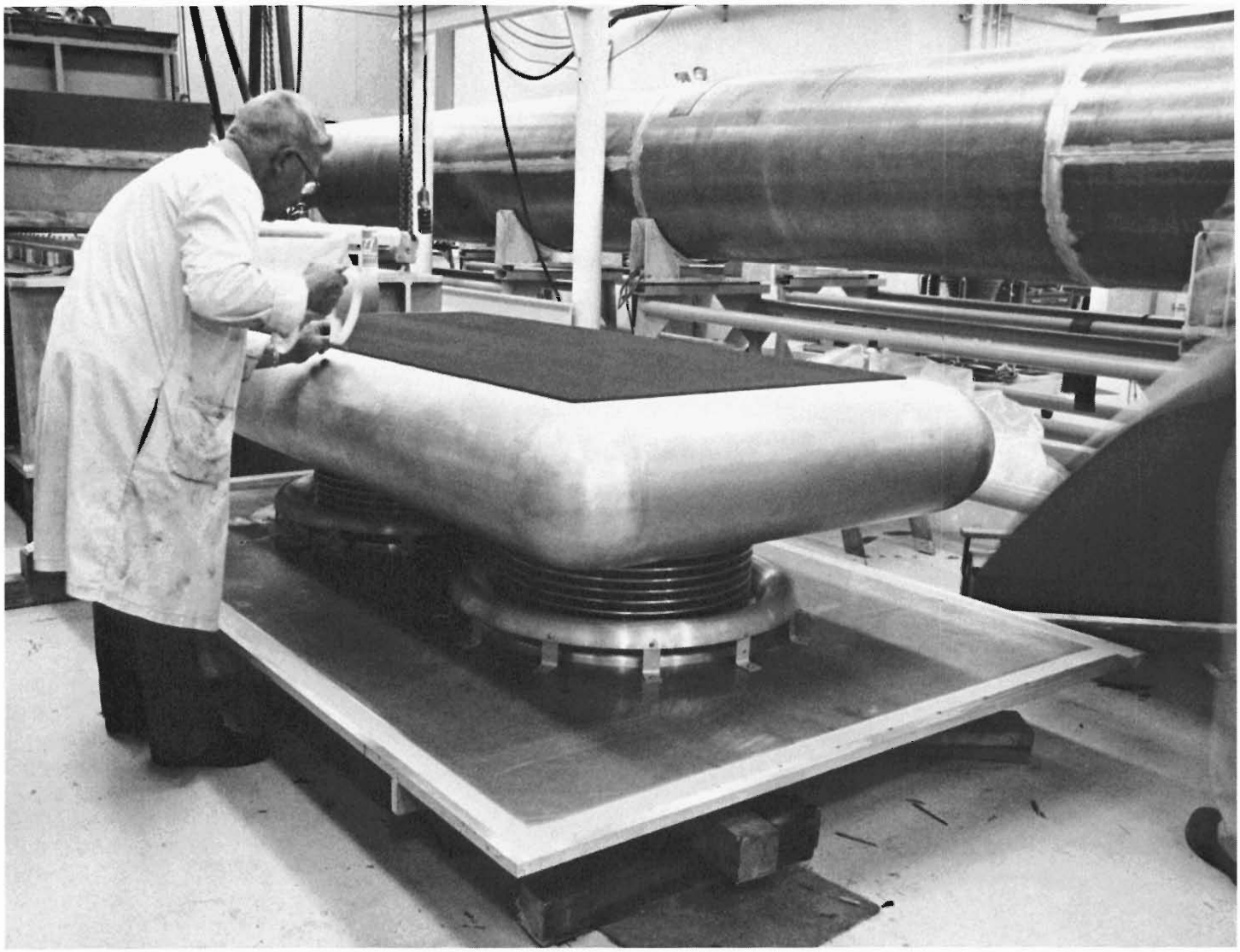


Fig. 8. One of the electron-beam cathodes for the LAM with the emitting surface upright.

of those cathodes is pictured in Fig. 8 lying on a workbench with the carbon felt emitter surface upright. Discharge of this amplifier produces ultraviolet power like that shown in the opening photograph.

The Aurora laser system will provide experience in nearly all of the issues involved in building a very large KrF laser fusion driver. We will gain experience not only in large KrF amplifier construction and operation but also in running a whole series of amplifiers with final flux close to the limiting flux for the system. A major issue in this, or any, large KrF laser system is damage to the windows and mirrors. We must develop coatings with good reflective or transmissive properties that also are resistant to fluorine attack and optical damage. At present, the size of apertures and, therefore, the overall system cost—depends very sensitively on the threshold for optical damage.

Results. Recent key results from Aurora include operation of the large final stage amplifier (LAM) at partial charge voltage. We were able to produce 5 kJ of high-quality laser output at the amplifier. This measure-

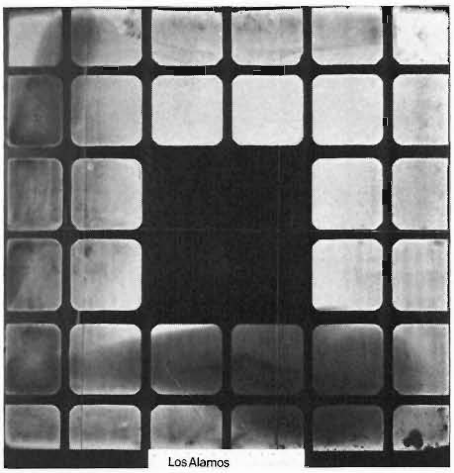
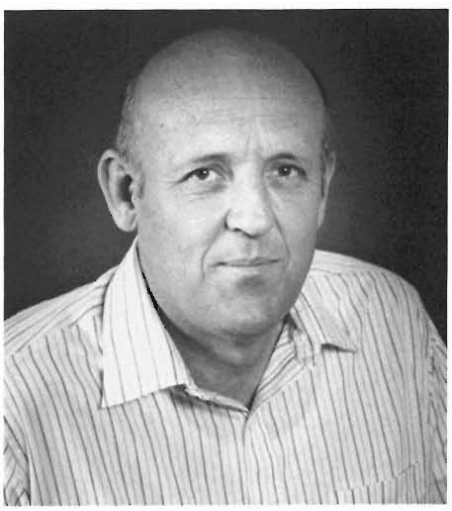


Fig. 9. A burn pattern produced by the LAM running as an oscillator. About 3 kJ of light produced this burn.

ment was made running the LAM as a resonator with a long build-up time rather than as an amplifier. The 5-kJ output was developed during the last 400 nanoseconds of the pump pulse. A typical burn pattern is shown in Fig. 9. The preliminary tests indicate that the goal of a 10-kJ output from Aurora should be exceeded when all of the preamplifiers are in full service.

The automatic alignment system needed to direct the 96 individual beamlets through the amplifier system to a target is based upon beam position sensing. Our technique uses rather ordinary television technology linked to inexpensive small computers. Key tests of the concept have been performed, and we expect to align all the beams simultaneously in a matter of minutes as part of our shot preparation procedure.

When we have successfully operated the complete Aurora system, most of the key issues in the use of efficient KrF lasers for fusion will have been resolved. We will then outfit the Laboratory's existing laser fusion facility (Antares) with the KrF ultraviolet laser system to begin to explore the promising realm of short-wavelength laser-driven fusion. ■

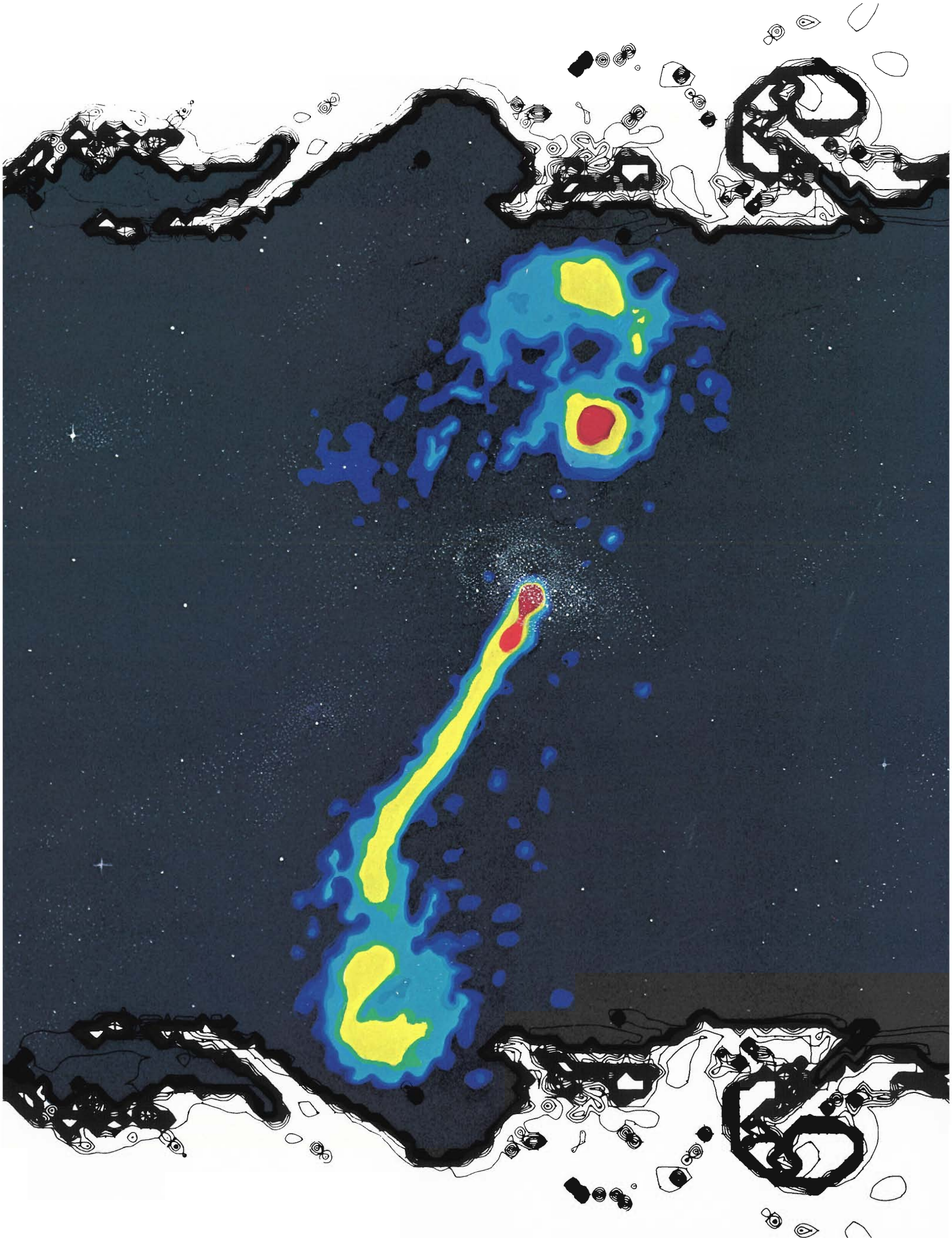


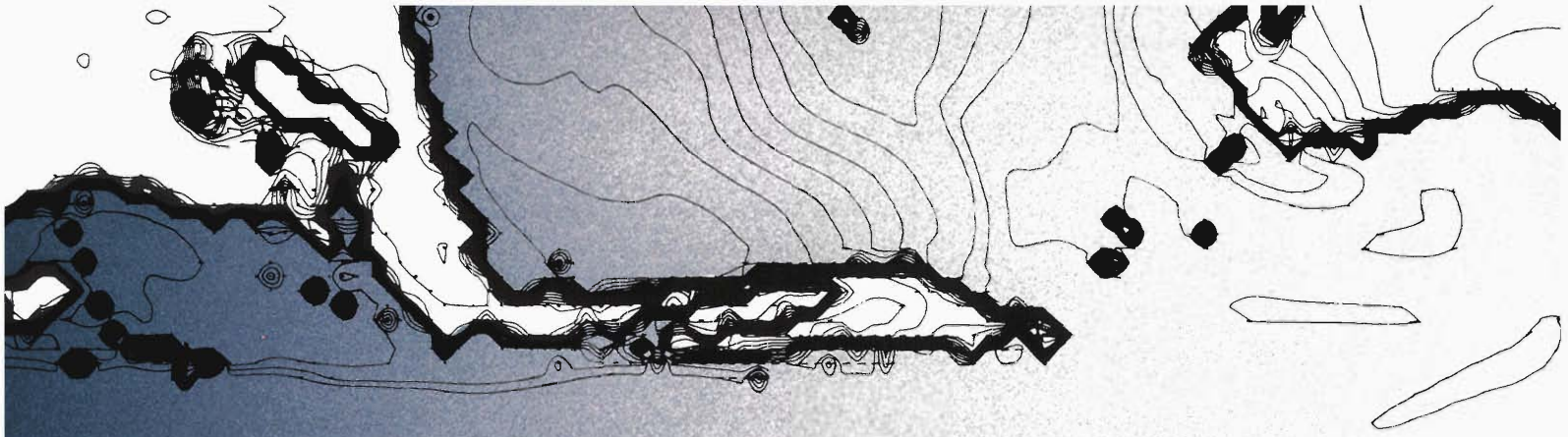
Reed J. Jensen received his Ph.D. in physical chemistry from Brigham Young University in 1965. After a postdoctoral appointment at the University of California, Berkeley, he joined the Laboratory's GMX Division. Following a two-year teaching experience at Brigham Young University, he returned to Los Alamos to initiate a program in chemical laser research. He played a key role in advancing the technology of high-energy pulsed chemical lasers. His experience with the interaction of laser radiation and chemical kinetics led him directly into the field of laser isotope separation. During this period he, along with several coworkers, advanced the concept for molecular laser isotope separation. His promotion to the post of Assistant Division Leader of the Laser Division was followed in 1976 by an appointment as Alternate Division Leader of the newly formed Applied Photochemistry Division. In 1981 he was named Division Leader of the Applied Photochemistry Division and Program Manager for molecular laser isotope separation. Under his direction approximately 225 scientists and technicians pursued research and development in laser isotope separation, laser-induced chemistry, applied photochemistry, spectroscopy, and related fields. In 1983 he became Associate Physics Division Leader and Program Manager for advanced fusion lasers. Current work includes efficient scaling of lasers to giant pulses for laser fusion.

Further Reading

J. E. Velazco and D. W. Setser. "Bound-Free Emission Spectra of Diatomic Xenon Halides." *Journal of Chemical Physics* 62 (1975):1990-1991.

J. J. Ewing and C. A. Brau. "Laser Action on the $^2\Sigma^+_{1/2} \rightarrow {}^2\Sigma^+_{1/2}$ Bands of KrF and XeCl." *Applied Physics Letters* 27 (1975):350-352.





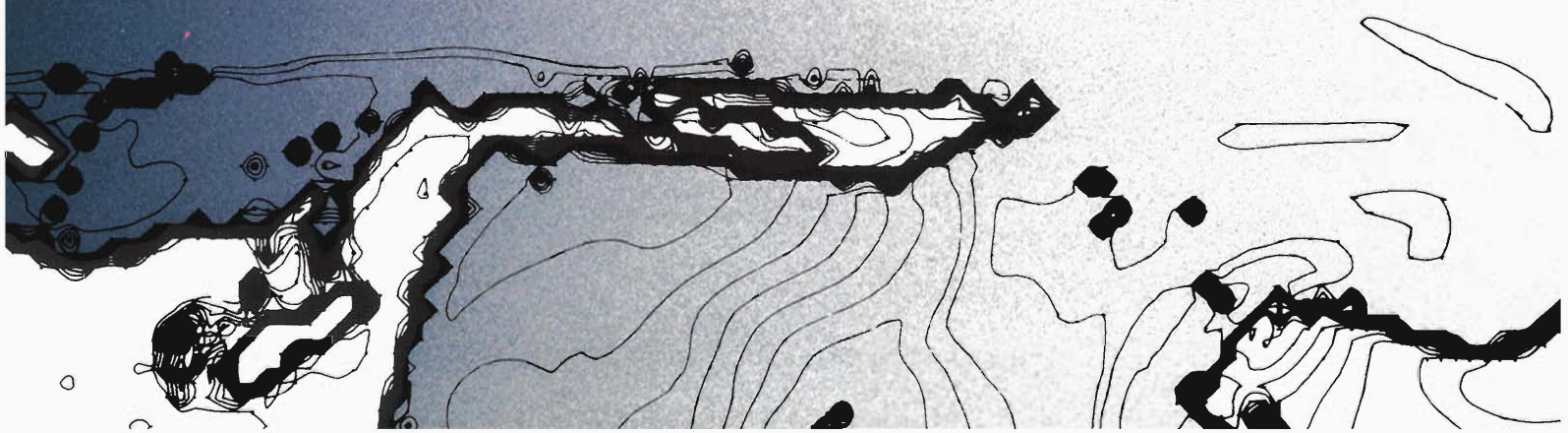
SUPERSONIC JETS

Nonlinear features of supersonic jet flow may explain the mysterious stability of extragalactic jets.

by Michael L. Norman and Karl-Heinz A. Winkler

To the man on the street, a supersonic jet is an aircraft that travels faster than the speed of sound and rattles windows and teacups with its sonic boom. In fact, he is doubly correct. Every aircraft like the Concorde or the F-15 not only *is* a supersonic jet but also creates one of another kind. The gaseous exhaust, which according to Newton's third law propels the aircraft forward, is expelled from the engine supersonically, that is, with a speed well in excess of the speed of sound in the gas. The gas forms a well-collimated jet that remains stable over a distance many times as long as its diameter (Fig. 1). This second type of supersonic jet occurs naturally in many contexts and over many distance scales. Moreover, it exhibits remarkable features of structure and stability that challenge the modeling capabilities of modern numerical hydrodynamicists.

Our interest in supersonic gas jets has been twofold. First, we wanted to exploit the inherent flexibility of numerical modeling to explore supersonic flow over a wider range of physical parameters than has been investigated experimentally. Second, we wondered whether through these simulations we could explain the extraordinary stability of astrophysical jets, which, as shown in the opening illustration, can bore their way through intergalactic space from the center of a radio galaxy to distances of thousands or millions of light years. Could we capture in numerical simulations the coherent nonlinear structures that seem to appear in both astrophysical and terrestrial supersonic jets?



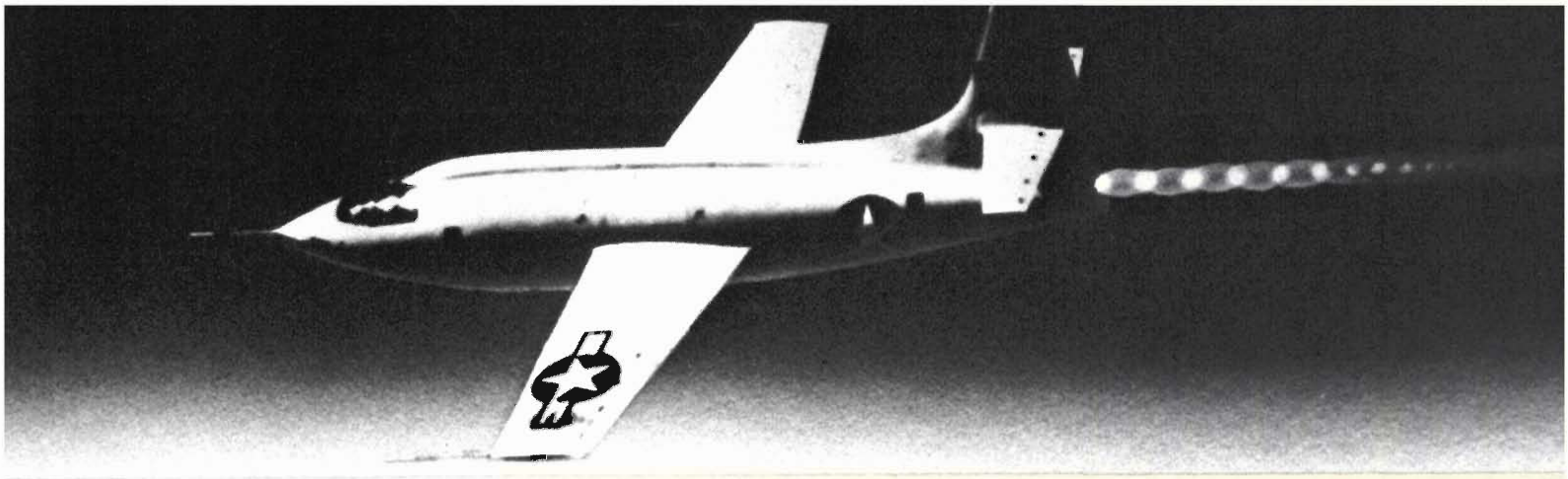


Fig. 1. A supersonic exhaust jet trailing the Bell X-1 rocket plane, the first manned aircraft to break the sound barrier. The beaded appearance of the jet is due to a network of internal shock waves that repeatedly compresses and reheats the exhaust gas. (From "Orders of Magnitude" by F. W. Anderson, NASA Special Publication 4403, 1981.)

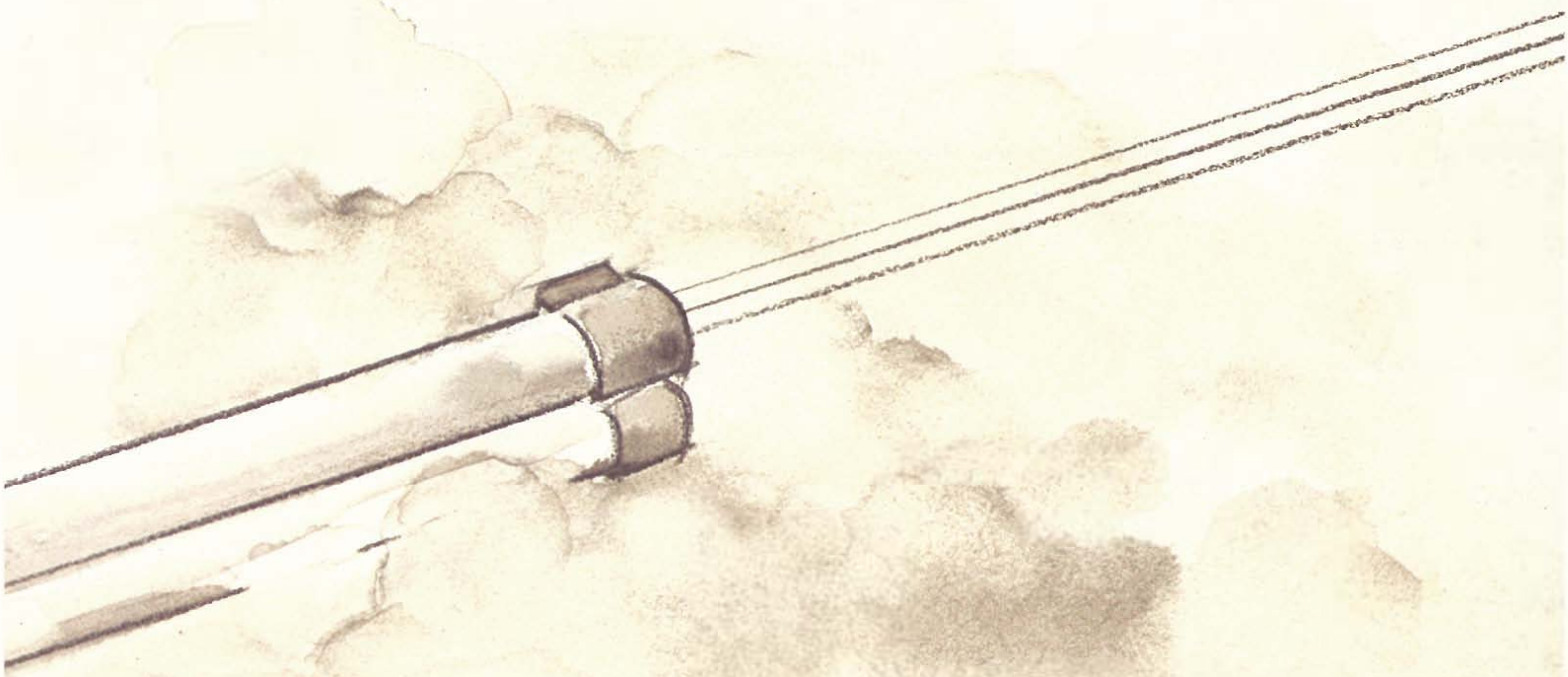
The results we present here were achieved in collaboration with Larry L. Smarr of the University of Illinois and are based solely on numerical simulations carried out on a Cray-1 computer at the Max Planck Institut für Physik und Astrophysik in Garching bei München. Working in West Germany, we were constantly reminded of the dual definition of a supersonic jet as NATO forces maneuvered overhead. This was strangely appropriate, however, for the history of research on supersonic gas jets goes back nearly a century to the pioneering work of Ernst Mach at the behest of the Austrian military.

Ernst Mach and the Military Connection

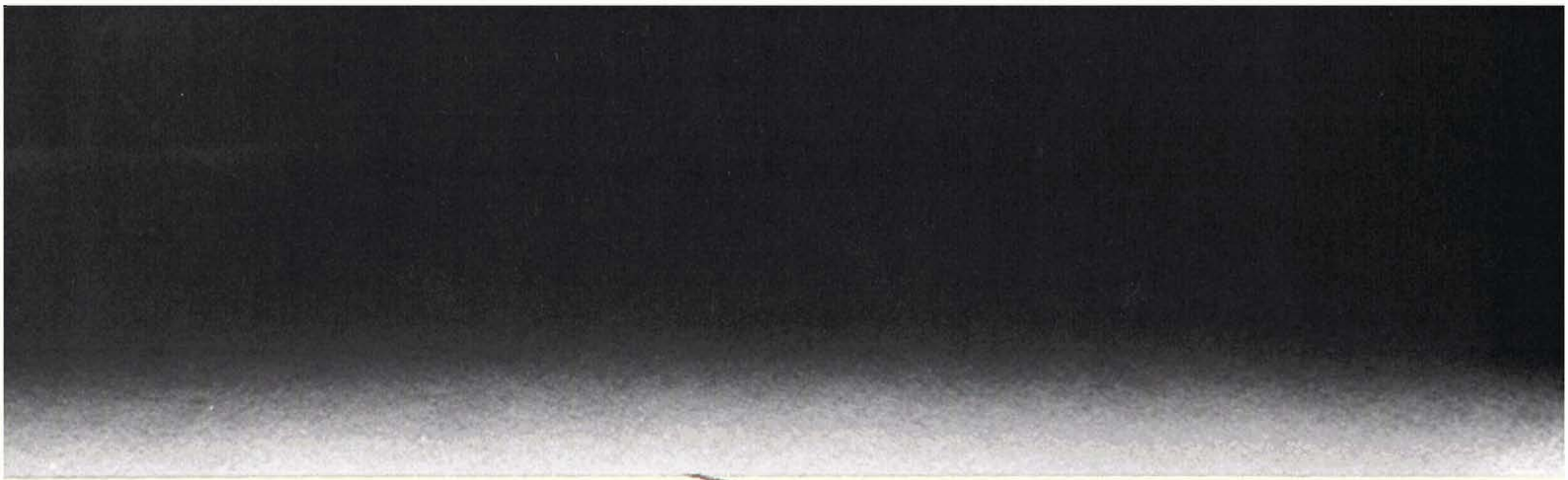
Back in the 1880s the concept of a shock wave was still brand new. In Europe the theoretical properties of these nonlinear compression waves were being hotly debated by the likes of Riemann, Rankine, Rayleigh, and Hugoniot. It was still uncertain whether it was energy or entropy that is conserved across the surface of discontinuity created by a shock front (see Sidebar 1, "Shock Waves versus Sound Waves").

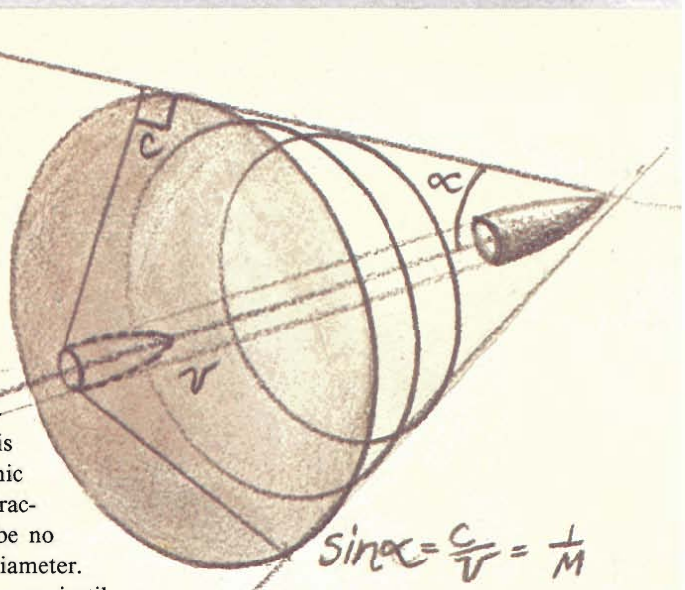
Meanwhile, artillery experts and ballisticians were groping for explanations of two curious phenomena brought to light during the Franco-Prussian War. The first was the double report heard when a shell was fired at high speed from an artillery piece; only a single report was heard when a projectile was fired at low speed. The second was the crater-like nature of the wounds inflicted by the French army's new high-speed bullets. Exploding bullets—expressly forbidden by treaty—were suspected as having been used.

The Belgian ballistician Melsens proposed an explanation for the second phenomenon



Opening figure (previous page): Looking between the dark pressure boundaries of a calculated supersonic jet, we see a radio-brightness map for a long extragalactic jet extending south some quarter of a million light-years from radio galaxy 3C200 and terminating in a radio lobe. In this map, which was taken at the Very Large Array (VLA) near Socorro, New Mexico, by Jack O. Burns, regions of high emission are depicted in red, intermediate emission in yellow, and low emission in blue. The jet powering the radio lobe north of the radio galaxy is for some mysterious reason not seen. (Collage by Jim Cruz.)





that so intrigued Ernst Mach that he set about to confirm it experimentally. Melsens conjectured that a high-speed projectile pushes ahead of itself a considerable mass of compressed air, which may cause explosion-like effects on the body it strikes. In other words, Melsens was proposing the existence of the bow shock that precedes a supersonic projectile, and, of course, Mach showed that Melsens was right.

In 1886 Mach and his experimental colleague Peter Salcher were the first to obtain schlieren photographs of a bow shock. These and other gas shocks are easy to record photographically because the abrupt change in density across a shock causes refraction anomalies in light transmitted through the shock. Mach correctly interpreted the bow shock as the envelope of sonic disturbances originating at the projectile. (The formation of shocks through convergence of sonic disturbances is discussed in Sidebar 2, "Shocks, Rarefactions, and Contact Discontinuities.") He derived the famous equation $\sin \alpha = c/v$ relating the Mach angle α (the opening half angle of the bow shock) to the projectile speed v and the ambient sound speed c (see drawing on this page).

Such a bow shock explains the first phenomenon, the double report: the first bang is produced by the powder gases escaping from the muzzle of the gun, and the second bang is the sonic boom associated with the bow shock of the supersonic projectile.

Ernst Mach is, of course, also remembered for his principle of relativity (Mach's principle), which states that physical forces exist only because of the motion of matter relative to the rest of the matter in the universe. Mach's belief in the relative nature of physical laws must have played a role in the design of a novel setup for his ballistics experiments: he held the projectile fixed in the laboratory and fired air past *it* supersonically.

The supersonic air stream was produced by allowing highly compressed air to escape from a reservoir through a collimating de Laval nozzle. (Essentially the same apparatus is used today to produce supersonic gas jets in the laboratory.) For practical reasons the nozzle could be no more than a few centimeters in diameter. Mach quickly discovered that a projectile immersed in this modestly sized supersonic air stream created a network of standing shock waves far more complex than the single bow shock he expected (see Fig. A in Sidebar 3, "Steady-State Jets"). To quote the 1889 discovery paper of Mach and Salcher, "The stream exhibited a customary appearance so remarkable, that a detailed study of the jet itself seemed worthwhile." Such was the serendipitous discovery of the complex structure of supersonic gas jets and the beginning of a systematic research program carried out by Ernst Mach's son Ludwig to determine the properties of the jets. Further research was also carried out by Emden and Prandtl around the turn of the century.

Steady-State Supersonic Jets in the Modern Era

Interest in supersonic jets essentially vanished for two generations until World War II, when military considerations again made them the topic of research—this time at the Aberdeen Proving Grounds by a group of Americans who were investigating the phenomenon known as gun flash.

When an artillery piece fires, gas from the burnt powder exits from the cannon muzzle at supersonic speeds (the cause of the first bang) and produces for a brief instant the network of oblique crisscrossed shocks that Mach had discovered. These standing shock waves are set up by nonlinear radial oscillations of the gas jet caused by a pressure imbalance at the nozzle orifice (in this case the end of the muzzle). Of particular concern to the Aberdeen researchers was the occurrence in the jet of a type of shock known as a Mach disk. Mach disks, which are perpendicular to the jet flow and much stronger than the oblique shocks, produce "gun flash" by reheating the powder gas.

Since these flashes could reveal the position of a gun to the enemy, Ladenburg, Van Voorhis, and Winckler were interested in designing a muzzle to eliminate them. They discovered that the crucial parameter is the pressure ratio at the nozzle, that is, the ratio of the pressure of the gas in the jet to the pressure of the ambient, or atmospheric, gas. As shown in Sidebar 3, pressure ratios far from unity produce Mach disks, and pressure ratios near unity produce X-shaped "regular" shock reflections. To eliminate Mach disks, one need only flare the muzzle wall outward so that the emerging gas expands and depressurizes before reaching the atmosphere.

continued on page 45

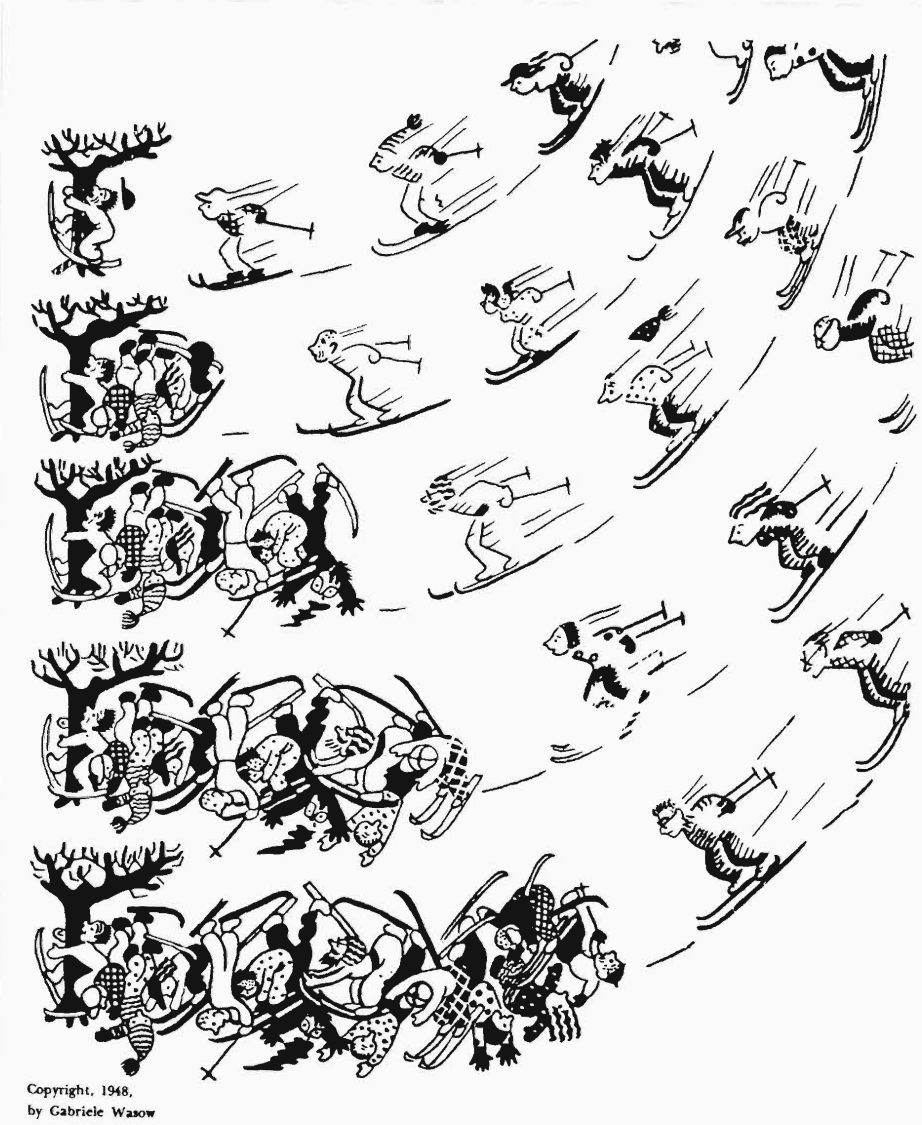
Shock Waves versus Sound Waves

SIDEBAR 1

Explosions, projectiles whizzing by at supersonic speeds, high-speed collisions of solids—what do these phenomena have in common?

They all create very large changes in local pressure over very short times, and these violent pressure changes self-steepen into shock fronts, or shock waves. Unlike a sound wave, which is a small-amplitude compression wave that propagates at the local sound speed and leaves the state of the medium unchanged, a shock front is a nonlinear wave that abruptly changes the state of the supersonically approaching gas. The gas is generally at a higher temperature and pressure after it has passed through the shock (or, equivalently, after the shock has passed through the gas). Moreover, the shock-heated gas moves subsonically with respect to the shock. As we describe below, the narrow region defined as the shock front is a region where thermodynamic processes are irreversible.

The cartoon from Courant and Friedrichs' classic book on supersonic flow illustrates the formation of a steep front in a discontinuous medium, namely, a train of skiers. The skiers, barreling single file down the narrow run, pile up in a heap as first one skier gets wrapped around a tree, and then, before he can warn the skier behind him to slow down, the next skier crashes into him, and so on. The pileup of human wreckage creates a steep front moving up the slope away from the tree analogous to a receding shock front. As in a continuous medium, formation of the front depends critically on the fact that the flow of skiers is "supersonic" in the sense that it is faster than the speed with which the medium, in this case the skiers, can respond to new boundary conditions. The high "pressure" of the skiers behind this front is analogous to the change of state experienced by shock-processed media.



An example of a receding shock wave. From* Supersonic Flow and Shock Waves *by R. Courant and K. O. Friedrichs (New York:Interscience Publishers, Inc., 1948).

About a hundred years ago Stokes, Earnshaw, Riemann, Rankine, Hugoniot, and Lord Rayleigh deduced the conditions that prevail at shock fronts in gases. The framework they used to analyze this compressible flow are the equations of ideal gas dynamics. These nonlinear equations describe conservation of mass, momentum, and energy in an ideal fluid, a fluid in which all changes in kinetic and internal energy are due to pressure forces. Heat conduction and viscous stress are ignored and entropy is constant, so all thermodynamic changes are adiabatic and reversible. It was known that, within this context, infinitesimal pressure changes generate linear compression waves, better known as sound waves, that travel with the local sound speed c ($c^2 = (\partial P/\partial \rho)_S$). But what happens when the amplitude of a compression wave is finite?

Figure A illustrates what happens. At time t_1 we have a finite-amplitude sound wave

such that the pressure variation δP is of order P. Variations in pressure imply variations in sound speed, since for an ideal gas $P \propto c^2$. Thus each point on the waveform propagates with its local sound speed, which is greater at the peaks than in the troughs. With time the waveform steepens (as shown at t_2) and eventually breaks (as shown at t_3) to produce multiple values for the state variables of the gas, P, ρ, and T. Of course this prediction is wrong. Instead nature inserts a shock front (dashed line) just before the wave breaks, and the flow variables remain single-valued.

How does this happen? On a microscopic level large gradients in temperature and velocity at the front of the steepening wave

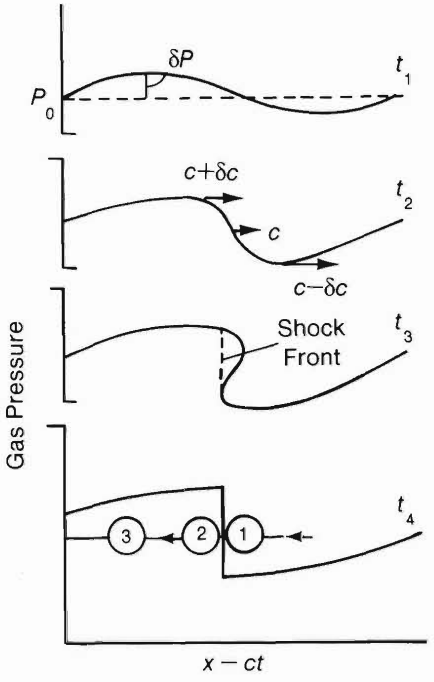


Fig. A. Self-steepening of a finite-amplitude sound wave. In the region where the state variables of the wave (here, pressure) would become multi-valued, irreversible processes dominate to create a steep, single-valued shock front (vertical dashed line).

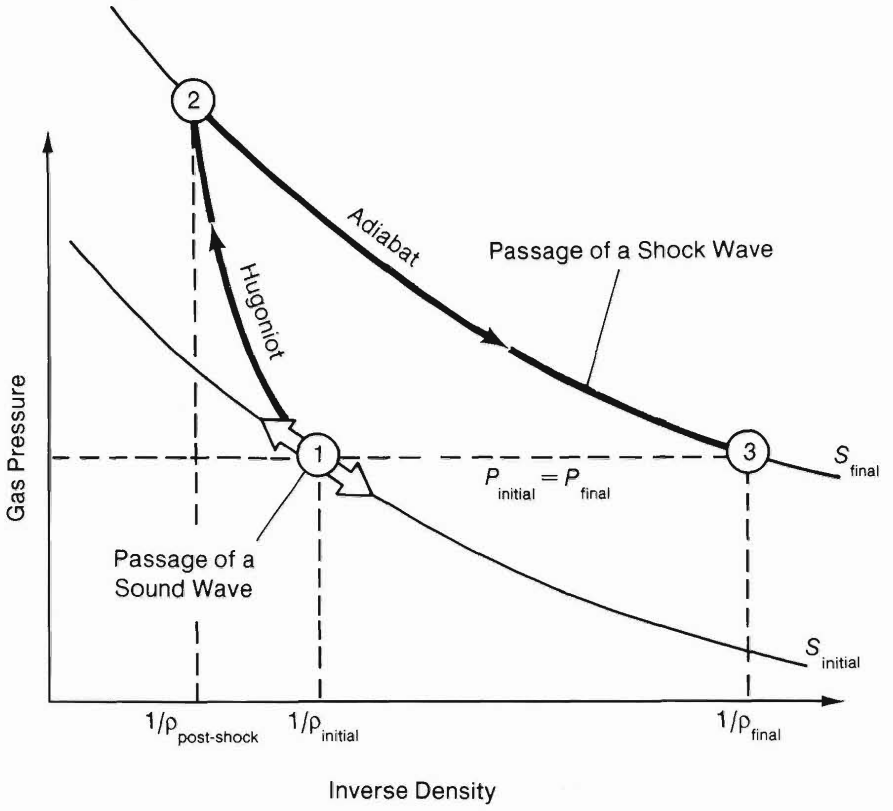


Fig. B. Effects of the passage of a sound wave and of a shock wave. As a sound wave passes through a gas, the pressure and density of the gas oscillates back and forth along an adiabat (a line of constant entropy), which is a reversible path. In contrast, the passage of a shock front causes the state of the gas to jump along an irreversible path from point 1 to point 2, that is, to a higher pressure, density,* and *entropy. The curve connecting these two points is called a Hugoniot, for it was Hugoniot (and simultaneously Rankine) who derived, from the conservation laws, the jump conditions for the state variables across a shock front. After passage of the shock, the gas relaxes back to point 3 along an adiabat, returning to its original pressure but to a higher temperature and entropy and a lower density. The shock has caused an irreversible change in the gas.

cause the irreversible processes of heat conduction and viscous stress to dominate in a region with a width equal to a few collision mean free paths and to counteract the self-steepening process so that a single-valued shock front forms. The net effect on a macroscopic level is that mass, momentum, and energy are conserved across the shock front, but entropy is not; it increases as relative kinetic energy is dissipated into heat through atomic or molecular collisions.

In 1864 Riemann was the first to analyze wave-steepening within the context of ideal gas dynamics. He mistakenly assumed that entropy was conserved (in other words, that all processes were adiabatic) across shock fronts as it is for finite-amplitude sound waves. Later, Rankine, Rayleigh, and Hugoniot showed that an adiabatic shock front would violate conservation of energy, and therefore shock fronts must be non-adiabatic and irreversible. Figure B shows the irreversible changes caused by the passage of a shock front in contrast to the reversible changes produced by a sound wave. To model these irreversible effects, the form of the Euler equations of gas dynamics we have adopted must be modified as described in Fig. 2 of the main text.

The dissipative nature of a shock front implies that it can maintain itself only in the presence of a driving force. A simple example of a driven shock front is given in Sidebar 2. ■

Shock Waves, Rarefactions, and

SIDEBAR 2

Contact Discontinuities

The nonlinear equations of ideal gas dynamics support three types of nonlinear "waves": shock fronts, rarefactions, and contact discontinuities.

Contact discontinuities are surfaces that separate zones of different density and temperature. By definition such a surface is in pressure equilibrium, and no gas flows across it. When, as often happens, the tangential velocity of the gas on one side of the surface differs considerably from that of the gas on the other side, the surface is called a slip discontinuity. The boundary between a supersonic jet and the ambient gas is an example of a slip discontinuity.

The other two types of nonlinear waves arise from abrupt changes in pressure. Shock fronts accompany compression of the medium, and rarefactions accompany expansion of the medium. As a simple example of how these waves arise, consider the one-dimensional flow of gas in a small-diameter cylindrical tube, which is fitted with a piston at one end ($x = 0$) and closed at the other end. Figure A shows the time-dependent flow that results when the piston is suddenly pushed into the cylinder and how a shock front forms as sonic disturbances generated by the piston converge. These sound waves travel along trajectories called characteristic paths, or simply characteristics. In general, the characteristics of a set of hyperbolic partial differential equations are the paths along which certain variables are conserved and are thus the paths along which information travels. The characteristics for the equations of ideal gas dynamics are (1) the streamlines along which matter flows and entropy is conserved and (2) the plus and minus

Fig A. Compression in a Shock Tube

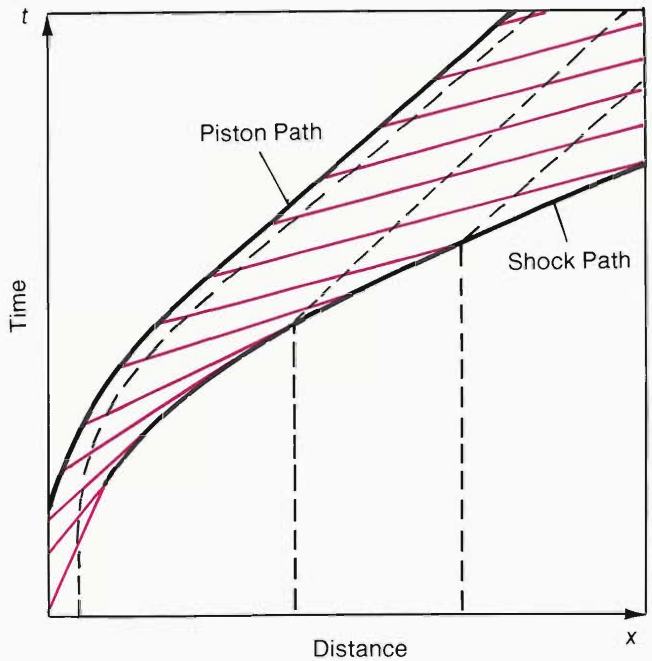
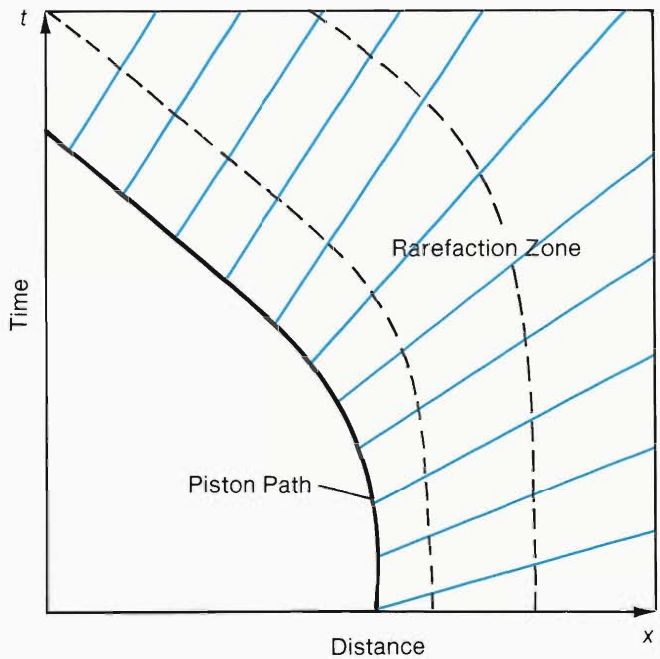
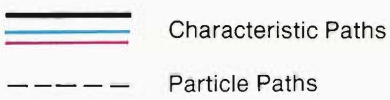


Fig B. Expansion in a Shock Tube





SIDEBAR 2

characteristics along which information travels from one point to another by sound waves. The so-called Riemann invariants, which are linear combinations of the fluid velocity and sound speed, are conserved along the plus and minus characteristics. In general, characteristics form a curvilinear coordinate system that reduces a complicated set of nonlinear partial differential equations to a coupled set of ordinary differential equations.

Since sonic disturbances travel in the gas at the local sound speed c, and since the gas immediately ahead of the piston is moving with the piston speed $v_{\text{piston}}(x)$, the sound waves emanating from each point along the piston path travel at a speed of $v_{\text{piston}}(x) + c$. Hence at each point x the characteristic makes the same angle with the piston trajectory. As a result, the characteristics (red) emanating from the curved portion of the piston path (the acceleration phase) converge to form a shock front. (The bow shock of a supersonic projectile also forms where sonic disturbances generated along the trajectory converge, except that in this case the sonic disturbances move through a stationary rather than a moving medium and therefore all travel at speed c.) The shock front ahead of the piston travels along the tube at supersonic speeds. Gas particles farther away from the tube remain stationary until the shock front passes, and then they travel at a speed equal to the final constant speed of the piston. The closer spacing of the particle paths (dashed lines) gives graphic evidence of shock compression.

Figure B shows the time-dependent flow that results when the piston is suddenly withdrawn from some position within the cylinder. As in Fig. A, characteristics drawn at a constant angle to the piston path represent the path of sonic disturbances generated in the gas immediately adjacent to the piston. In this case the characteristics (blue) emanating from the curved portion of the piston path diverge to form a rarefaction zone (gray). The greater separation of the particle paths in the rarefaction zone indicates that the gas is expanding. ■

continued from page 41

Other data on supersonic flows are available from studies of jet and rocket propulsion. Since these studies were aimed at designing more efficient, higher thrust engines, virtually all of them have focused on the steady-state, or time-averaged, properties of the jet flow, as has our discussion so far.

Time-Dependent Jets in Nature

The time-dependent evolution of jets becomes important when we consider supersonic "jets" in nature. One example is the supersonic jet produced by the eruption of a gas-rich volcano, such as the eruptions of Krakatoa and Mount St. Helens. The blast wave that accompanies the formation of such a volcanic jet produces widespread devastation. Experiments have shown that these transient supersonic jets have structural properties resembling those of steady-state jets.

Other examples are the numerous jets emanating from astrophysical objects. In our galaxy protostars, old binary star systems associated with supernova remnants, and planetary nebulae produce jets of ionized gases. A number of these stellar jets have knots of bright emission that remind one of the bright spots in the exhaust jet shown in Fig. 1. Spectral lines from highly excited atoms in the knots indicate that these bright spots are regions of high temperature, probably produced by shock heating. The Doppler shifts of the spectral lines demonstrate that these jets are still evolving with time.

Our final example of time-dependent jets, and the ones that exhibit the most remarkable and mysterious stability, are the enormous jets emanating from quasars and other galaxies. These jets, known as radio jets because most of them can be observed only with radio telescopes, range from thousands to millions of light years in length, or about fifty to one hundred jet diameters. Many of them exhibit bright knots of radio emission in the beam. A radio jet typically ends in a "hot spot"—a small region of intense radio emission—embedded in a larger amorphous lobe of fainter emission (the opening illustration is an exception).

The emission spectra of radio jets, which are continuous and polarized, are produced by synchrotron emission from a nonthermal population of relativistic electrons gyrating in a magnetic field. This nonthermal component probably coexists with a diffuse thermal plasma in some uncertain mixture to form the "stuff" of extragalactic radio sources. We know little else about the composition of the jets and the lobes and, in the absence of spectral lines, have no direct measure of jet speeds from Doppler shifts.

In a landmark paper, written in 1974 when only the radio lobes, and not the jets themselves, had been observed, Roger Blandford and Martin Rees interpreted the hot spots in the lobes as the point of impact of a supersonic gas jet impinging on the ambient medium as the jet advances. In other words, the radio lobes are powered by fluid-like beams, or jets, originating at the centers of radio galaxies. They suggested that the main structural features of the jets could be explained by ideal gas dynamics, that is, by the nonlinear differential equations of compressible fluid flow. Magnetic fields and complex plasma dynamics that might affect local regions could be ignored when considering the overall structure of a jet.

The fluid-beam hypothesis of Blandford and Rees provides a simple physical explanation for radio observations and a framework for detailed model building, but within this framework the issue of stability becomes paramount. We know from laboratory experiments that Kelvin-Helmholtz shear instabilities grow rapidly on the boundary of supersonic gas jets and effectively decelerate and disrupt the flow. How do extragalactic jets propagate such great distances without suffering the same fate? Are they, or are they not, governed primarily by the equations of gas dynamics? If so, what are the critical parameters determining the stability and the observed morphology of the flow? We undertook to answer these questions by

continued on page 50

SIDEBAR 3

Steady-State Jets

Supersonic gas jets are created in the laboratory by allowing a highly compressed gas to escape through a nozzle into the atmosphere or some other ambient gas. In contrast to astrophysical jets, laboratory jets usually have high density ratios ($\eta \equiv \rho_{jet}/\rho_{ambient} > 1$), moderate Mach numbers ($M \equiv v_{jet}/c_{jet} \leq 3$), and large pressure mismatches at the source. Investigations of laboratory jets have, until recently, focused on their steady-state behavior rather than the initial time-dependent behavior of interest in astrophysical contexts. Nevertheless, our understanding of the steady-state structures observed in laboratory jets provides a point of departure for interpreting our numerical simulations of time-dependent jet flow.

An Idealized Supersonic Jet

Figure A shows the characteristic structure taken on by a slightly underexpanded supersonic jet, that is, one for which the pressure of the gas at the nozzle orifice, P_n, is slightly greater than the ambient gas pressure, P_a. (A jet is referred to as underexpanded if $K \equiv P_n/P_a > 1$, overexpanded if $K < 1$, and pressure matched if $K = 1$.)

This complicated axisymmetric structure has several remarkable features. First, the jet boundary oscillates as the jet gas periodically overexpands and reconverges in its attempt to match the ambient pressure. The gas continually overshoots the equilibrium position because the effects of the boundary are communicated to the interior of the jet by sound waves, which, by definition, travel more slowly than the bulk supersonic flow. The characteristic paths of the sound waves converge to form the second remarkable feature of the jet, the network of crisscrossed shock waves, or shock diamonds (red lines). These standing shocks alternate with rarefaction fans (blue lines). The gas in the jet interior expands and cools (shades of blue) as it flows through the rarefaction fans and is compressed and heats (shades of red) as it passes through the shock diamonds. The figure clearly illustrates that the jet interior is always out of step with the jet boundary. For example, the positions of greatest gas compression (dark red) do not coincide with the

SIDEBAR 3

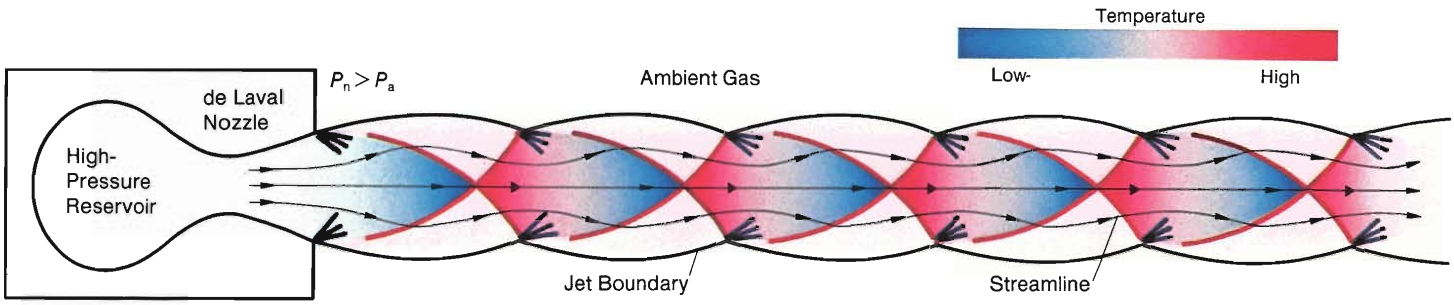


Fig. A. Idealized steady-state structure of a slightly underexpanded (P_n slightly greater than P_a) supersonic jet. The red lines represent incident and reflected shocks (see Fig. B), and the blue lines indicate the beginnings of rarefaction fans. The gas temperature varies according to the key. Black streamlines follow the oscillating flow path of the jet gas.

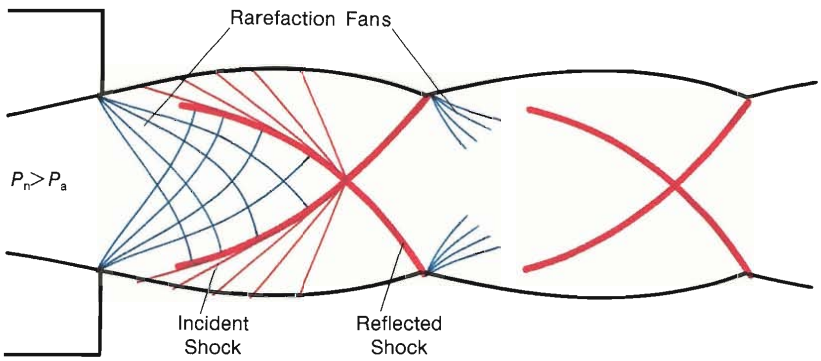
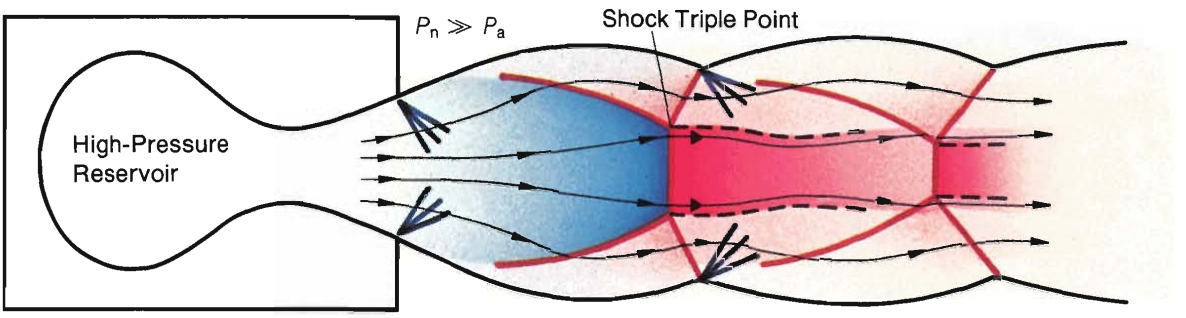


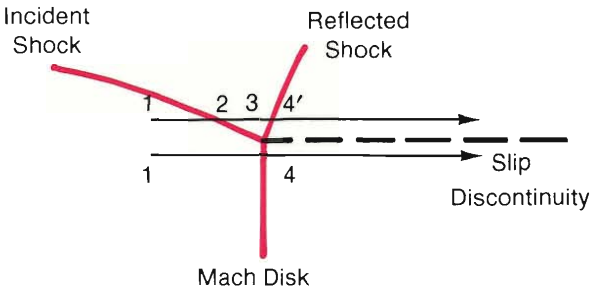
Fig. B. Regular reflection in a slightly underexpanded jet. The pattern of crisscrossed shocks in Fig. A can be understood in terms of characteristics. Diverging characteristics (blue) form rarefaction zones. Converging characteristics (red) from the boundary form the incident conical shocks, which reflect off the jet axis to form the reflected shocks.

positions of minimum jet diameter. The black streamlines in the figure indicate the flow paths of the gas. The gas bends out toward the boundary as it passes through rarefaction fans and bends back toward the axis as it passes through shock fronts.

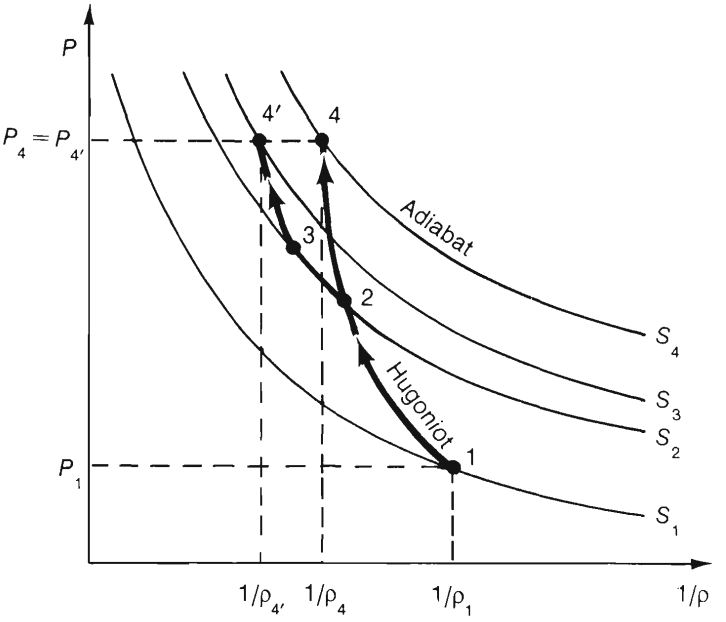
Regular Reflections

Figure B shows how the shock structure in Fig. A can be understood in terms of characteristics. As the gas leaves the nozzle, it expands and a rarefaction fan (diverging blue characteristics) emanates from the nozzle orifice. The gas overexpands, and the pressure of the ambient gas at the boundary, acting like the piston in Sidebar 1, pushes the jet gas back toward the axis, creating the red characteristics. These characteristics form a converging conical shock. When this so-called incident shock reaches the jet axis it undergoes a regular reflection; that is, it forms a diverging shock. At the point where this reflected shock reaches the jet boundary, it knocks the boundary outward, creating a new rarefaction fan, and the process begins all over again. The geometric pattern of the (x, y) characteristics that form the shocks and rarefaction fans in this steady-state two-dimensional flow is similar to that of the





(1) Shock Triple Point



(2) Thermodynamic History

Fig. C. Mach reflection in an underexpanded ($P_n \gg P_a$) supersonic jet. A Mach reflection creates a shock triple point where three shocks meet. Figures C1 and C2 show how passage through the Mach disk (from point 1 to point 4) slows down the beam much more than passage through the incident and reflected shocks (from point 1 through points 2 and 3 to point 4′). As a result, a slip discontinuity is formed (dashed line).

(x, t) characteristics for one-dimensional time-dependent flow described in Sidebar 2.

Mach Reflections

A striking change in flow structure occurs when the pressure mismatch at the orifice is large ($K \gg 1$ or $K \ll 1$). As shown in Fig. C for an underexpanded jet, the incident shock, rather than converging to a point on the axis, reflects at the perimeter of a Mach disk—a strong shock normal to the flow direction (so named because Ernst Mach was the first to record its existence). The angle between the incident shock and the jet axis determines the type of reflection: small angles of incidence yield the regular reflections shown in Figs. A and B, and large angles of incidence yield Mach reflections. When gas passes through a shock, its velocity component normal to the shock is greatly reduced but its parallel component remains unchanged. Thus shocks with large angles of incidence relative to the flow axis are much more effective at slowing down the flow than shocks with small angles of incidence. The critical angle for transition from regular reflections to Mach reflections is approximately the angle that yields a sonic relative velocity for the gas downstream of the shock.

SIDEBAR 3

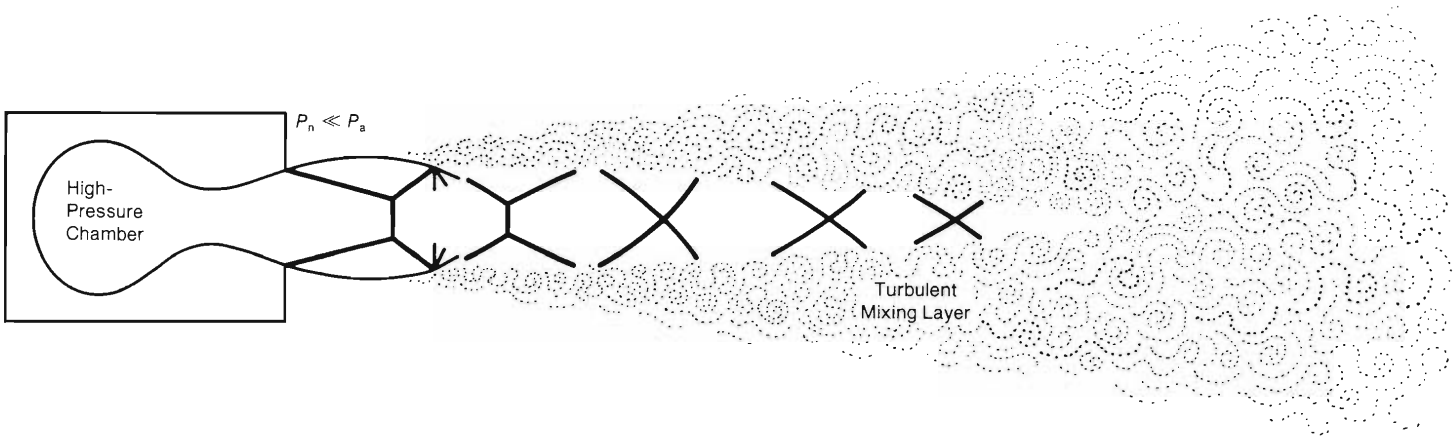


Fig. D. Realistic steady-state structure of an overexpanded ($P_n \ll P_a$) supersonic jet, showing the presence of both Mach reflections and regular reflections. Since this is an overexpanded jet, shocks rather than rarefactions emanate from the nozzle orifice. Shear instabilities create a mixing layer that grows until it reaches the beam axis.

A prominent feature of Mach reflections is the emergence of a slip discontinuity (dashed line) from the shock triple point where the incident shock, reflected shock, and Mach disk meet. The flow velocity, density, and temperature are discontinuous across this contact surface. This slip discontinuity arises because the thermodynamic pathway through the incident and reflected shocks *does not equal* the pathway through the Mach disk.

In Fig. C1 we display two adjacent streamlines, one on each side of the shock triple point, and in Fig. C2 we display the corresponding thermodynamic pathways.

Adjacent fluid elements at some initial point on either streamline have the same state variables and the same total (kinetic plus internal) specific energy, that is, energy per unit mass. This quantity is given by

$$\frac{1}{2}v^2 + \frac{\gamma}{\gamma - 1}\frac{P}{\rho},$$

where γ is the ratio of the specific heat of the fluid at constant pressure to the specific heat at constant volume.

By Bernoulli's principle the total specific energy remains constant along a streamline. Therefore, when the two adjacent fluid elements arrive at points 4 and 4′ in Fig. C1 they must still have the same total specific energy. They must also have the same pressure because they are still adjacent. However, we see from Fig. C2 that their densities and entropies are different. The element at point 4 has been shock-heated along a Hugoniot to a higher entropy and lower density than the element that passed through the incident and reflected shocks along points 2 and 3 to point 4′. The lower density of the fluid element at point 4 implies that its internal energy (which is proportional to P/ρ) is greater than that of the fluid element at point 4′. Bernoulli's principle then implies that the fluid element at point 4 must have a correspondingly lower kinetic energy, and hence flow velocity, than the adjacent element at point 4′. A slip discontinuity results from this difference in flow velocities.

Laboratory Supersonic Jets

The jet structures shown in Figs. A, B, and C are idealizations in that real supersonic jets do not have sharp, stable boundaries but turbulent boundaries where jet and ambient gases mix. Figure D shows a more realistic steady-state structure for an overexpanded laboratory jet. Near the orifice, where the pressure mismatch is large, Mach reflections occur, but farther downstream the reflections are regular. The mixing layer, which grows as a result of Kelvin-Helmholtz (shear) instabilities, progressively eats its way into the supersonic core of the jet. When the mixing layer reaches the axis of the jet, the flow is subsonic and fully turbulent. It is then susceptible to twisting and bending motions, like a smokestack plume in a crosswind.

Strictly speaking the wave structures within the supersonic core are not steady since they are buffeted by the turbulent boundary layer. However, their average positions, those in Fig. D, are well defined. ■

continued from page 45
solving the equations of gas dynamics through numerical simulation.

Numerical Jet Laboratory

The numerical problem we set ourselves is well defined. We start with the equations for compressible fluid flow displayed in Fig. 2a, assume that at $t = 0$ the jet gas starts to flow continuously through a narrow orifice into a region filled with quiescent gas as shown in Fig. 2b, and compute the time-dependent flow numerically with finite-difference techniques. We assume that the flow is axisymmetric and nonswirling (that is, the gas has zero angular momentum) and thus reduce a three-dimensional problem to two dimensions (R and Z). This reduction is necessary to achieve reliable numerical results with the computing power of a Cray-1. Moreover, since many extragalactic jets are remarkably straight over most or all of their lengths, axisymmetry is a reasonable assumption. The importance of nonaxisymmetric effects is discussed toward the end of this article.

The interesting thing about this problem is that it is scale-invariant, and therefore the results of the calculation can be scaled from laboratory to cosmic dimensions by varying the only dimension that appears in the problem, the jet diameter. Four scale-invariant (dimensionless) parameters determine the flow: η, K, M, and θ. The parameter η is the ratio of the beam, or jet, gas density ρ_b to the ambient gas density ρ_a,

$$\eta \equiv \rho_b / \rho_a ;$$

the parameter K is the ratio of the jet gas pressure P_b to the ambient gas pressure P_a,

$$K \equiv P_b / P_a ;$$

the parameter M, the Mach number, is the ratio of the jet speed v_b to the speed of sound in the jet gas c_b,

$$M \equiv v_b / c_b ;$$

(a) Euler Equations for Compressible Flow

Conservation of Mass
$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0$$

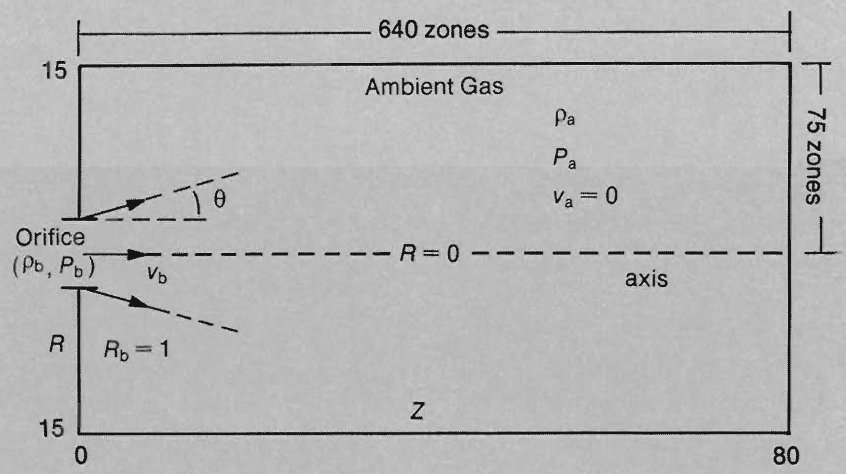
Conservation of Momentum
$$\frac{\partial (\rho \mathbf{v})}{\partial t} + \nabla \cdot (\rho \mathbf{v} \cdot \mathbf{v}) + \nabla P = \vec{\mathcal{D}}$$

Internal Energy Equation
$$\frac{\partial (\rho e)}{\partial t} + \nabla \cdot (\rho e \mathbf{v}) + P(\nabla \cdot \mathbf{v}) = \mathcal{Q}$$

Equation of State for Ideal Gas
$$P = \rho c^2 / \gamma = (\gamma - 1)\, \rho e$$

ρ = density
$\mathbf{v}$ = velocity
$\mathbf{P}$ = pressure
e = internal specific energy
c = sound speed
γ = ratio of specific heats
$\vec{\mathcal{D}}$ = artificial viscous stress
$\mathcal{Q}$ = artificial viscous heating

(b) Computational Region



Inflow boundary conditions in terms of dimensionless parameters:
$\eta = \rho_b / \rho_a$, $M = v_b / c_a$, $K = P_b / P_a = 1$, $\theta = 0$.

Fig. 2. (a) The Euler equations of ideal compressible flow modified by source terms $\vec{\mathcal{D}}$ and $\mathcal{Q}$ to simulate viscous stress and heating in a narrow region behind the shock front. These terms have zero value elsewhere so that farther away from the shock front all processes are reversible and adiabatic, in agreement with experimental observations. (b) The computational grid and boundary conditions for numerical simulation of axisymmetric pressure-matched jets. Gas enters the computational region through a small orifice on the axis. Ambient gas is subsequently allowed to escape through the outer boundaries if required by advance of the jet.

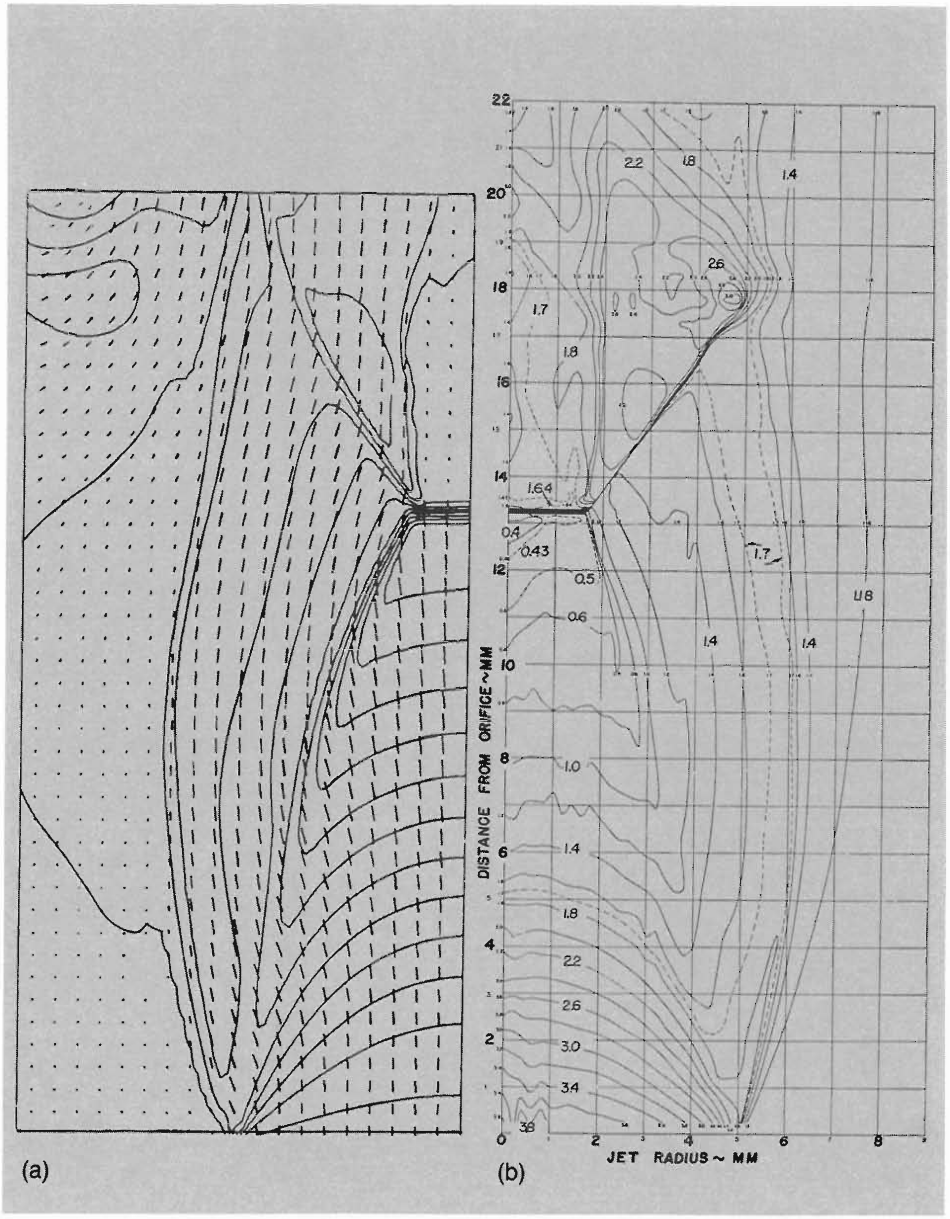


Fig. 3. A comparison of computed and measured steady-state flow fields in an underexpanded axisymmetric supersonic jet with parameters $\eta = 3.36$, $M = 1$, $K = 2.75$, ***and*** $\theta = 0$. ***The numerical solution given in (a) shows density contours and velocity vectors at every third grid point in a 60-by-180 calculation grid. The experimental results in (b), which were adapted from R. Ladenburg*** et al., **Physical Review** *76(1949):662, show density contours and the triple-shock configuration. Note the good agreement between calculation and experiment on the position and geometry of the embedded shock waves.***

and the parameter θ is the opening angle of the jet,

$$\theta \equiv \tan^{-1}(v_R / v_Z) ,$$

where v_R and v_Z are the components of the beam velocity in the R and Z directions.

For extragalactic jets it is reasonable to reduce the number of parameters to two. To mimic the small observed opening angles of these jets, we start with a perfectly collimated beam at the source ($\theta = 0$). We also assume that the jet gas pressure equals the ambient gas pressure ($K = 1$) because x-ray data from satellites tell us that hot thermal gas pervades intergalactic space and is adequate in most cases to pressure-confine the jets. The remaining two-parameter set (M,η) defines a solution space that has scarcely been explored in the laboratory because of the difficulties of achieving very high Mach numbers ($M > 10$) and very low density ratios ($\eta \ll 1$). The ability to explore solutions corresponding to arbitrary values of these parameters is one of the great advantages of numerical simulation.

Solution of the two-fluid problem posed by jet flow depends critically on the motion of the interface separating the two fluids. Fluid interfaces are generally subject to Kelvin-Helmholtz instabilities, which are driven by shear motion parallel to the interface, and by Rayleigh-Taylor instabilities, which are driven by accelerations normal to the interface. Instabilities lead to mixing of the two fluids and can modify the jet flow significantly.

Techniques for tracking interfaces are a specialty at Los Alamos and Livermore, where complicated multi-fluid problems have been handled on large computers for three decades. Although multi-fluid problems are abundant in astrophysics as well, they have usually been approximated as one-fluid problems. Consequently, the use of interface-tracking techniques is somewhat of a novelty in numerical astrophysics, although in this application it is essential.

We adapted an interface-tracking technique developed by James M. LeBlanc and James R. Wilson at Livermore. This calculation is Eulerian; that is, the calculation grid is fixed. The interface between the two fluids is defined by those computational zones with fractional fluid volumes. The fractional volumes are carried along by the bulk flow, and their consequent redistribution on the grid is then used to construct an approximate geometry of the displaced fluid interface. Although an approximate treatment, the technique is robust and very effective at preventing an undesirable consequence of most Eulerian computations, the artificial "blending" of disparate materials at moving boundaries.

We tested the code through comparisons with experiment. Figure 3 shows one such comparison, in which we attempted to reproduce not only the general shape of the flow, but also the position and geometry of the embedded shock waves. The surprisingly good agreement gives us confidence that we can reproduce accurately the steady-flow structures measured in laboratory supersonic jet experiments and, with perhaps some luck, the transient jet structures as well.

A Supersonic Jet Up Close

Having described our numerical simulation apparatus, let us now use it to observe the establishment and propagation of a pressure-matched supersonic jet, that is, one for which $K = 1$.The following analysis relies heavily on computer graphics. Plots of several different flow variables will be required to get a complete picture of the flow and to dig out the essential physics.

Figure 4 shows four snapshots in the time evolution of a low-density jet with initial values of $(M,\eta) = (3,0.1)$. In these plots the color-coded density values show that the entire jet is substantially less dense than the ambient gas throughout its evolution. Pressure confinement is nevertheless maintained because the jet gas is correspondingly hotter than the ambient gas.

The penetration of hot, diffuse supersonic jets into cold, dense continuous media is just beginning to be investigated in the laboratory (Fig. 5), but this physical situation appears to be the relevant one for astrophysics. A novel aspect of such jets, already predicted by Blandford and Rees in 1974, is the establishment of a cocoon—a sheath of gas surrounding the central beam that "splashes back" from the head of the jet. The cocoon is visible in Fig. 4a as a blue layer above and below the green supersonic beam. The mechanics governing deflection of the beam gas into the cocoon at the so-called working surface is quite complicated and will be discussed below. Nevertheless, the existence of cocoons in low-density jets is easy to understand on simple physical grounds.

As shown in Fig. 6, the jet head advances at a speed determined by a balance of ram pressure (or, equivalently, momentum flux) along the beam axis in the rest frame of the working surface. If we let W be the speed of advance of the working surface, then ram-pressure balance is given by $\rho_b (v_b - W)^2 = \rho_a W^2$. Solving for W and using the definition $\eta \equiv \rho_b / \rho_a$, we find that

$$W = \frac{v_b \sqrt{\eta}}{1+\sqrt{\eta}}. \qquad (1)$$

Equation 1 implies that the advance speed W

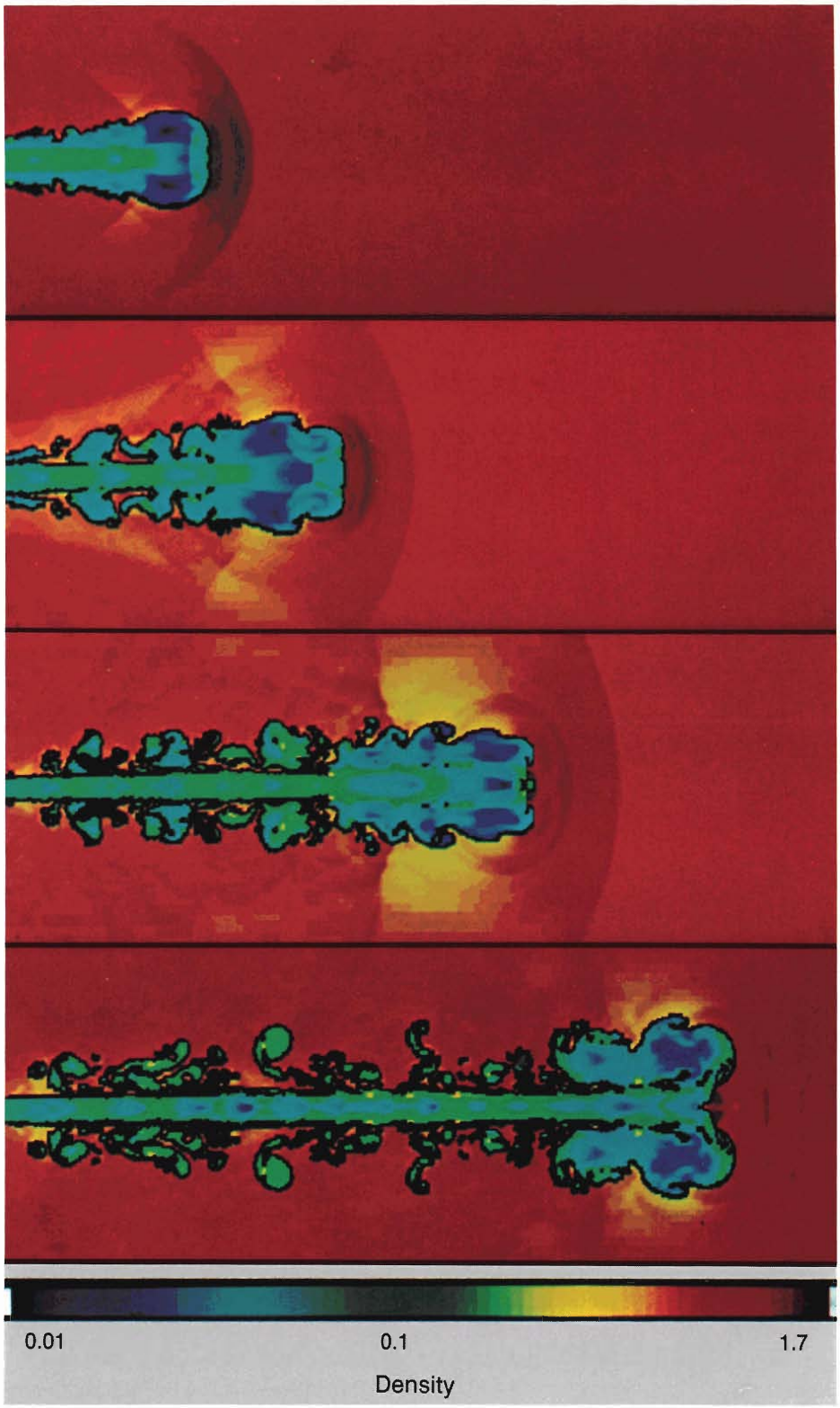


Fig. 4. Numerical simulation of the time evolution of a pressure-matched ($K = 1$) axisymmetric supersonic jet. The jet gas is admitted continuously from the left with an opening angle θ of zero. For this jet $M = 3$ and $\eta = 0.1$. The calculated gas density varies according to the color code. Frames (a) and (b) show the establishment and turnover of the jet cocoon, which is composed of gas that has passed through the terminus of the jet. Frames (c) and (d) show the subsequent mixing of cocoon and ambient gases, which leaves intact only a lobe of cocoon gas at the jet head. The bow shock driven by the supersonically advancing jet head is revealed by the jump in density it produces. The jet image has been reflected about the axis of symmetry for ease of visualization.

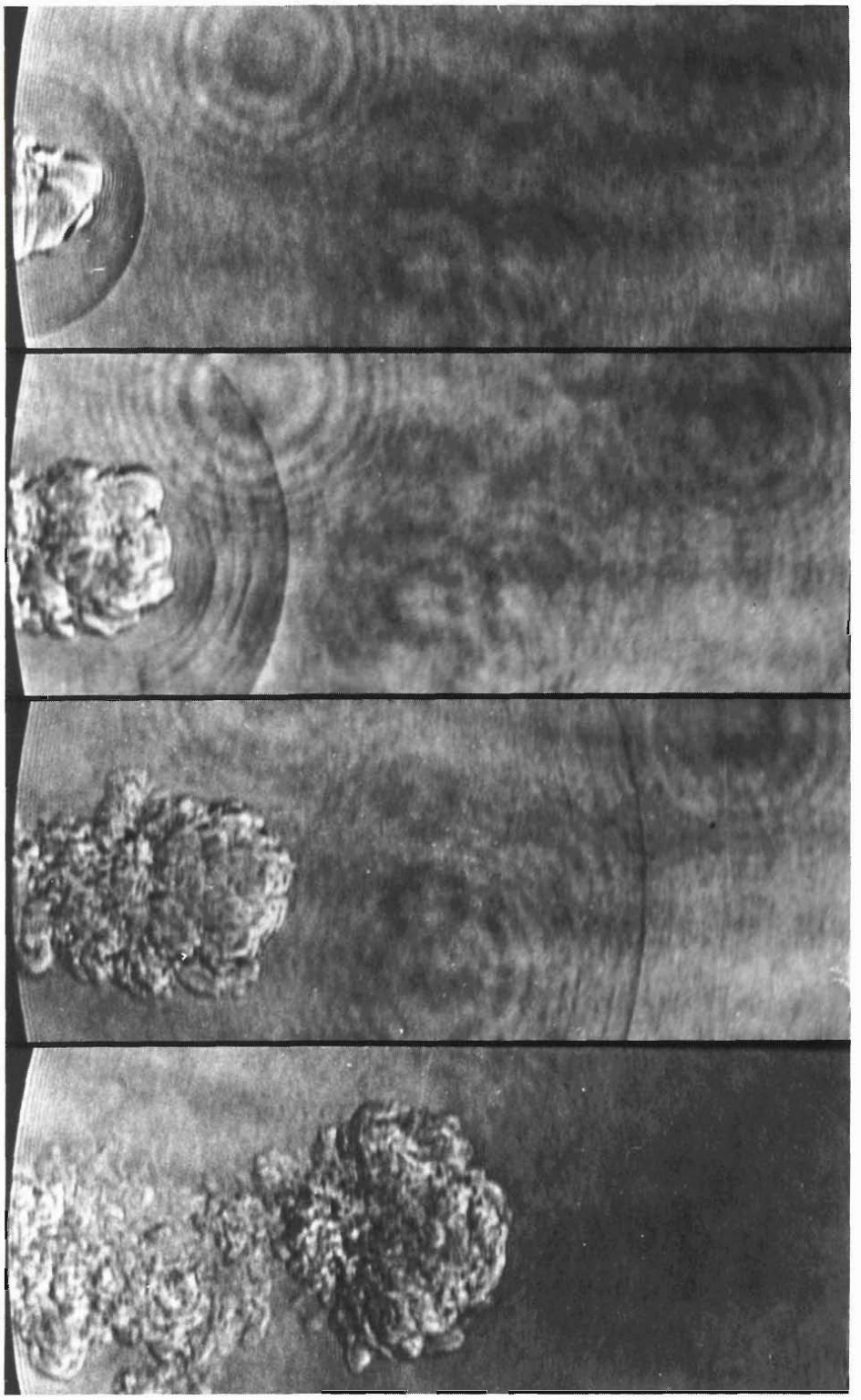


Fig. 5. Laser-schlieren photographs of the time evolution of a low-density supersonic jet. The jet was created by release of 3 joules of energy in a spark, which caused ejection of a plasma pulse into ambient air. Shown here, top to bottom, are short-exposure photographs taken 15, 35, and 80 microseconds after the pulse and at a very late stage in the evolution of the jet. The short exposure time reveals the transient behavior of the flow and emphasizes the turbulent structure of the jet cocoon. The bow shock preceding the jet can be seen as an expanding arc in the first three frames; it has moved out of view in the last frame. The length of the area viewed is 4.5 centimeters. (Reproduced with the permission of J. A. Cavolowsky and A. K. Oppenheim, Department of Mechanical Engineering, University of California, Berkeley.)

of the head of a diffuse jet ($\eta \ll 1$) is less than the beam speed v_b, which means that the beam gas constantly overtakes the jet head. Rather than accumulate there, the incoming beam gas continuously escapes into the cocoon after being shock-decelerated and deflected sideways at the working surface. Surprisingly, the cocoon gas actually flows backward toward the inlet. Backflow is a simple consequence of mass conservation and is encountered on earth when supersonic water jets are used to drill into rocks.

The relative velocity between the backward flowing cocoon gas and the ambient gas at the cocoon boundary is substantial, but it is subsonic for the jet shown in Fig. 4. Consequently, the cocoon boundary is subject to the Kelvin-Helmholtz shear instability, which overturns and mixes the cocoon and ambient gases, as can be seen in Figs. 4b and 4c. As a result of the instability, only a lobe of cocoon gas is left intact at the jet head in Fig. 4d. The central supersonic beam nevertheless appears to propagate without significant mixing at its boundary. We can verify that this is indeed the case with the aid of additional plots that elucidate the flow field in the jet (Fig. 7).

Details of the Flow Field

In Fig. 7a a novel mode of display reveals the large-scale structures of the flow field for the ($M = 3, \eta = 0.1$) jet of Fig. 4. Kinematically distinct parts of the flow inside and outside the jet are assigned different colors. Within the jet boundary (black) we can identify the central supersonic beam (green), the backward flowing ($v_Z < 0$) cocoon gas (white), and the forward flowing ($v_Z > 0$) cocoon gas (yellow). Outside the jet we can identify the undisturbed ambient gas (dark blue) and the bow shock (red). The beam consists by definition of those regions where the local kinetic energy flux $\frac{1}{2} \rho v_Z^3$ is at least 5 percent of its maximum value. Aside from a few yellow patches within the beam, which we shall comment on below, we can follow the supersonic channel uninterrupted to the very end of the jet.

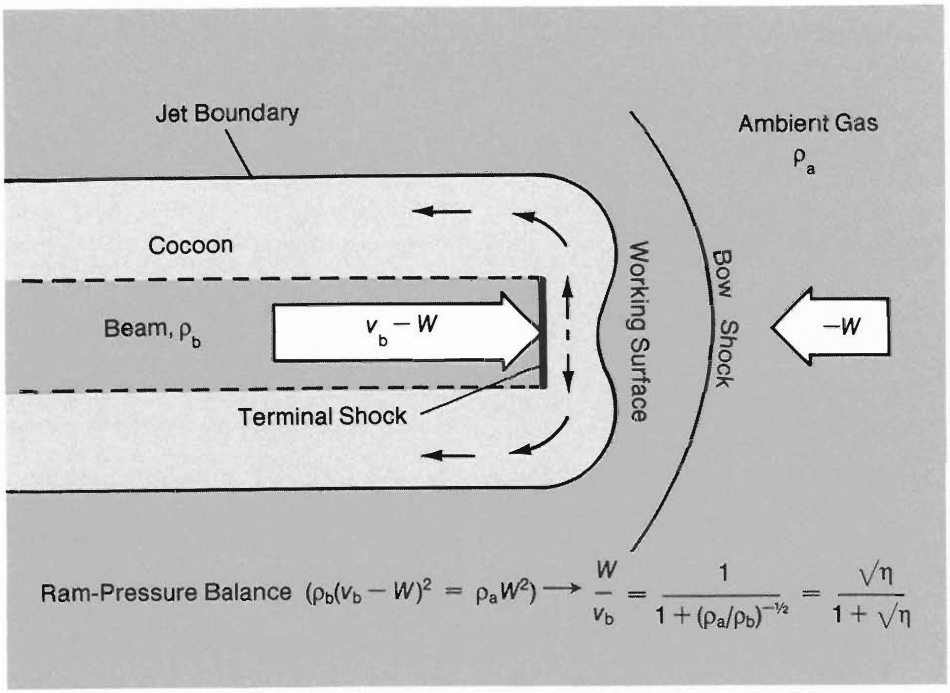


Fig. 6. Flow schematic of the jet head in the rest frame of the working surface. The speed of advance W of the working surface is determined by equating the ram pressure, or momentum flux, of the beam to the ram pressure of the ambient gas in a frame in which the working surface is stationary. In this frame the beam speed is $v_b - W$ and the speed of the ambient gas is $-W$. The expression derived for W/v_b implies that, for jets with low density ratios ($\eta \ll 1$), the working surface advances much more slowly than the beam gas. However, the beam gas does not collect at the working surface but is shock decelerated and deflected sideways to form a low-density backward flowing cocoon around the forward moving beam.

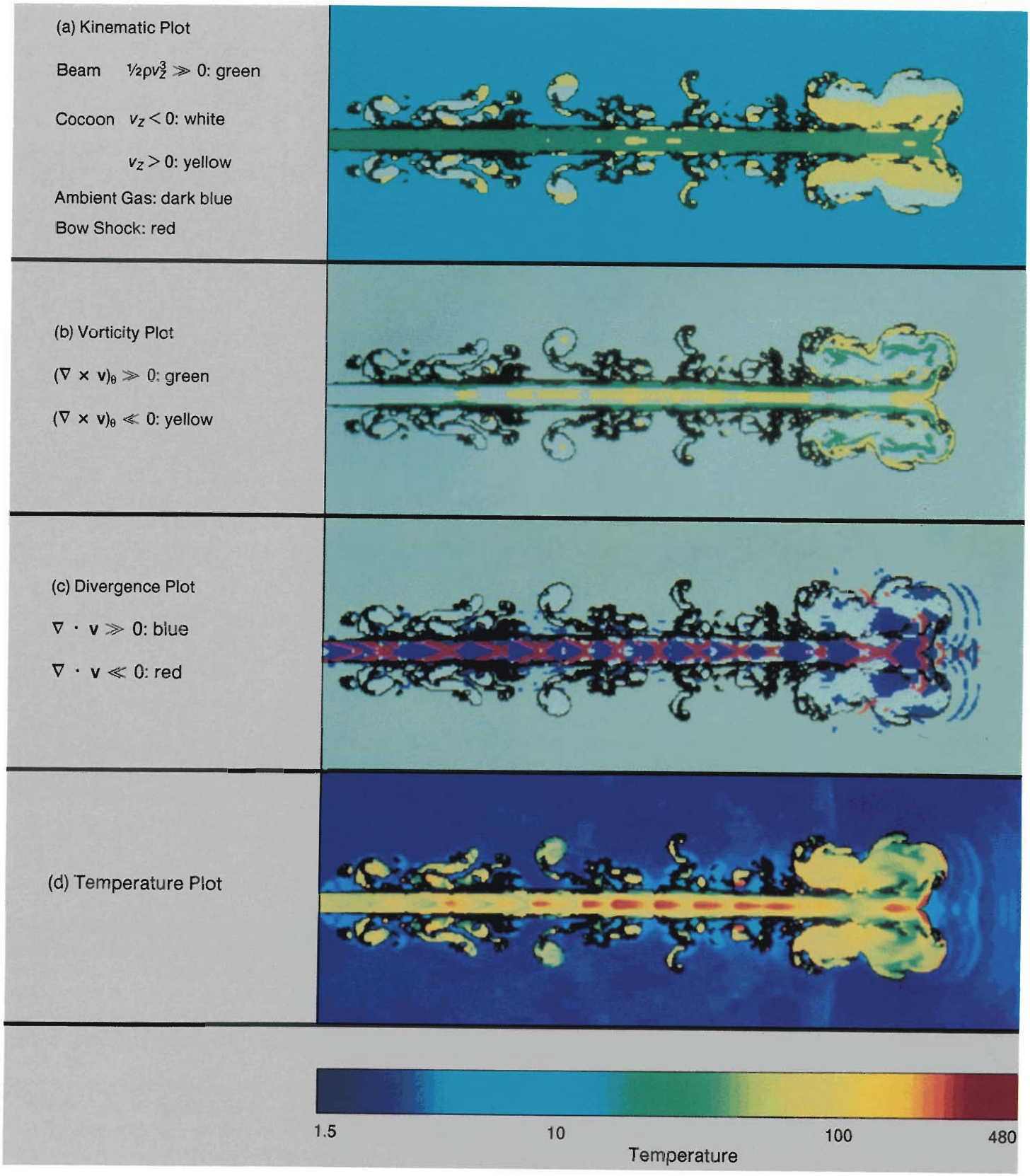
Fig. 7. Computer-graphic diagnostics of the flow field at $t = 60$ for the ($M = 3, \eta = 0.1$) jet of Fig. 4. (a) The kinematic plot depicts kinematically distinct regions in different colors. The beam boundary (black) encloses the beam (green), which by definition consists of those regions for which the kinetic energy flux $\frac{1}{2}\rho v^3$ exceeds 5 percent of the global maximum of that quantity. The backward flowing ($v_Z < 0$) and forward flowing ($v_Z > 0$) cocoon regions are shown in white and yellow, respectively. The undisturbed ambient gas is dark blue, and the bow shock is red. (b) The vorticity plot depicts regions of large positive and negative azimuthal vorticity ($\nabla \times \mathbf{v})_\theta$ in green and yellow, respectively, and reveals the vortex sheet at the beam boundary (green) and the shear layer at the cocoon boundary (yellow). (c) The divergence plot depicts regions of large positive and negative divergence of the velocity ($\nabla \cdot \mathbf{v}$) in blue and red, respectively, and reveals a network of internal shock waves (red) and rarefaction zones (blue). (d) The temperature plot reveals the alternate heating and cooling of the gas by shock and rarefaction waves, respectively.

The stability of the beam boundary is verified in Fig. 7b, where we plot the azimuthal vorticity $(\nabla \times \mathbf{v})_\theta$. Regions with a large positive vorticity are green, and regions with a large negative vorticity are yellow. The beam boundary, a region of high positive vorticity (shear), can be seen as a stable green layer, just inside the black jet boundary, that continues beneath the forward cocoon lobe up to the working surface. Note the coincidence of the beam boundary as defined by the two complementary techniques of Figs. 7a and 7b. At the end of the jet, the shear layer is carried into the cocoon, where it breaks up into isolated vortices. The existence of vortices in the cocoons of diffuse jets is one of our general findings and accounts for the forward flowing cocoon gas indicated in yellow in Fig. 7a.

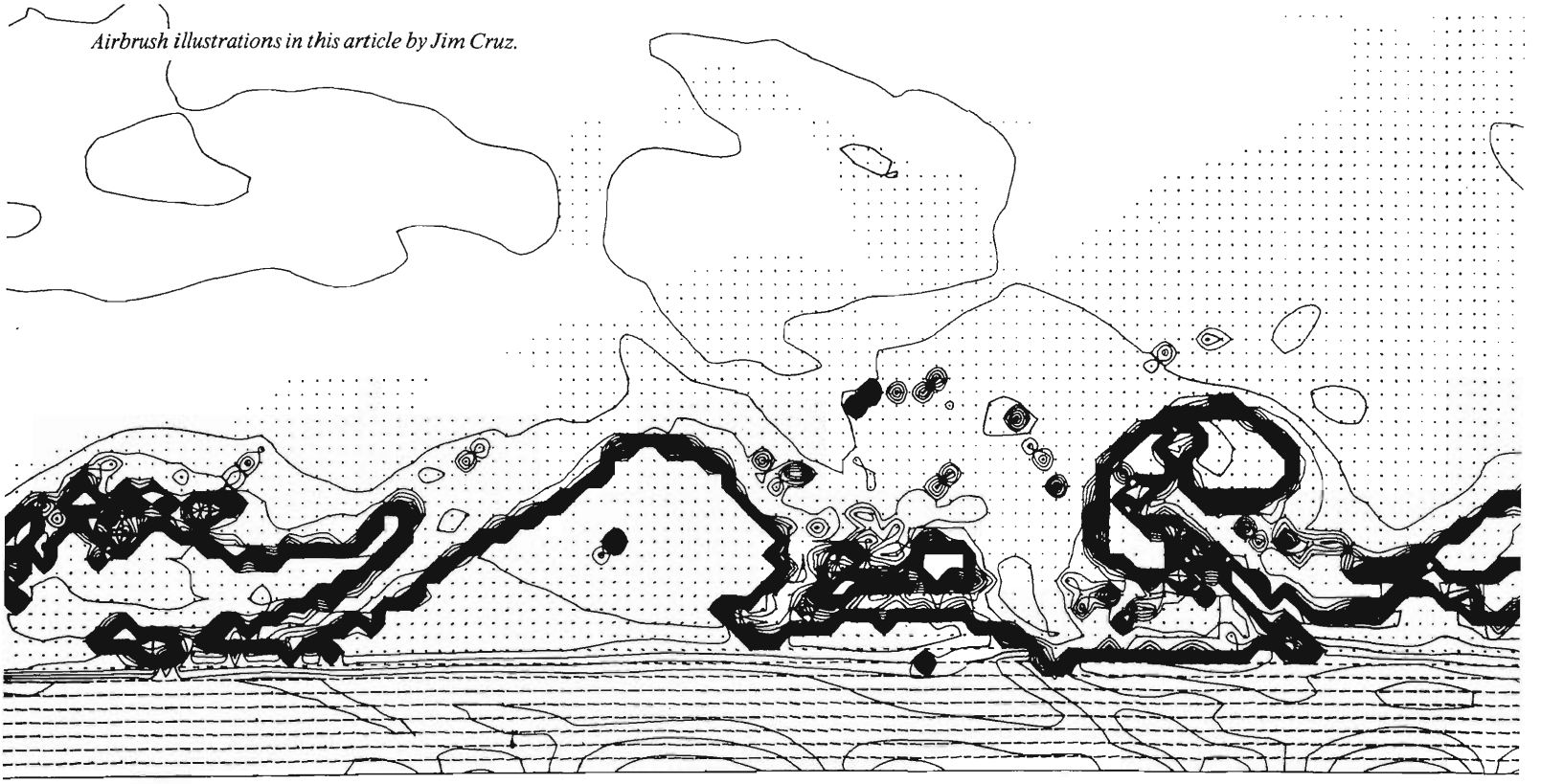
In addition to positive vorticity, Fig. 7b reveals that a substantial amount of negative vorticity is being created within the beam. What is the mechanism responsible for this? The answer is given in Fig. 7c, where regions of large positive flow divergence $\nabla \cdot \mathbf{v}$ are colored blue, and regions of large negative $\nabla \cdot \mathbf{v}$ are colored red. The blue regions are rarefaction zones, and the red regions encompass shock waves (see Sidebars 2 and 3). We see within the beam a well-developed network of crisscrossing shock waves separated by strong rarefaction zones. The negative vorticity is generated by the oblique shock waves, which are highly nonuniform in strength because of their cylindrical convergence toward the axis. The terminal shock at the working surface can also be seen in Fig. 7c just behind the leading contact surface, which is more or less normal to the incoming beam. This leading contact discontinuity separates the jet from the ambient gas. It propagates into the ambient gas, and no gas flows across it in either direction.

A color-coded temperature plot (Fig. 7d) clearly shows that the effect of the embedded shock waves is to compress and heat the beam gas. Knots of shock-heated gas can be seen downstream of each oblique shock intersection. Such knots frequently appear in supersonic rocket plumes as bright patches, which are due to enhanced thermal emission or to chemiluminescence.

Although the shock pattern in our simulation of a pressure-matched jet bears remarkable similarities to the shock waves characteristic of the overexpanded and underexpanded jets produced in the laboratory (see Sidebar 3), there is a vital difference. In the laboratory jets the shock pattern is stationary, whereas in the pressure-matched jet it moves downstream at a speed equal to a significant fraction of the beam speed. We shall return to this point later.


(a) Kinematic Plot
Beam $\frac{1}{2}\rho v_z^3 \gg 0$: green
Cocoon $v_z < 0$: white
$v_z > 0$: yellow
Ambient Gas: dark blue
Bow Shock: red
(b) Vorticity Plot
$(\nabla \times \mathbf{v})_\theta \gg 0$: green
$(\nabla \times \mathbf{v})_\theta \ll 0$: yellow
(c) Divergence Plot
$\nabla \cdot \mathbf{v} \gg 0$: blue
$\nabla \cdot \mathbf{v} \ll 0$: red
(d) Temperature Plot
1.5
10
100
480
Temperature

Airbrush illustrations in this article by Jim Cruz.



The Working Surface

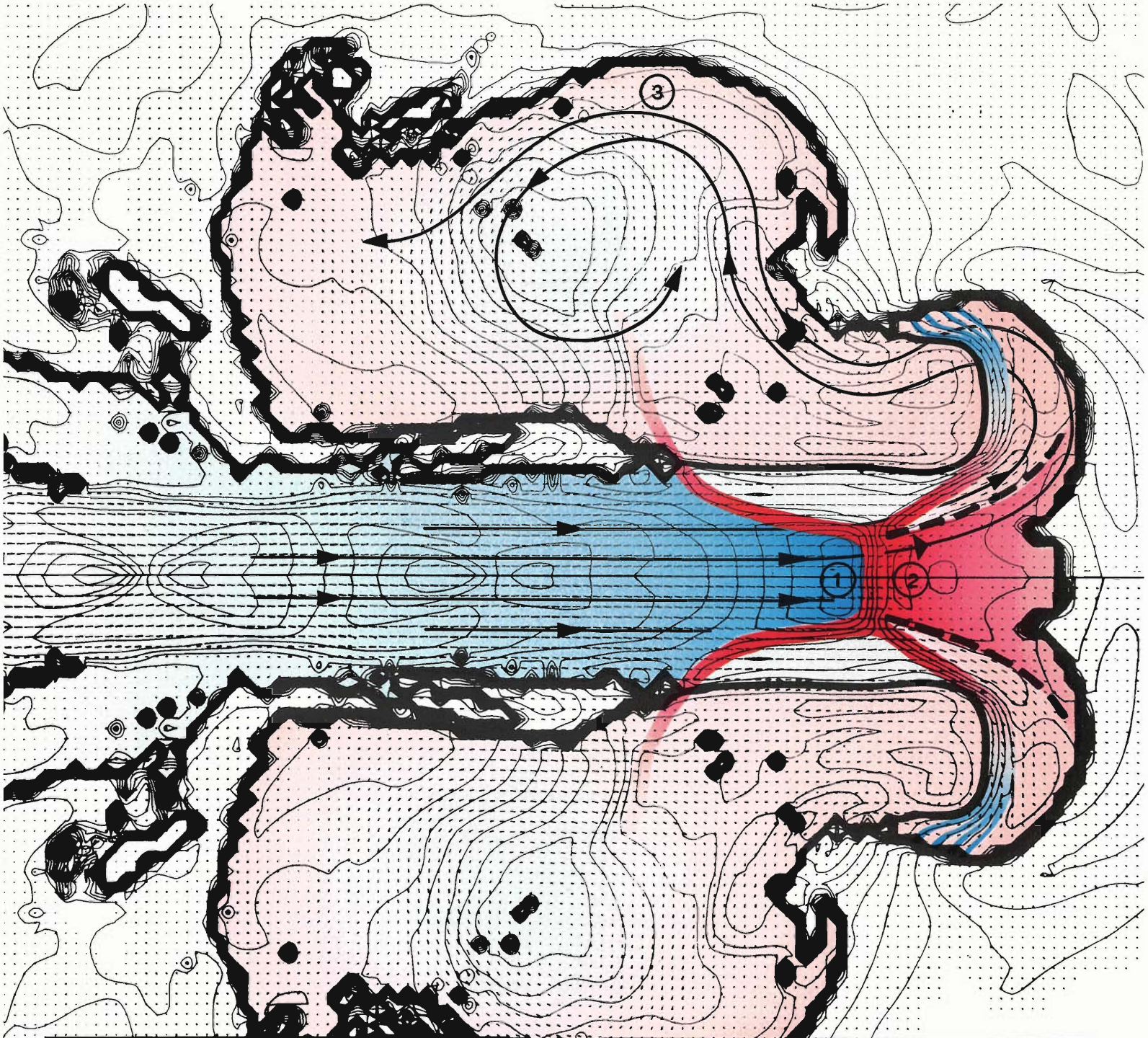
We now turn our attention to the working surface itself and inquire how a supersonic flow can be deflected sharply backward to supply the cocoon with gas. The detailed structure of the working surface (Fig. 8) reveals a complex network of nonlinear waves and flow patterns that propagates as a coherent whole at the head of the jet.

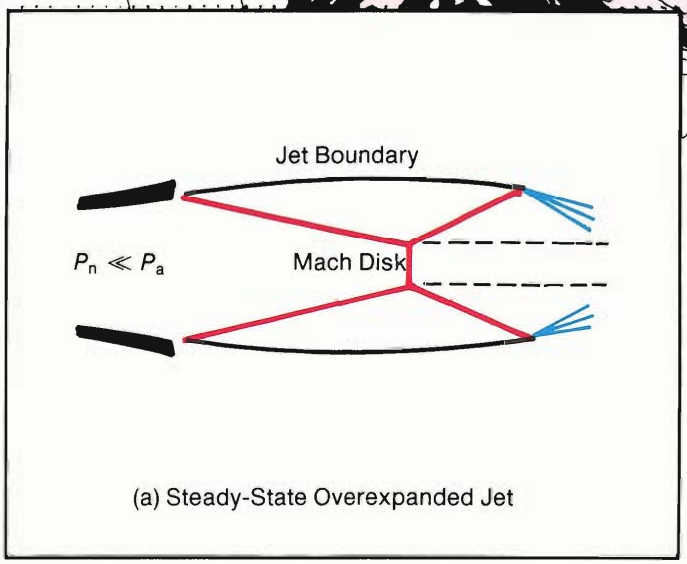
Quite generally, the terminal shock system looks just like the triple-shock structure formed by Mach reflection in steady-state jets (see inset (a) in Fig. 8 and Sidebar 3), except that the structures downstream of the Mach disk (the reflected shock and rarefaction zones) are swept back by the oncoming ambient gas. Upon passing through the Mach disk, beam gas is virtually brought to rest with respect to the leading contact discontinuity, which is itself advancing into the ambient gas. The Mach disk converts the directed kinetic energy of the beam into internal energy, thereby creating a region of very high pressure and temperature between the Mach disk and the contact discontinuity. It is generally believed that this region corresponds to the hot spots observed in the radio lobes of extragalactic jets. The shocked high-pressure gas within this zone can only escape sideways, being confined longitudinally by the ram pressure of the opposing beam and ambient streams. Once beyond the radius of the beam, however, a fluid element will be blown back into the cocoon by the ram pressure of the ambient gas, which is there not balanced by the ram pressure of the beam. As the gas expands into the cocoon, its pressure decreases until it reaches the ambient pressure. Inset (b) in Fig. 8 traces the irreversible thermodynamic evolution of a fluid element following the streamline in Fig. 8 that passes through the Mach disk and into the cocoon. A key consequence of the irreversible nature of the terminal shock front is that the cocoon gas is generally less dense than the beam gas after pressure balance has been re-established.

Vorticity is generated at the shock triple point in the form of a slip discontinuity (the dashed line in Fig. 8), is carried into the cocoon, and contributes to the formation of a large-scale toroidal vortex surrounding the beam. The streamline in Fig. 8 that flows above the shock triple point circles the vortex to produce a forward flowing stream of high-pressure cocoon gas. This high-pressure environment in the cocoon drives the oblique incident-shock component of the beam's terminal triple shock. This intimate relation between the flow structures at the working surface and the adjacent cocoon constitutes a nonlinear self-consistent solution to the problem of flow deceleration and deflection in supersonic jets.

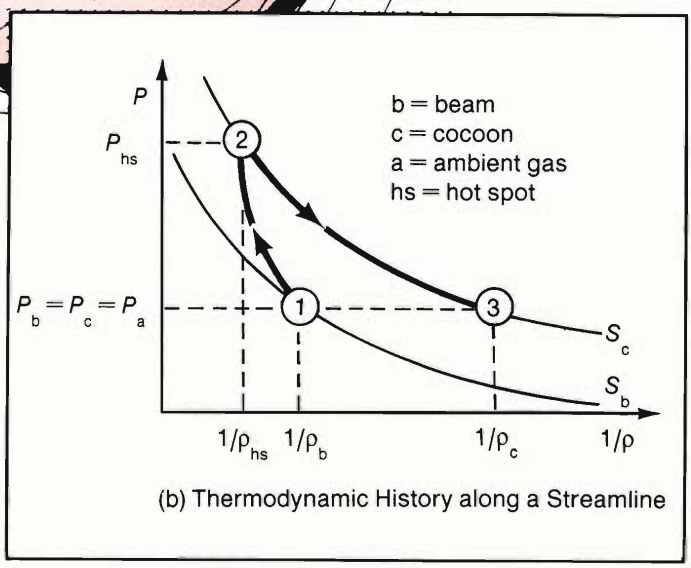
How can we be sure that our numerical techniques are giving us the correct solution to this time-dependent problem when no

Fig. 8 The working surface of a supersonic jet. The terminal shock structure in a calculated isodensity contour plot of the ($M = 3$, $\eta = 0.1$) jet of Figs. 4 and 7 has been highlighted to reveal its resemblance to the triple-shock configuration of a Mach reflection. The incident shock, the Mach disk, and the reflected shock are dark red, and the rarefaction fan is dark blue. This terminal shock structure is similar to the Mach reflection in a steady-state overexpanded ($K < 1$) jet shown in inset (a) (and to the Mach reflection in a steady-state underexpanded jet presented in Sidebar 3), except that the reflected shock and rarefaction fan are swept back by the pressure of the oncoming ambient gas. The gas temperature varies from low (blue) to high (red) according to the scale shown in Sidebar 3. Inset (b) shows the thermodynamic evolution of a gas element as it travels along a streamline through the Mach disk (point 1 to point 2) and around to the vortex in the cocoon (point 3). By the time the gas element reaches point 3, its pressure matches that of the beam and ambient gases, but its density is much lower than before being shock-heated by the terminal shock. (This irreversible process was discussed in Sidebar 1.)



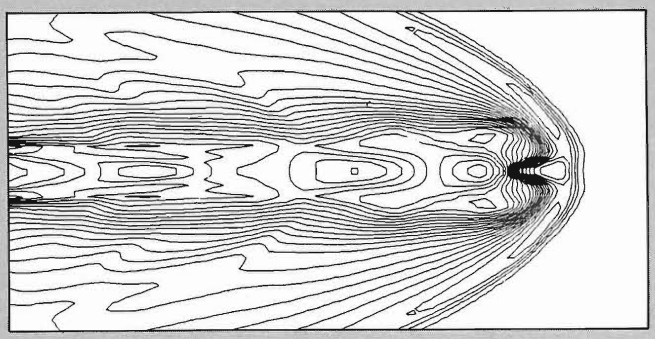


(a) Steady-State Overexpanded Jet



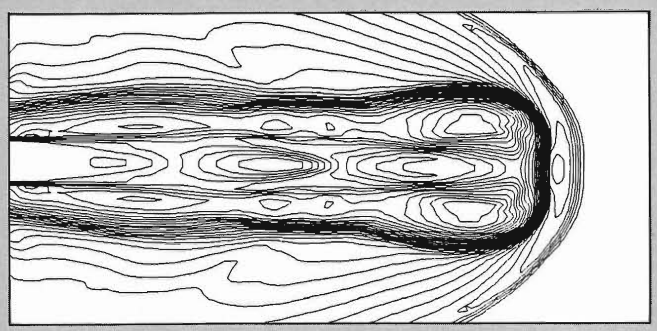
(b) Thermodynamic History along a Streamline

(a)



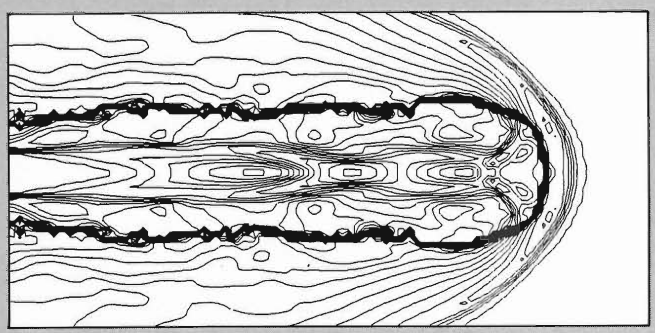
Grid size: one-quarter of the beam radius in both dimensions.
Algorithm: donor-cell advection.

(b)



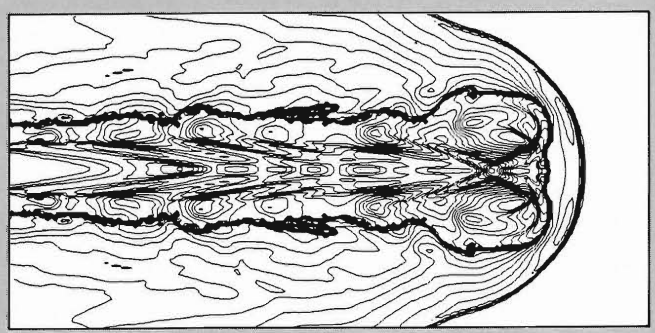
Grid size: one-quarter of the beam radius in both dimensions.
Algorithm: van Leer monotonic advection.

(c)



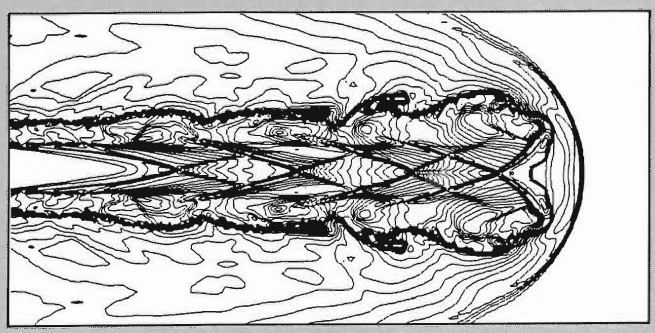
Grid size: one-quarter of the beam radius in both dimensions.
Algorithm: van Leer monotonic advection with a multi-material interface to track the jet boundary.

(d)



Grid size: one-eighth of the beam radius in both dimensions.
Algorithm: van Leer monotonic advection with a multi-material interface to track the jet boundary.

(e)



Grid size: one-sixteenth of the beam radius in both dimensions.
Algorithm: van Leer monotonic advection with a multi-material interface to track the jet boundary.

analytic solutions or experimental data exist for comparison? In principle, we cannot be sure; in practice, however, we *can* verify, through a convergence study, that our simulation captures the important physical structures and processes. Figure 9 displays the results of a convergence study in which we investigated the effects of numerical techniques and grid resolution on the structure of a ($M = 6, \eta = 0.1$) jet. Scanning the figure from top to bottom, one sees the flow field come into focus as the accuracy and capability of the algorithms are improved and as the grid resolution in each dimension is doubled and then doubled again. Note that all of the key flow features (beam, cocoon, working surface, and embedded shocks) revealed at the high resolution of Fig. 9e also appear in the solution computed at half that resolution, Fig. 9d. The same cannot be said of Fig. 9c.

We used the resolution of Fig. 9d for our simulations but increased the number of zones by a factor of 4 to extend the spatial domain beyond that of Fig. 9. Many extragalactic jets are very long, with length-to-width ratios of up to 100.Our 75-by-640 grid allows us to study jets with lengths of up to 40 jet diameters—well within the range of observed extragalactic jets. Figure 9, as well as the consistency of Fig. 8 with "textbook" supersonic flow structures, gives us confidence that our numerical simulations are producing reliable results.

We have delved into the structural details of a single jet to illustrate the similarities and differences between laboratory jets and our simulations. In the remainder of this article we explore how both jet structure and stability change as the parameters M and η are varied.

We did numerical simulations corresponding to thirteen points in the (M, η) parameter space. We varied M between 1.5 and 12 and η between 0.01 and 10. Jet power, which is proportional to $M^3/\sqrt{\eta}$, varied by roughly four orders of magnitude between the lowest power, ($M = 1.5, \eta = 1$) jet and the highest power ($M = 12, \eta = 0.01$) jet. Although jets of even lower density and higher Mach number may be relevant to astrophysics, our survey has revealed the important trends that seem applicable to astrophysical objects. It should be noted that our survey covers a range of parameters that far exceeds what has been investigated in the laboratory. Laboratory jets rarely exceed Mach 3 and are almost exclusively more dense than the ambient gas. As we shall see, the density ratio—a parameter that is seldom varied systematically in laboratory studies—has a profound effect on jet morphology and stability.

Jet Morphology

Figure 10 tells the story of jet morphology in snapshots of our simulated jets arranged according to their values of M and η. All the jets are pictured at the same length.

Two general trends are apparent. Scanning the third column from high to low density ratios, we see a marked thickening of the jet. A jet with a low density ratio ($\eta < 1$) builds a thick cocoon of backward flowing gas around itself as it propagates, whereas a jet with a high density ratio ($\eta > 1$) is just a naked beam, lacking a cocoon altogether. Scanning the third row from low Mach number to high Mach number, we see that the spatial extent of the cocoon increases. At low Mach numbers the cocoon is but a lobe of material at the head of the jet, whereas at high Mach numbers the cocoon surrounds the central supersonic beam along its entire length. The combined effect of these trends is that only jets with low density ratios and high Mach numbers possess extensive cocoons.

These qualitative trends can be understood with the help of a few simple concepts and equations. As mentioned earlier, the cocoon is supplied by gas that flows through the jet head at a rate proportional to $v_b - W$. Equation 1 tells us that this rate is small in high-η jets, since $W \simeq v_b$ for $\eta \gg 1$, and large in low-η jets, since $W \ll v_b$ for $\eta \ll 1$. Thus, on kinematic grounds alone one would expect an inverse correlation between density ratio and the size of the cocoon.

This inverse correlation is enhanced by the shock heating of the jet gas at the terminal shock. The amount of shock heating at the terminal shock is related to M_W^2, where the relative Mach number M_W is defined as $(v_b - W)/c_b$. With the help of Eq. 1 and the definition of M, we find that $M_W = M/(1 + \sqrt{\eta})$. We can see that for a given M, M_W, and thus the amount of shock heating, is largest when η is smallest. The greater the shock heating, the more the gas flowing into the cocoon must expand to match the ambient pressure. Thus, a small η implies not only a large rate of flow into the cocoon but also a greater expansion of the shock-processed gas. The result, in agreement with our simulations, is a cocoon with a relatively large volume and low density.

A similar argument, but now with η held fixed, suggests that larger Mach numbers imply greater shock heating, greater pressure differences between the working surface and

Fig. 9 Effects of resolution (dimensions of finite-difference grid) and algorithmic accuracy on jet structure, as revealed by isodensity contours, for a supersonic jet with initial parameters $M = 6$, $\eta = 0.1$, $K = 1$, and $\theta = 0$. (a) A low-resolution, first-order-accurate computation shows little internal structure of the jet and excessive smearing of the jet boundary. (b) A low-resolution, second-order-accurate computation shows a sharpening of internal structure and boundary. (c) A low-resolution, second-order-accurate computation with interface tracking shows a sharply defined jet boundary but little change in internal structure. (d) A computation with the same algorithms as in (c) but twice the resolution reveals considerable internal structure, including the central supersonic beam, internal shock waves, the cocoon, and waves at the boundary. (e) A computation with the same algorithms as in (c) and (d) but twice the resolution of (d) shows sharply focused internal structure with qualitatively the same features as in (d).

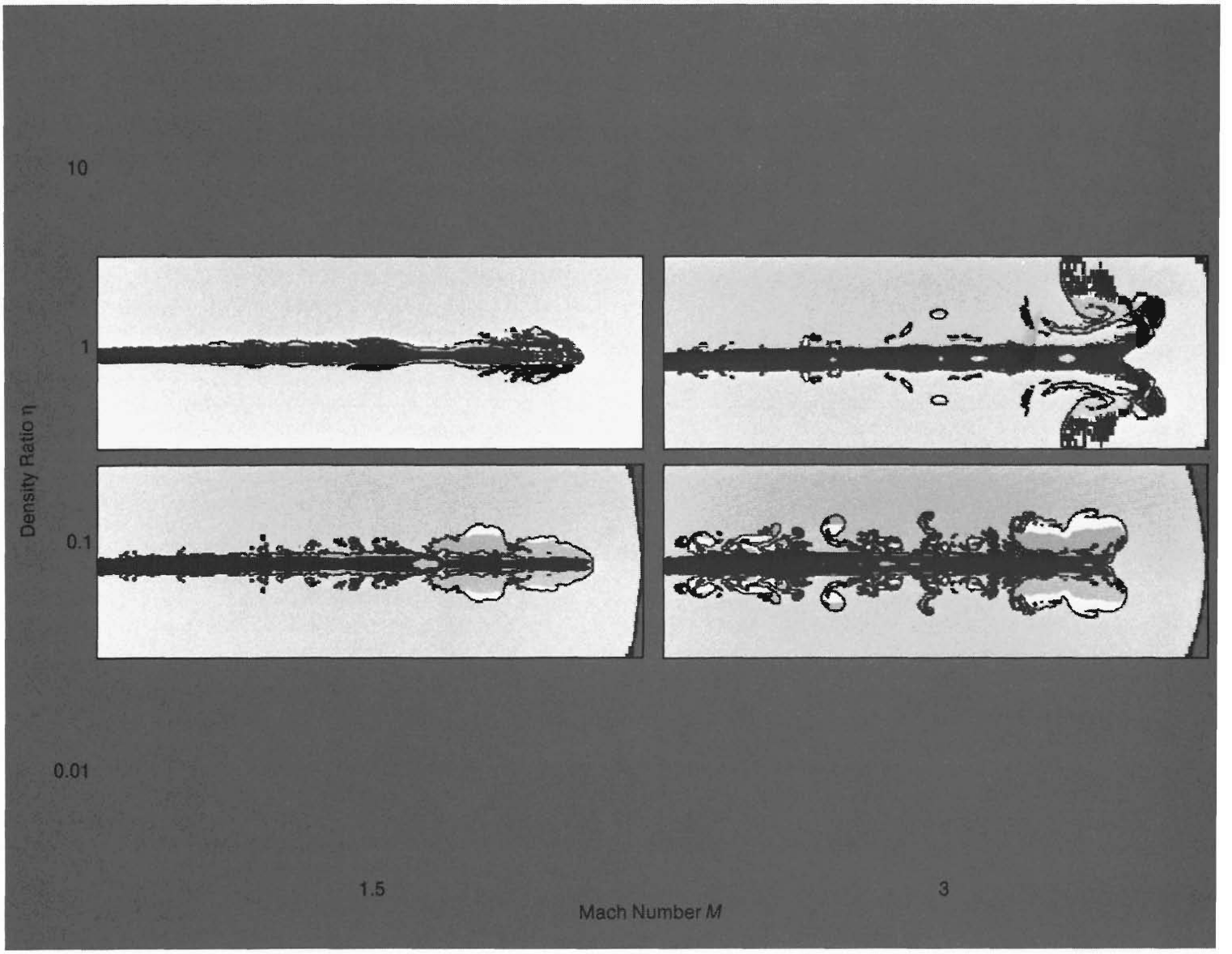
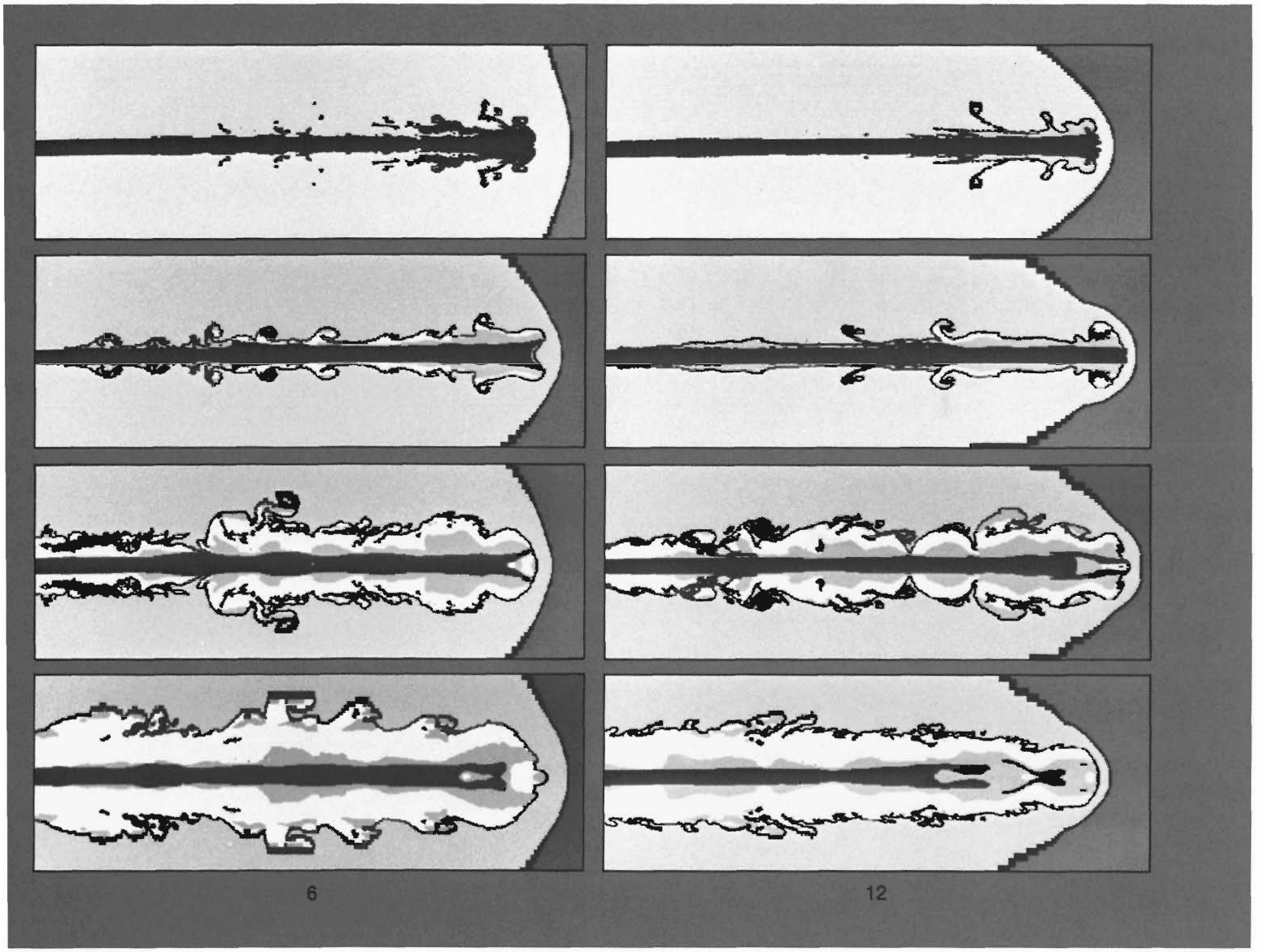


Fig. 10. Dependence of the flow morphology of pressure-matched supersonic jets on the dimensionless parameters M and η. In these kinematic plots the supersonic beam and the bow shock are shown in black, the backward flowing ($v_z < 0$) cocoon gas in white, the forward flowing ($v_z > 0$) cocoon gas in gray, the undisturbed ambient gas in dark gray, and the disturbed ambient gas in light gray. The simulations reveal an inverse dependence of cocoon thickness on density ratio η and a direct dependence of cocoon length on Mach number M.

the ambient gas, and thus cocoons with larger volumes.

Shear instabilities on the cocoon boundary tend to limit the extent of cocoons for jets with low Mach numbers. The gas in the cocoon generally flows back toward the source, as indicated by the white regions in Figure 10. This backflow is resisted by corrugations in the cocoon boundary, such as are produced by the Kelvin-Helmholtz instability. At low Mach numbers the combined effects of low pressure difference and instability-induced flow resistance serve to localize the cocoon gas to a lobe near the head of the jet.

Can we use these morphological trends to infer the physical conditions in extragalactic radio sources? In principle, yes, but translating from a fluid flow field to a radio surface-brightness distribution is not straightforward. The main uncertainty is the quantitative relationship between synchrotron emissivity and thermal energy in the flow field. In the simplest model one assumes that a small but constant fraction of the thermal energy density is equally divided between the radiating relativistic electrons and the mag-



netic field. One then can derive a completely local prescription relating the total synchrotron emissivity to the thermal pressure. This prescription has the form P^q, where $1.75 \leq q \leq 2$. We have derived hot-spot morphologies by applying this prescription to the working surface and have reproduced many features observed in the hot spots of radio lobes. However, mapping lobes and cocoons will require a more detailed emission model that takes into account the sources and sinks of the energetic electrons, in particular, the radiative and adiabatic expansion losses experienced by a parcel of synchrotron-emitting plasma as it flows from the working surface into the cocoon.

When observed at low frequencies, powerful double radio sources generally exhibit extensive cocoons with mean diameters far in excess of their hot-spot diameters, which we may take as measures of the beam diameters. The morphological trends discussed above imply very large Mach numbers ($M \gg 10$) and very low density ratios ($\eta \ll 0.01$) for these jets. Since the kinetic power of a jet varies as $M^3/\sqrt{\eta}$, jets with the thickest, most extensive cocoons should be the most powerful, in agreement with observation.

Although we understand the trends in morphology on a qualitative level, our attempts to quantify these effects analytically have not been successful. Nonlinear and time-dependent effects are just too complicated. For example, the terminal-shock structure is complex and time-dependent, which implies that the cocoon gas is highly inhomogeneous. These inhomogeneities lead to mass motions (vortices) within the cocoon that perturb the central beam and ultimately contribute to the unsteadiness of the terminal shock. In the face of such complex and coupled effects, numerical simulations are the best predictive tool we have for the structure of time-dependent jets.

Pinch Instabilities

SIDEBAR 4

and the Bernoulli Effect

Jet Stability

Numerical simulation is also a powerful tool for studying the effects of instabilities in time-dependent jets. This approach complements the traditional analytic tool of linear perturbation theory, which completely characterizes all modes of instability only in the small-amplitude limit. In contrast to linear theory, our simulations show which modes grow to large amplitudes and how they ultimately affect the flow. We now have some tantalizing hints that Kelvin-Helmholtz instabilities, which disrupt laboratory jets with low Mach numbers, may be reduced or totally absent in the high-Mach-number jets emerging from radio galaxies and quasars. In addition, our simulations confirm recent analytic predictions of the existence in supersonic jets of a new family of modes of the Kelvin-Helmholtz instability and characterize for the first time the nonlinear behavior of the new modes. To motivate our discussion, we review the key results gleaned from analytic studies of jet stability.

Linear Stability Theory

A simple analysis of the growth of axisymmetric, or pinch, modes of instability in supersonic beams is given in Sidebar 4, "Pinch Instabilities and the Bernoulli Effect." This analysis predicts that the relation

$$M = 1 + \sqrt{\eta} \qquad (2)$$

defines a stability boundary in the (M,η) plane. For $M < 1 + \sqrt{\eta}$, the pinch modes grow, whereas for $M > 1 + \sqrt{\eta}$, they damp. This analysis applies to a purely longitudinal mode, that is, one that introduces no transverse structure within the perturbed channel. Such a mode corresponds to the fundamental axisymmetric mode of instability.

In 1981 Attilio Ferrari and coworkers in Bologna determined from perturbation theory that a whole family of unstable (that is, capable of growth) higher order modes exists

What happens to a fluid beam when its flow is slightly constricted? Does the flow resist the constriction and remain stable? Or does the constriction grow and pinch off the flow? Here we use a simple argument based on Bernoulli's principle to derive the conditions under which supersonic flow is susceptible to pinch instabilities, or, in other words, is unstable to longitudinal constrictions.

To establish the basic mechanism that drives a pinch instability, we consider first in Fig. A1 the steady flow of an incompressible fluid, such as water, in a pipe with a localized constriction. The velocity and pressure of the fluid far from the constriction are v_0 and P_0, respectively. At steady state the mass flux, $\rho v A$, is constant along the length of the pipe. Since the fluid is incompressible (that is, ρ is constant) the fluid velocity v must increase when the cross-sectional area A decreases. The Bernoulli principle states that along a fluid streamline the total specific energy E remains constant:

$$E = \frac{1}{2}v^2 + \frac{\gamma}{\gamma - 1}\frac{P}{\rho} = \text{constant}. \qquad (1)$$

Consequently, if ρ remains constant and v increases, the pressure P in the vicinity of the constriction must decrease. Figure A2 shows what happens if the channel wall is replaced by a static background fluid of uniform pressure P_0. Since the beam pressure P near the constriction is less than the background pressure P_0, the channel boundary will move inward, further constricting the flow. By the above reasoning further constriction leads to even higher velocities, lower pressures, and even further constriction.

Precisely this sequence of events is responsible for pinch instabilities in subsonic jets. In principle, an initial constriction of arbitrarily small amplitude will be amplified by the Bernoulli effect until the constriction completely pinches off the flow. In actuality, the ambient fluid following the channel boundary inward at the point of constriction is swept into the channel and carried downstream by the flow as in Fig. A3. As a result, pinch instabilities contribute to the mixing of a subsonic jet with its environment.

When the beam is supersonic, two other effects come into play. First and most important, the compressibility of the fluid can no longer be ignored. Thus the constancy of the mass flux no longer implies an increase in velocity with a decrease in area. Instead the density increases so rapidly as the area decreases that the velocity decreases also. A general mathematical expression relating velocity change to area change in a flow channel of slowly varying cross section is

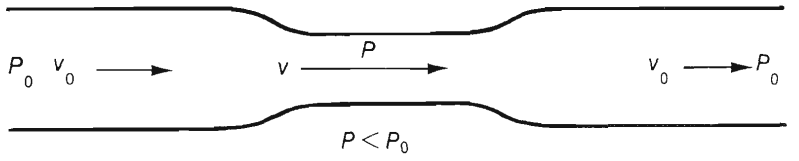
$$\frac{dv}{v} = \left(\frac{v^2}{c^2} - 1\right)^{-1}\frac{dA}{A}, \qquad (2)$$

where c is the speed of sound in the fluid. We see that if the flow is subsonic ($v < c$), its velocity increases ($dv > 0$) at a constriction ($dA < 0$) just as it does in an incompressible fluid, whereas if the flow is supersonic ($v > c$), its velocity decreases ($dv < 0$) at a constriction ($dA < 0$).

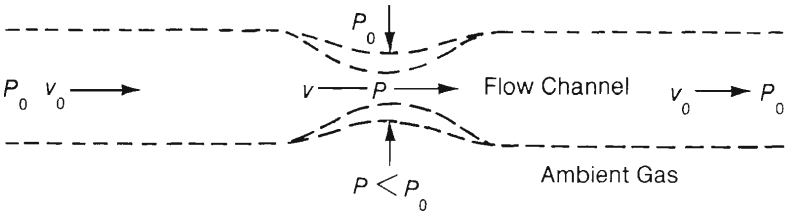
A naive application of the Bernoulli principle would predict that all supersonic beams are stable to the pinch instability, since by Eqs. 1 and 2 the pressure now *increases* at a constriction and thus resists further pinching. This prediction is incorrect because we have failed to take into account the effects of the motion of the boundary in the direction of the flow. As shown in Fig. B1, the instability created when one fluid moves past

Fig. A. Incompressible Flow

(1) One-Dimensional Flow in a Pipe



(2) Pinch Instability in Subsonic Flow



(3) Entrainment of Ambient Medium

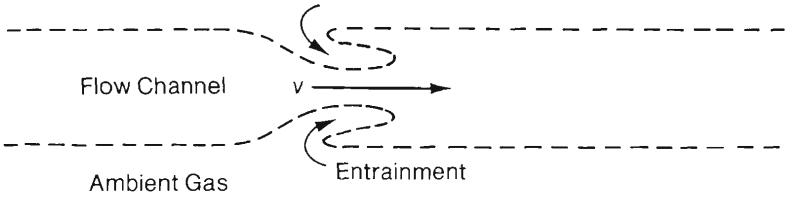
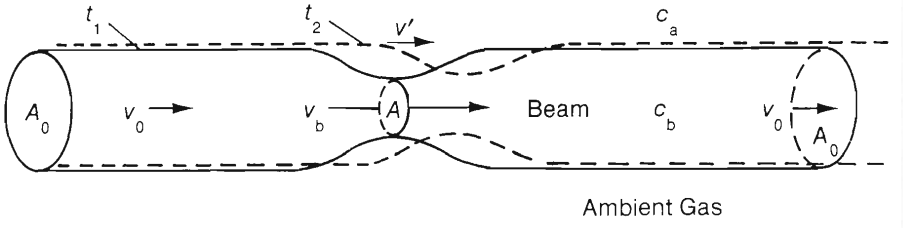
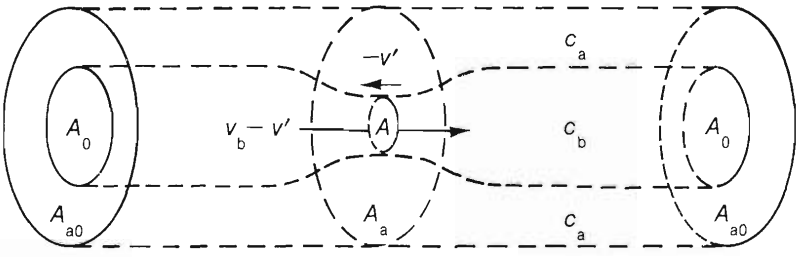


Fig. B. Convective Instability in Compressible Flow

(1) Laboratory Frame



(2) Comoving Frame



SIDEBAR 4

another (the KelvinHelmholtz instability) is a *convective* instability; that is, the waveform of the perturbed boundary is carried downstream with some velocity v'.

Figure B2 shows the flow in a frame moving at velocity v'. In this frame the beam flows through the constriction with a relative speed of $v_b - v'$, and the ambient gas flows past the constriction with a relative speed of $-v'$. Now consider the Bernoulli effect in the two flow channels shown in Fig. B2—one inside the jet boundary and one surrounding it. To guarantee pinching, we require that at the constriction the pressure decrease inside the jet boundary and increase outside the jet boundary. *This will be the case if both the jet gas and the ambient gas move subsonically with respect to the constriction*, since, from Eqs. 1 and 2, only for relative Mach numbers (ratios of flow velocities in the moving frame to sound speed) less than unity will the channel area and the pressure vary in the same direction. Therefore, a sufficient condition for instability is that $v_b - v' < c_b$ and $v' < c_a$, where c_b and c_a are the speeds of sound in the beam gas and the ambient gas, respectively. Adding the two inequalities eliminates the unknown v' and yields the condition

$$v_b < c_b + c_a . \qquad (3)$$

Thus, on very simple physical grounds instability requires that the jet speed be subsonic with respect to the sum of the internal and external sound speeds.

To express this condition in terms of the dimensionless parameters Mach number M and density ratio η, we eliminate c_a by invoking the pressure balance between the jet and ambient gases ($\rho_a c_a^2 = \rho_b c_b^2$) and obtain

$$M < 1 + \sqrt{\eta} . \qquad (4)$$

This relationship, which was first deduced by David Payne and Haldan Cohn from a dispersion analysis of the linearized equations of motion, thus follows simply from the fact that the growth of pinch instabilities requires subsonic flow of both beam gas and ambient gas relative to the moving constriction. ■

at all Mach numbers greater than unity. These so-called reflection modes derive their name from a special property of a supersonic shear surface, namely, for certain angles of incidence, a sound wave will be amplified upon reflection from the surface. A reflection mode instability in a supersonic beam follows from the repeated amplification of perturbations reflected back and forth between the channel walls. A recent analysis of the reflection modes by David Payne and Haldan Cohn has shown that at the resonant wavelength of each mode, the wave fronts of the perturbation meet the beam boundary at the Mach angle. Thus, the reflection modes are largely transverse in character, in contrast to the fundamental mode. Figure 11 shows the structure, as computed by Payne and Cohn with linear theory, of the fundamental pinch instability and of the first and second reflection pinch instabilities.

A key finding of Payne and Cohn is that Eq. 2 defines a *transition* boundary (Fig. 12) in the dominant mode of instability: for $M < 1 + \sqrt{\eta}$, reflection modes are absent and only the fundamental mode exists, in agreement with our naive analysis; for $M > 1 + \sqrt{\eta}$, the reflection modes dominate. This boundary is indicated in Fig. 12 by the upper dashed line. Note that the majority of our simulated jets (solid circles) reside in the reflection-mode regime; in particular, five cases straddle the transition boundary. We can therefore investigate whether indeed a transition in the modal character of nonlinear pinch instabilities occurs, and if so, what effect the transition has on jet propagation.

Instability in the Nonlinear Regime

How do we detect an instability in a numerical simulation? Figure 13 illustrates the growth of a reflection-mode instability in the $(M = 3, \eta = 0.1)$ jet discussed in detail above. Figure 13a is a color-coded plot of gas pressure for this jet at the same instant in its evolution as that depicted in Fig. 7. The continuity of color at the orifice indicates that we are indeed fulfilling our pressure-

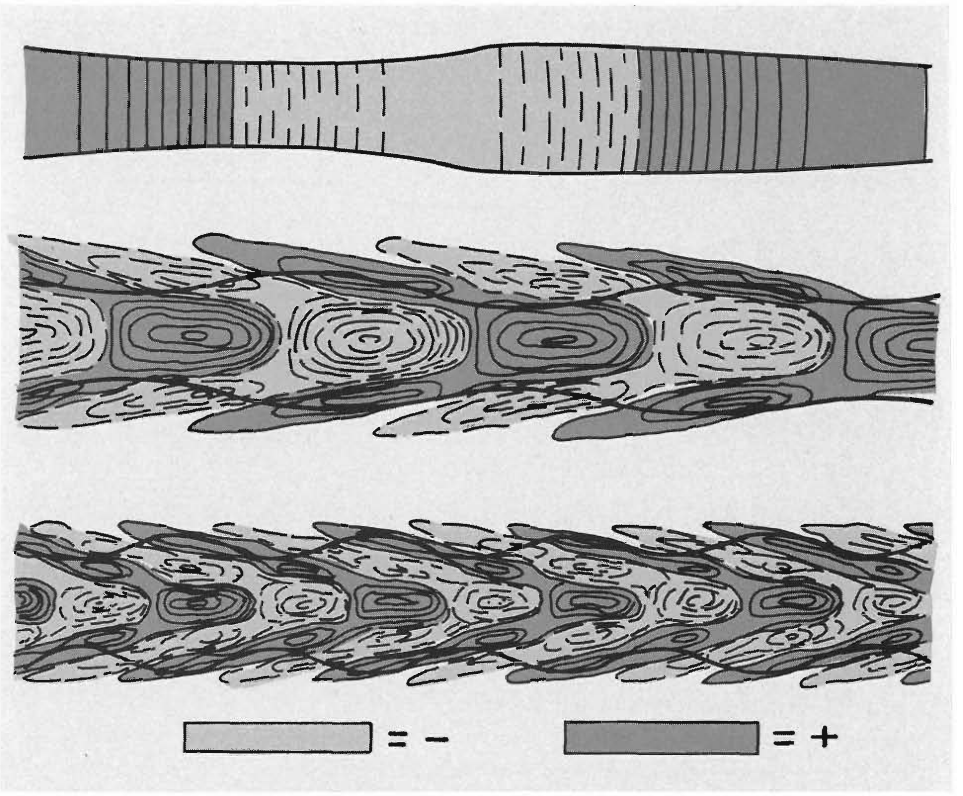


Fig. 11. Structure of the fundamental and first and second reflection pinch instabilities in a $(M = 5, \eta = 0.1)$ supersonic beam as predicted by linear perturbation theory. Dark and light shades of gray correspond to regions of positive and negative pressure about the mean. Whereas the fundamental pinch mode is purely longitudinal, the nth reflection mode possesses n nulls in the radial perturbation function. Adapted from David G. Payne and Haldan Cohn, **Astrophysical Journal** ***291(1985):655-667.***

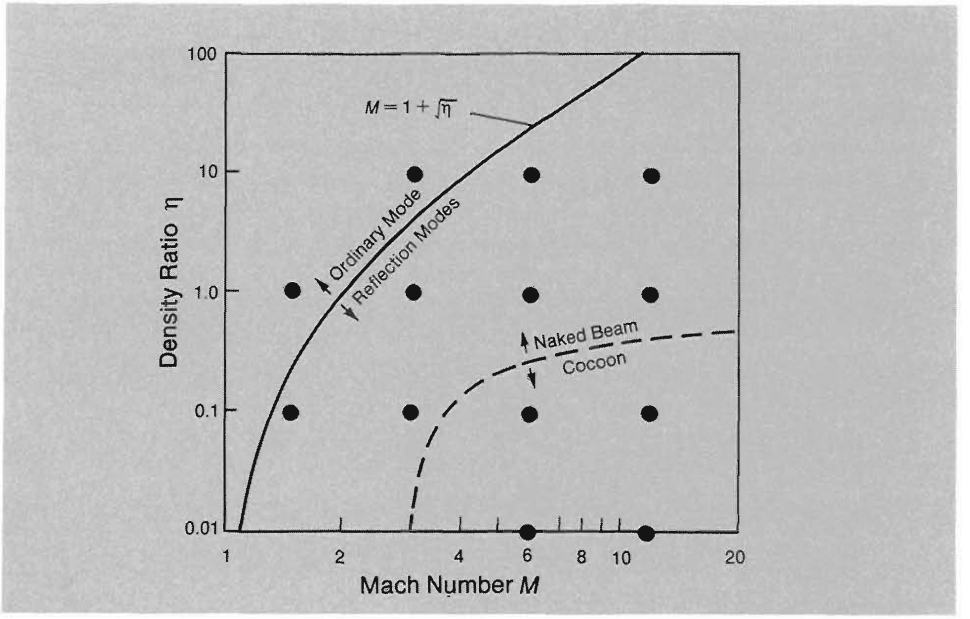


Fig. 12. Transition boundaries in (M, η) parameter space. The transition boundary separating the regions of dominance of the ordinary mode $(M < 1 + \sqrt{\eta})$ and of the reflection modes $(M > 1 + \sqrt{\eta})$ of the Kelvin-Helmholtz pinch instability was analytically determined (see Sidebar 4). The approximate location of the line of demarcation between jets with cocoons and those without cocoons (naked beams) was determined from a visual inspection of the jets displayed in Fig. 10. The solid circles indicate the initial (M, η) values for those jets.

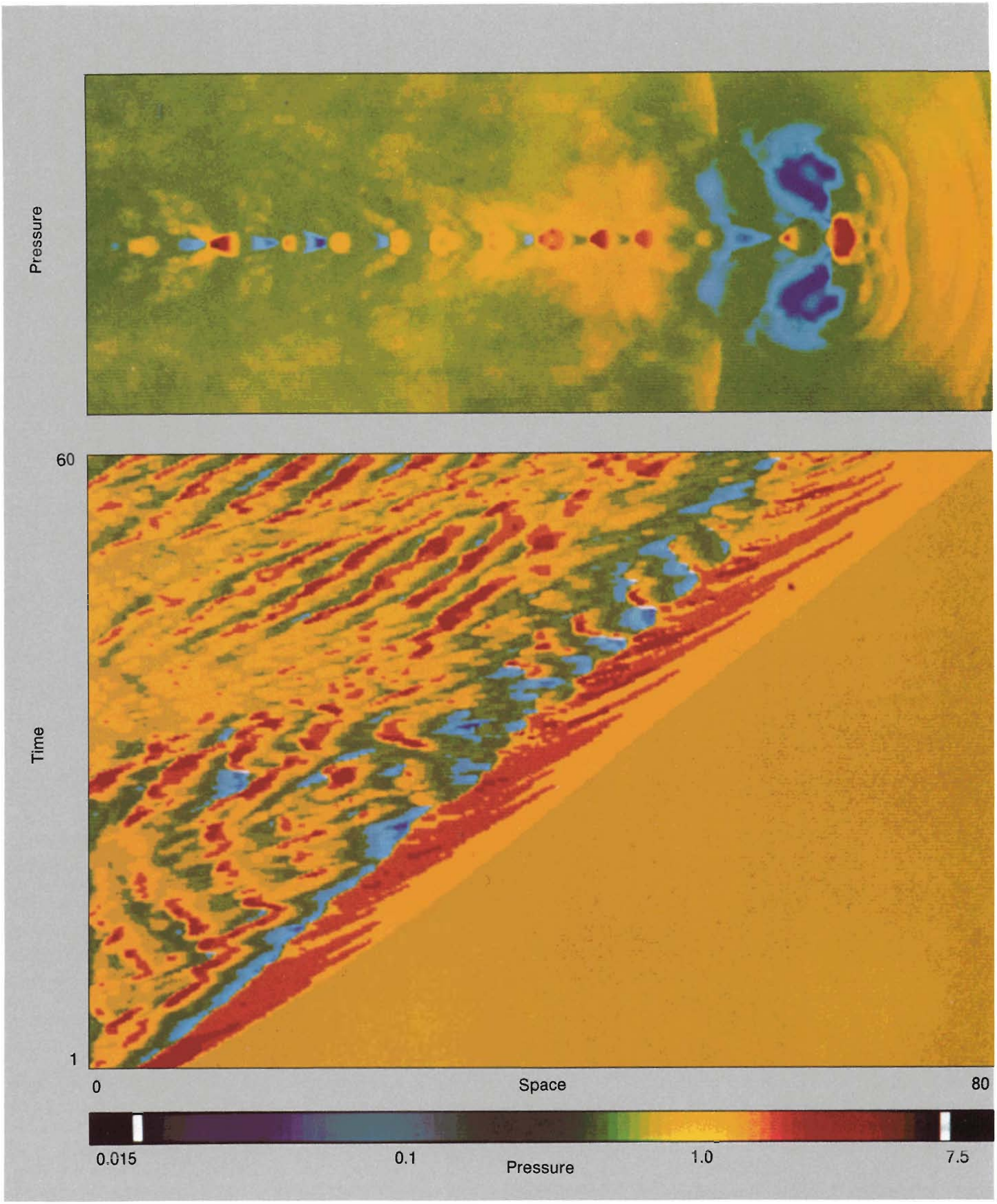


Fig. 13. Pressure fluctuations in the ($M = 3, \eta = 0.1$) supersonic jet of Figs. 4 and 7 caused by the nonlinear growth of a reflection-mode pinch instability. Pressure values vary according to the color code. (a) The spatial pressure distribution at $t = 60$ shows the high-pressure knots and terminal "hot spot" produced by the embedded shock waves displayed in Fig. 7c. (b) The space-time distribution of axial pressure shows the advance of the bow shock into the constant-pressure ambient gas and the emergence and subsequent motion of the knots (red diagonal ridges).

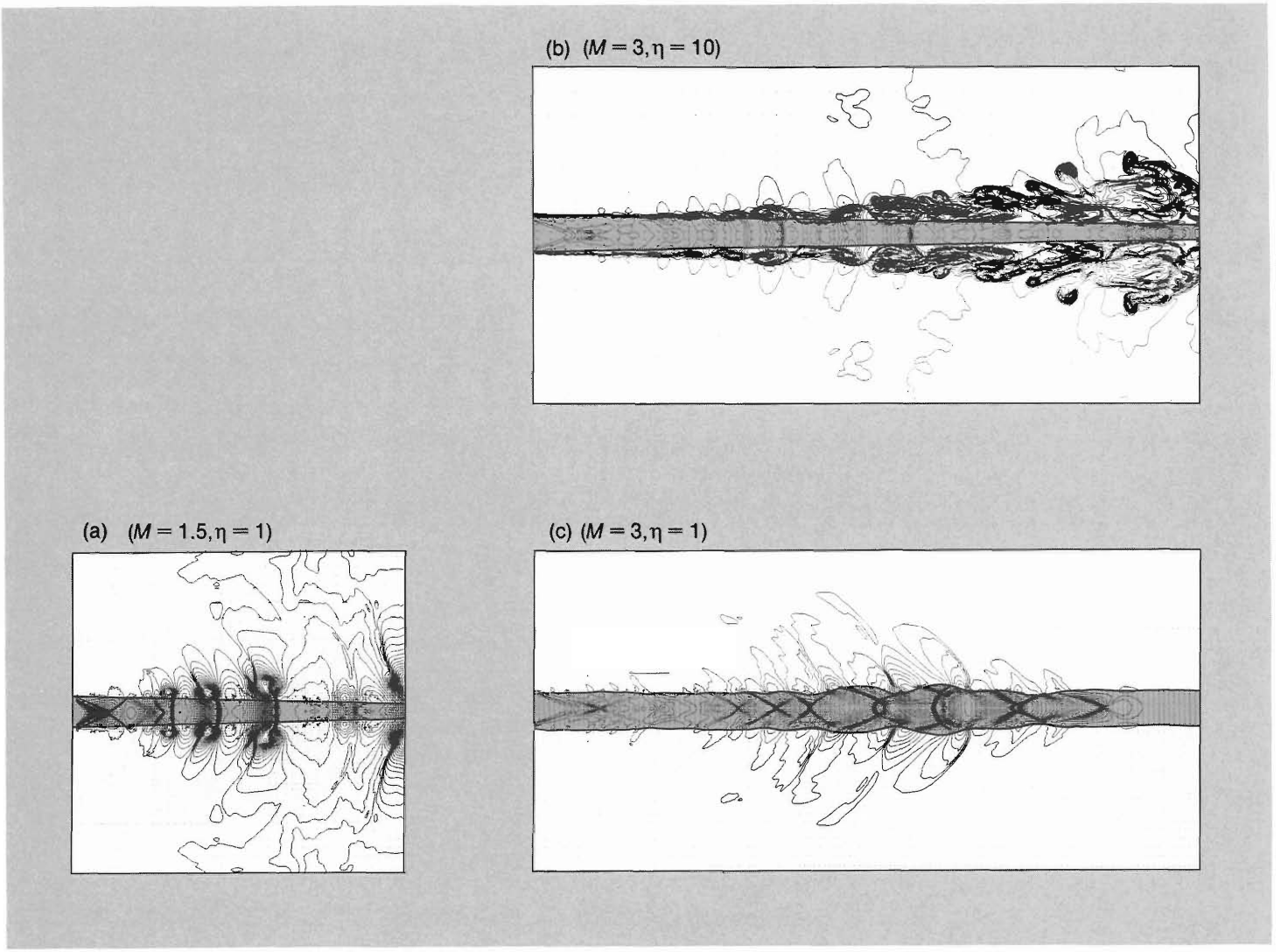


Fig. 14. High-resolution simulations in the nonlinear regime for supersonic jets on both sides of the transition boundary between the fundamental mode and reflection modes of the Kelvin-Helmholtz pinch instability. At $t = 0$ the beam extends across the computational domain, flowing in at the left with fixed parameters and flowing out at the right. The initial jet radius is resolved into twenty computational zones. A perturbation is applied at the inlet by continuously admitting beam gas with a 1-percent pressure excess (that is, $K = 1.01$). The subsequent evolution of the beam is computed by using a numerical fluid-interface technique to track the beam boundary until the instability is well established. (a) Growth of a fundamental pinch instability in a ($M = 1.5, \eta = 1$) beam leads to constriction of the beam boundary (dark line) and entrainment of tongues of ambient gas. Isopressure contour lines reveal the formation of normal shock fronts upstream of the channel constrictions. (b) Growth of a fundamental pinch instability in a ($M = 3, \eta = 10$) beam establishes a mixing layer at the beam boundary. The isodensity contour lines reveal that this mixing layer rapidly spreads into the supersonic core. (c) Growth of the first reflection pinch instability in a ($M = 3, \eta = 1$) beam causes only a mild undulation of the beam boundary and, as revealed by the isopressure contour lines, excites a network of oblique internal and external shock waves.

matching boundary condition ($K = 1$). Nevertheless, large-amplitude pressure fluctuations appear within a few beam diameters downstream of the orifice, caused by the presence there of embedded rarefactions and shock waves (see Fig. 7a). These fluctuations are not tied to the orifice, in marked contrast to the pressure fluctuations in laboratory supersonic jets, which are excited with large amplitudes at the nozzle by the pressure mismatch there and subsequently decay downstream.

We can now understand the origin of the network of oblique shock waves that appears in the beam in Fig. 7a—it is the result of the oblique acoustic disturbances (reflection-mode instabilities) pictured in Figs. 11b and 11c, which have grown to finite amplitude and steepened into shock fronts. The superficial resemblance of the resulting shock network to the diamond-shaped steady-state shock pattern displayed in Sidebar 3 breaks down when we investigate the time evolution of nonlinear reflection-mode instabilities. This evolution is traced in Fig. 13b, a plot of the axial pressure distribution within the jet as a function of time. To orient the reader, we point out that the color discontinuity along the diagonal line from lower left to upper right is the space-time trajectory of the bow shock as it is driven into the ambient gas (which has uniform pressure) by the advance of the jet head. The sharp jump from blue and green to orange and red traces the motion of the terminal shock front. The red diagonal ridges trace the movement of

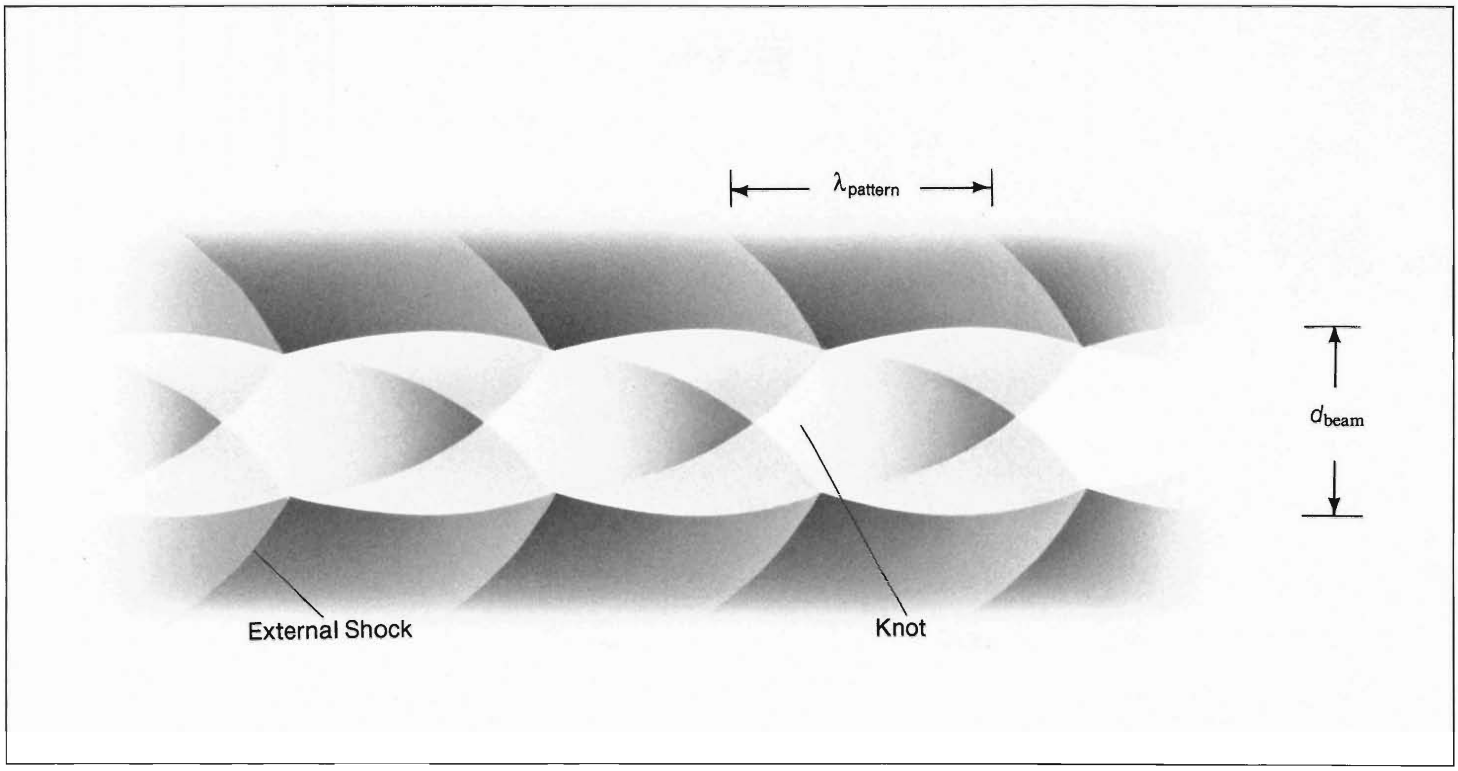


Fig. 15. The shock-wave structure of a nonlinear reflection-mode pinch instability and the resulting temperature distribution. Temperature values are depicted according to an arbitrary scale in which hot is white and cold is dark gray. The pattern created by the saturated oscillation has a wavelength of about 2.5 beam diameters and travels downstream at a speed v_{knot} that is a significant fraction of the beam speed.

the high-pressure knots produced by the embedded shock network and reveal a remarkably well-ordered space-time flow. This pattern indicates that the reflection mode is a convective instability, that is, one that grows in amplitude as the perturbation is convected downstream.

Payne and Cohn suggest that the stationary shock diamonds and high-pressure knots arising in underexpanded ($K > 1$) supersonic jets are a signature of the reflection mode at marginal stability, that is, at a state of zero growth rate and zero phase velocity. Zero growth implies zero damping, and this may well account for the remarkable longevity of the oscillations observed in the plumes of some rocket engines (see Fig. 1). Of course, the irreversible processes of shock heating and turbulence do introduce some damping, and the growth of the turbulent mixing layer at the jet boundary ultimately overcomes the oscillations, as shown in Fig. D of Sidebar 3.

An ever-broadening mixing layer is conspicuously absent in Fig. 7a; the clumps of material above and below the beam are what is left of the cocoon as it mixes with the ambient gas (see Fig. 4). In addition, the nearly constant thickness of the green vortex layer in Fig. 7b shows very little spreading of the beam boundary. Reflection-mode instabilities thus seem to be far less disruptive than the fundamental-mode instability.

To further investigate this interesting result, we performed high-resolution simulations of supersonic beams with parameters on both sides of the transition boundary. The results of these calculations (Fig. 14) support the general conclusion that boundary mixing is greatly reduced in supersonic beams satisfying $M > 1 + \sqrt{\eta}$. A more extensive numerical study of the transition from fundamental mode to reflection mode is in progress.

The key features of the saturated reflection mode are summarized in Fig. 15. The mild undulation of the jet boundary is coupled to the network of oblique internal shock waves. Two subtle but important differences between this wave pattern and the steady-state jet depicted in Fig. A of Sidebar 3 are (1) the reflection of shocks from the boundary as shocks rather than rarefactions and (2) the transmission of a shock into the gas external to the beam. The external shock wave is the mechanism by which the energy of relative motion amplifies the internal reflected shock. Since there is no relative motion between ambient gas and boundary corrugations in underexpanded supersonic jets, no external shock wave is excited in such jets.

We observe the growth of nonlinear reflection modes in all but one of the jets lying within the parameter space below the instability transition boundary and above the boundary for cocoon formation (lower dashed line in Fig. 12). In the one exception, a ($M = 1.5, \eta = 0.1$) jet, the shock waves were too weak to be detected numerically. Since the pattern wavelength of the saturated oscillation is about 2.5 beam diameters in the Mach 3 and Mach 6 jets and only slightly greater at Mach 12, the dependence of wavelength on Mach number is quite weak. This result is consistent with perturbation theory, which predicts a maximum growth rate at wavelengths near the beam circumference. The speed of advance of the patterns varies between 10 and 50 percent of the beam speed and is consistent with the density-ratio scaling given in Eq. 1. This is not surprising because ram-pressure balance in a frame comoving with the oblique shocks determines the advance speed of the patterns, just as it determines the advance speed of the working surface. Quite generally, therefore, the high-pressure knots advance with approximately the same speed as the jet head, as can be seen in Fig. 13b.

In principle, cocoon-shrouded beams are also subject to reflection-mode instabilities. We find, however, that cocoon dynamics puts its own imprint on the internal structure of the beam. At high Mach numbers the backflow within the cocoon is also supersonic, and components of this momentum flux normal to the beam drive oblique shock

waves into the beam (Fig. 16). The perturbations are associated with toroidal vortices within the cocoon that drift back toward the source. The oblique shocks they excite possess a less regular pattern than shocks induced by reflection-mode instabilities, and consequently the high-pressure knots are also less regular.

Astrophysical Implications

Knots of enhanced emission, frequently regularly spaced, are observed in numerous extragalactic radio jets (Fig. 17). The nonlinear reflection-mode instability is an attractive candidate mechanism for producing these knots because of our finding that this instability is ubiquitous in highly supersonic jets. The high-pressure knots created by the instability would appear as regions of enhanced emission because of the P^2 scaling of the synchrotron emissivity mentioned previously.

A prediction of our model is motion of the emission knots. In principle, a measurement of their apparent motion with respect to the galactic nucleus could be used to infer the jet speed and density ratio. Unfortunately, the time scales associated with the evolution of kiloparsec-scale radio jets are too great for this motion to be detectable. Some parsec-scale extragalactic jets and stellar jets also exhibit regular strings of emission knots, and the motion of these knots is detectable with current radio and optical telescopes. Future high-resolution observations of stellar jets with the Hubble Space Telescope and of compact radio jets with the the Very Large Array of radio telescopes can put our model to the test.

Returning to the large-scale radio jets and their remarkable ability to propagate without disruption over intergalactic distances, we propose that the strings of emission knots may be a signature of nature's way out of the stability problem. This may sound contradictory, since the reflection modes that produce the knots are a kind of instability. However, like many other instabilities in

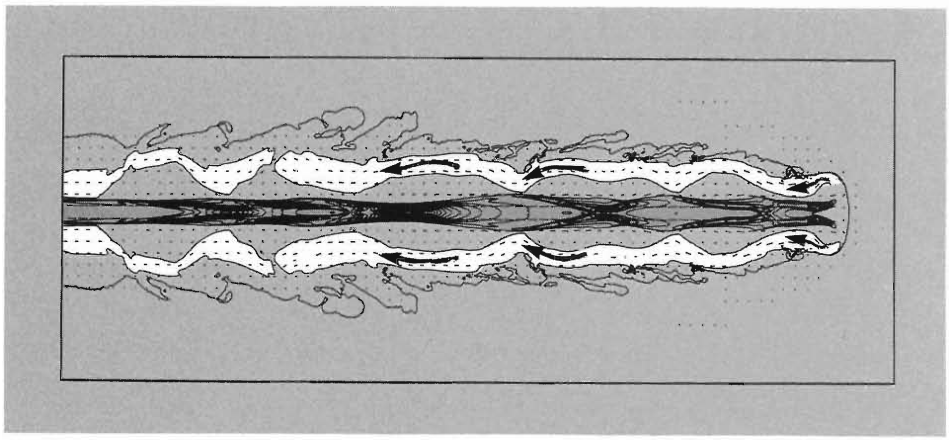


Fig. 16. Interaction of backward flowing cocoon gas (white) with forward flowing supersonic beam gas produces oblique internal shock waves in a ($M = 12, \eta = 0.01$) jet. Isocontours of the axial kinetic energy flux $\frac{1}{2}\rho v_z^3$ are plotted within the jet boundary (heavy line).

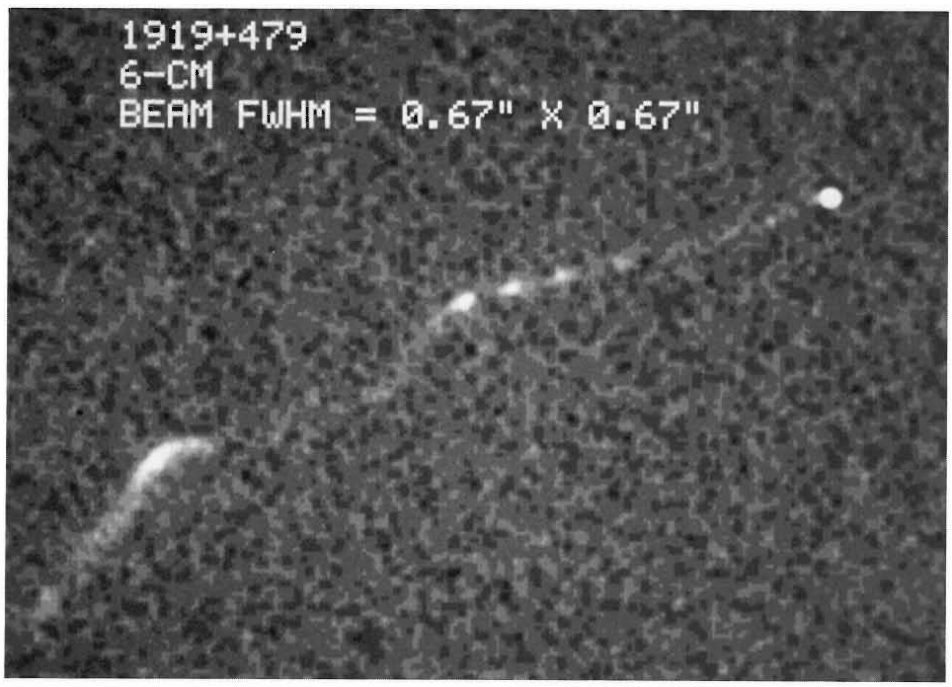
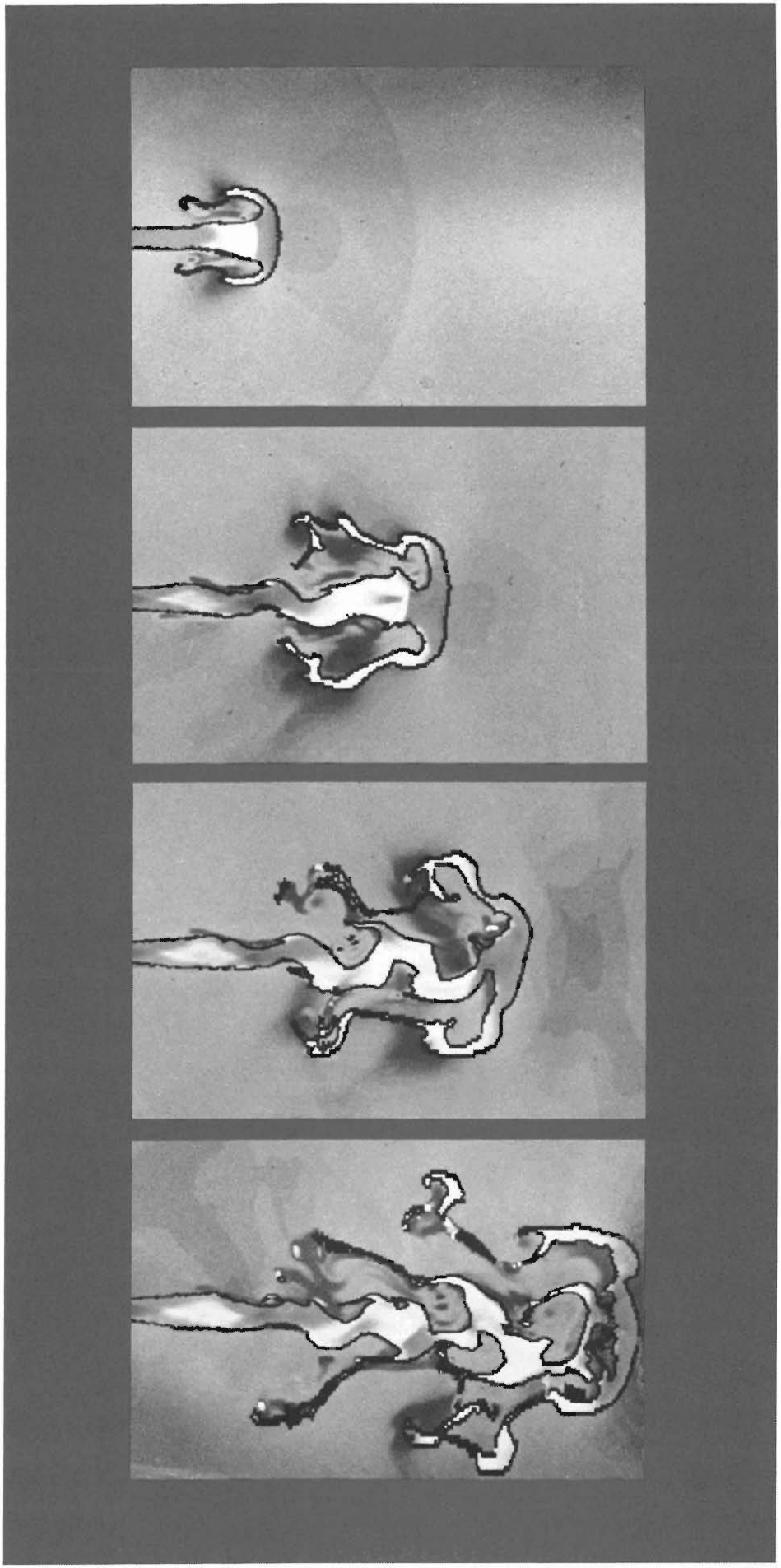


Fig. 17. Radio emission knots in the extragalactic jet 1919+479. Regions of high radio brightness are shown in white. The bright spot to the upper right is the unresolved radio "core" (the source of the jet), which coincides with the nucleus of a giant elliptical galaxy. The onset of large-scale bends in the jet signals its incipient disruption. The radio jet is roughly 300,000 light-years long. (Unpublished VLA image courtesy of J. O. Burns.)



fluids and plasmas, the reflection modes appear to saturate at finite amplitudes and do minimum damage to the flow.

The enhanced boundary stability of the reflection modes is almost certainly exaggerated by our two-dimensional simulations. It is nevertheless known experimentally, but not understood, that the spreading rate of supersonic shear layers drops precipitously beyond Mach numbers of about 5. Another important feature of reflection modes is the existence of oblique internal shock waves. Their oblique character allows the postshock axial flow speed to remain supersonic even though the component normal to the shock is subsonic. These two effects—enhanced boundary stability and minimal flow deceleration—make the reflection mode a benign instability in terms of jet disruption.

In very long jets these small effects may add up to a big effect and ultimately disrupt the flow. Quite often the signature of incipient disruption is the onset of large-scale wiggles in the jet, as can be seen in Fig. 17. This kind of behavior, which is observed experimentally in fluid jets and magnetically confined plasma columns, is a consequence of the growth of nonaxisymmetric instabilities to amplitudes at which nonlinear effects become important. These so-called garden-hose instabilities are fundamentally three-dimensional in character and therefore cannot be studied within the context of two-dimensional axisymmetric numerical simulations. However, two-dimensional

Fig. 18. Numerical simulation of the growth of a kink instability in a two-dimensional "slab" jet with $M = 3$ and $\eta = 10$. The variable plotted is density. The instability is excited at the inlet by imposing on the beam a sinusoidally varying transverse velocity with an amplitude equal to 5 percent of the beam's axial velocity. The frequency of the perturbation corresponds to the most unstable wavelength: $\lambda = 2d\sqrt{M^2 - 1}$. The jet diameter d is divided uniformly into forty zones.

simulations of a planar "slab" jet do allow us to study the effects of kinking on the flow.

We performed such a simulation for a slab jet with parameters in the ordinary-mode regime, namely $M = 3$ and $\eta = 10$. As shown in Fig. 18, the amplitude of the oscillations grows rapidly, thereby destroying the directionality of the flow. A similar calculation for a slab jet with values of M and η in the reflection-mode regime also shows growth of kinks, but the kinks grow much more slowly and are of longer wavelength. As we might expect, oblique internal shock waves accompany the bends. These shocks become quite strong as the bend angle becomes large and produce high-pressure knots at the bend. This is a likely explanation for the origin of the bright emission knot at the bend in the radio jet shown in Fig. 17.

Both the axisymmetric and nonaxisymmetric instabilities derive from the interaction between a moving fluid stream and a confining medium. Although the pressure of the intergalactic gas appears adequate to confine the vast majority of radio jets, evidence is accumulating that the jets emerging from distant quasars are too highly pressurized to be confined by their local environment. Magnetic confinement, which is often invoked to explain these observations, would require that the jets carry substantial currents. As we have no unambiguous evidence for the presence or absence of such currents in radio jets, magnetic confinement remains just a hypothesis. It is curious and perhaps significant that Payne and Cohn find that the properties of the linear reflection modes are quite insensitive to the confinement mechanism, be it pressure or magnetic forces. If this finding holds also in the nonlinear regime, strings of emission knots may well be a universal signature of confined supersonic jets.

Outlook

In this article, and in our research, we have emphasized the phenomenology of a set of basic physical laws operating in an astrophysical context and the dependence of this phenomenology on the underlying parameters in order to uncover trends in morphology and stability that may aid in the interpretation of extragalactic radio sources. Far from being detailed source models, our numerical simulations are rather thought-experiments carried out by a computer that show us the consequences of self-consistency in a nonlinear, multidimensional, time-dependent context. An important outcome of our investigation is the discovery of large-scale coherent structures associated with the supersonic beam, the working surface, and the cocoon and of the essential role these structures play in contributing to the phenomenological trends.

We foresee future research in the physics of extragalactic jets proceeding along two complementary tracks: 1) further phenomenological studies aimed at systematically investigating the new dynamical effects that arise when magnetic forces, bulk relativistic motion, and turbulence are incorporated into two- and three-dimensional models, as well as understanding the impact of inhomogeneities, pressure stratifications, and crosswinds in the intergalactic environment; and 2) the development of detailed source models aimed at bridging the gap between simulation and observation by incorporating self-consistent synchrotron emission models into the dynamical calculations and followed by mapping of the resultant flow field for observable quantities such as the distribution of radio brightness, degree of polarization, spectral index, and magnetic field orientation. Only through such a twin-track approach will we be able to select between competing hypotheses concerning the composition, internal processes, state of motion, and ultimately the origin of extragalactic radio sources and their associated jets, which have fascinated and continue to fascinate twentieth-century astronomers and astrophysicists. The increase in resolving power of our respective tools—radio telescopes and computers—is continuing, if not accelerating, and promises exciting research on into the twenty-first century and progress toward understanding these spectacular cosmic eruptions. ■

Further Reading

R. Courant and K. O. Friedrichs. *Supersonic Flow and Shock Waves.* New York: Interscience Publishers, Inc., 1948.

H. Reichenbach. "Contributions of Ernst Mach to Fluid Mechanics." *Annual Review of Fluid Mechanics* 15(1983):1-28.

R. Ladenburg, C. C. Van Voorhis, and J. Winckler, "Interferometric Studies of Faster Than Sound Phenomena. Part II. Analysis of Supersonic Air Jets." *Physical Review* 76(1949):662-677.

R. D. Blandford and M. J. Rees. "A 'Twin-Exhaust' Model for Double Radio Sources." *Monthly Notices of the Royal Astronomical Society* 169(1974):395-415.

Roger D. Blandford, Mitchell C. Begelman, and Martin J. Rees. "Cosmic Jets." *Scientific American,* May 1982, pp. 124-142.

Norman J. Zabusky. "Computational Synergetics." *Physics Today,* July 1984.

Michael L. Norman, Larry Smarr, Karl-Heinz A. Winkler, and Michael D. Smith. "Structure and Dynamics of Supersonic Jets." *Astronomy and Astrophysics* 113(1982):285-302.

Larry L. Smarr, Michael L. Norman, and Karl-Heinz A. Winkler. "Shocks, Interfaces, and Patterns in Supersonic Jets." In *Fronts, Interfaces, and Patterns—Proceedings of the Third Annual Conference of the Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, New Mexico, May 2-6, 1983.* Edited by Alan R. Bishop, Laurence J. Campbell, and Paul J. Channell. *Physica* 12D(1984):83-106.

Michael L. Norman, Larry Smarr, and Karl-Heinz A. Winkler. "Fluid Dynamical Mechanisms for Knots in Astrophysical Jets." In *Numerical Astrophysics.* Edited by J. Centrella, J. LeBlanc, and R. Bowers. Portola Valley, California: Jones and Bartlett, 1985.

Michael L. Norman, Karl-Heinz A. Winkler, and L. Smarr. "Knot Production and Jet Disruption via Nonlinear Kelvin-Helmholtz Pinch Instabilities." In *Physics of Energy Transport in Extragalactic Radio Sources—Proceedings of NRAO Workshop No. 9, Green Bank, West Virginia, August 1984,* pp. 150-167. Edited by Alan H. Bridle and Jean A. Eilek. Green Bank, West Virginia: National Radio Astronomy Observatory, 1985.

David G. Payne and Haldan Cohn. "The Stability of Confined Radio Jets: The Role of Reflection Modes." *Astrophysical Journal* 291(1985):655-667.

Michael L. Norman joined the Laboratory in 1984 as a staff member in the Advanced Concepts and Plasma Applications Group of the Applied Theoretical Physics Division. He received his B.S. in astronomy from the California Institute of Technology in 1975 and his M.S. in 1976 and Ph.D. in 1980 in engineering and applied sciences from the University of California, Davis. While earning his Ph.D., he performed numerical calculations of protostar formation. From 1980 until he joined the Laboratory in 1984 he was a staff member at the Max Planck Institut für Physik und Astrophysik in Munich where he did the work described in this article. Mike is also an Adjunct Assistant Professor in the Physics and Astronomy Department of the University of New Mexico and is directing the research of a number of graduate students on extragalactic jets.

Karl-Heinz A. Winkler earned his Dr. rer. nat. in physics from the Georg-August Universität Göttingen in 1976. He then became a staff member at the Max Planck Institut für Physik und Astrophysik in Munich until August 1984, where he did research on radiation hydrodynamics, star formation, and supersonic jets. In 1978 he became a consultant to Los Alamos and Livermore. Karl-Heinz joined the Laboratory as a staff member in the Applied Theoretical Physics Division during the summer of 1984. He is designing numerical methods and applying them to interesting problems in physics and astrophysics.

Genes by Mail

Participants discuss the Gene Library Project and its benefits to research on the genetic endowment of the human species.

An interview with L. Scott Cram, Larry L. Deaven, Carl E. Hildebrand, Robert K. Moyzis, and Marvin Van Dilla

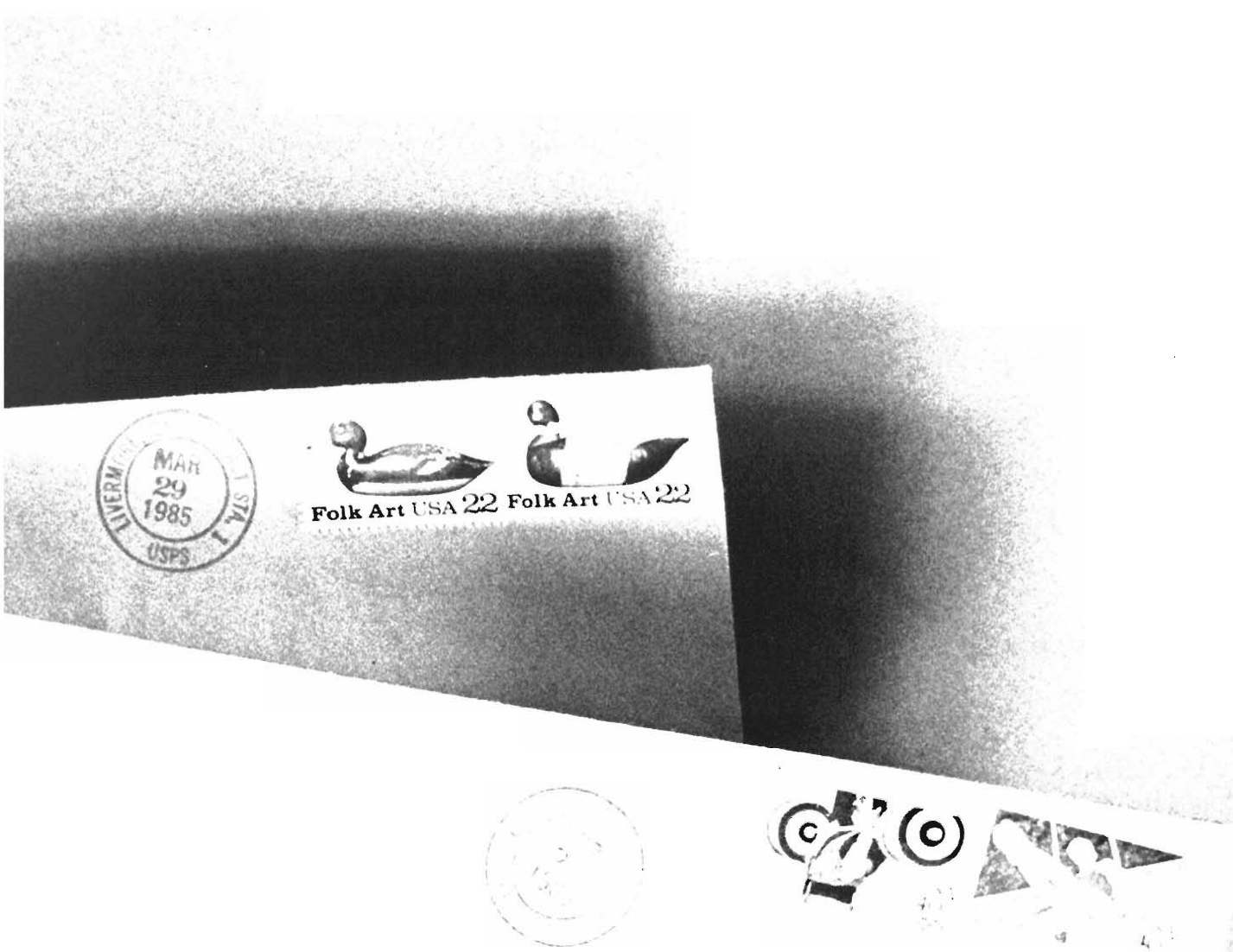
SCIENCE: *Could you start by explaining exactly what the National Laboratory Gene Library Project provides to scientists studying human chromosomes, genes, and DNA?*

VAN DILLA: We always have to start with an explanation because the name "gene library" is misleading. It conjures up a vision of a storehouse stocked with many little vials, one for each of the hundred thousand or so human genes and all neatly arranged and ready for checking out. This vision will probably be realized in the future—the distant future—but for the time being researchers must be content with something less.

What we do provide is a selection from only twenty-four vials—twenty-four libraries. Each library contains DNA specific to one of the twenty-four types of human chromosomes, the two sex chromosomes and the twenty-two pairs of autosomes. The DNA is in the form of a set of fragments cut by a restriction enzyme. So the contents of a vial are a library only in the sense of being a collection—of DNA fragments—and are in no way labeled or organized for retrieval.

DEAVEN: The fragments are not truly a "gene" library either, since each may include one gene, more than one gene, a portion of a gene, or no gene at all.

MOYZIS: The vials we mail out to users don't contain naked DNA fragments though. There's even more to making a chromosome-specific library than culturing cells, sorting the chromosomes from the cells, and letting a restriction enzyme chop up the DNA molecules from the chromosomes. A library must contain enough copies of the set of restriction fragments that the user has a reasonable amount of material to work with. One could obtain that many copies by chopping up the DNA from lots and lots of sorted chromosomes from lots and lots of cultured cells, but that's a hard and expensive way to go. Instead we put a biological amplifier to work and clone the fragments.



Dr. J. L. Hsueh
Institute of Genetics
Fudan University
Shanghai
People's Republic of China

First the human DNA fragments are each spliced into a DNA molecule of a bacterial virus—a bacteriophage, or phage for short. Then these recombinant DNA molecules are each wrapped up in the bodily trappings of the phage, and the phages are allowed to infect some unfortunate bacteria. The bacteria then proceed to work themselves to death making many duplicate viruses, each containing a human DNA fragment in a viral DNA molecule. We gather up the plaques, or colonies of viruses, suspend them in some liquid, and ship them off. So a library actually consists of human DNA fragments encased in bacteriophages.

HILDEBRAND: The phage bodies make convenient, protective packages for the fragments. And more important, if a user wants to work one at a time with members of the set of restriction fragments in a library—and that's usually the case—all he needs to do is infect some more bacteria with the phages in his library and pick off one plaque at a time.

SCIENCE: *How is such a library used?*

HILDEBRAND: It depends on what the user has in mind. In some cases the fragments may be investigated directly—their base sequences determined, for example—and the information pieced together to give some picture of the DNA molecule as a whole. You see, DNA molecules are so huge that details of their organization and functioning can, with present technology, be learned only by studying them in pieces, and preferably not pieces generated by random breaks in the molecules but pieces cut at the same points along each molecule. That's what restriction enzymes do, and it's difficult to overstate how absolutely essential they are to molecular genetics.

Getting back to the question, in most applications a library is used as a source of probes. A probe is a radioactively or fluorescently labeled DNA fragment that is used to detect and locate, in a sample of DNA being investigated, portions of DNA with a complementary base sequence. Of course, any fragment in a library can be labeled and used as a probe, but it may not be detecting anything very exciting. The challenge is to

What Is a Chromosome?

The hereditary information in each cell of an organism is awesome in extent, complex in organization, and beautiful in the dynamics of its expression. Chromosomes are the packages housing this information. Most of us think of chromosomes as the *X*-shaped objects viewed through a microscope in a high-school biology course. This familiar manifestation of chromosomes exists, however, only during a certain brief stage in the life of a cell. Since chromosomes—in particular, human chromosomes—are the essential raw material for the Gene Library Project, a brief review of what is now known about these carriers of heredity is worthwhile.

Located in the nucleus, each chromosome includes, as a single molecule, some fraction of the cell's complement of DNA and hence of the cell's genes. The number of chromosomes in a normal cell depends on the species of origin of the cell. Although all organisms of a species have the same chromosome number, that number is not unique to the species. For example, the evening bat (*Nycticeius humeralis*), the red squirrel (*Tamiasciurus hudsonicus streatori*), and the human all have the same chromosome number. The dependence of chromosome number on species is not straightforward but exhibits a general downward trend with progression up the evolutionary ladder.

In higher organisms another variation in chromosome number occurs: all gametes (cells that participate in sexual reproduction, such as egg and sperm cells) have the same "haploid" number of chromosomes, and all other cells are "diploid," with exactly twice that number. Human gametes, for example, possess twenty-three chromosomes, and all other human cells possess forty-six.

How is the chromosome number of a species maintained constant from one cellular and organismal generation to another? In the case of humans, a new organism begins with the formation of an ancestral diploid cell by fusion of two haploid gametes: an egg carrying an X sex chromosome and a sperm carrying an X or a Y sex chromosome. (The new organism is a female (male) if an X-bearing (Y-bearing) sperm participated in the fusion.) Other cells are then formed by successive waves of mitosis, a type of cell division resulting in two diploid cells, each containing duplicates of the chromosomes of both the maternal and the paternal gametes. (The maternal and paternal versions of each chromosome are referred to jointly as a homologous pair.) The organism grows and develops as mitosis continues and certain of the cells become differentiated in function. At some stage or stages in the organism's life, special diploid cells (germ cells) undergo meiosis, a type of cell division resulting in formation of four haploid gametes. Each chromosome in these gametes is a duplicate of either the

maternal or the paternal chromosome or, more usual, is some combination of the two arrived at by exchange of hereditary material between homologous chromosomes during meiosis. The cycle begins again with fusion of one such gamete with another from an organism of the opposite sex.

A chromosome includes, in addition to a single DNA molecule, a mass of special proteins (histones) roughly equal to that of the DNA molecule (some 15 percent of the total chromosomal mass), small amounts of many other DNA-binding proteins, and RNA. The histones make the DNA molecule more flexible, by neutralizing the negative charges of its phosphate groups, and thus play an essential role in packaging the very long DNA molecules (which, for humans, average 5 centimeters in length) within the confines of the cell nucleus (which is between 5×10^{-4} and 10^{-3} centimeter in diameter). Many details of this remarkable engineering feat are as yet unknown, but several levels of packaging are involved (Fig. 1).

The packaging begins with the formation of "nucleosomes," each consisting of a length of linking DNA and a length of DNA wound around a core of histones. Next, some regular packing of the nucleosomes, also involving histones, leads to the structure known as the 30-nanometer chromatin fiber (the double helical strand of DNA itself is about 2 nanometers in diameter). This fiber in turn is organized into a sequence of "looped domains," each of which is thought to be activated as a unit during gene expression. The sequence of looped domains is further condensed in some fashion to form the "interphase" chromosome, portions of which decondense and recondense as various genes are activated and deactivated during biosynthesis by the cell.

The ultimate in DNA packaging occurs as a germ cell prepares for meiosis or as a somatic cell (any cell other than a germ cell) prepares for mitosis. Then each interphase chromosome is duplicated, and the two identical "sister chromatids," which are joined at some point along their lengths (the "centromere"), form a longitudinally sym-

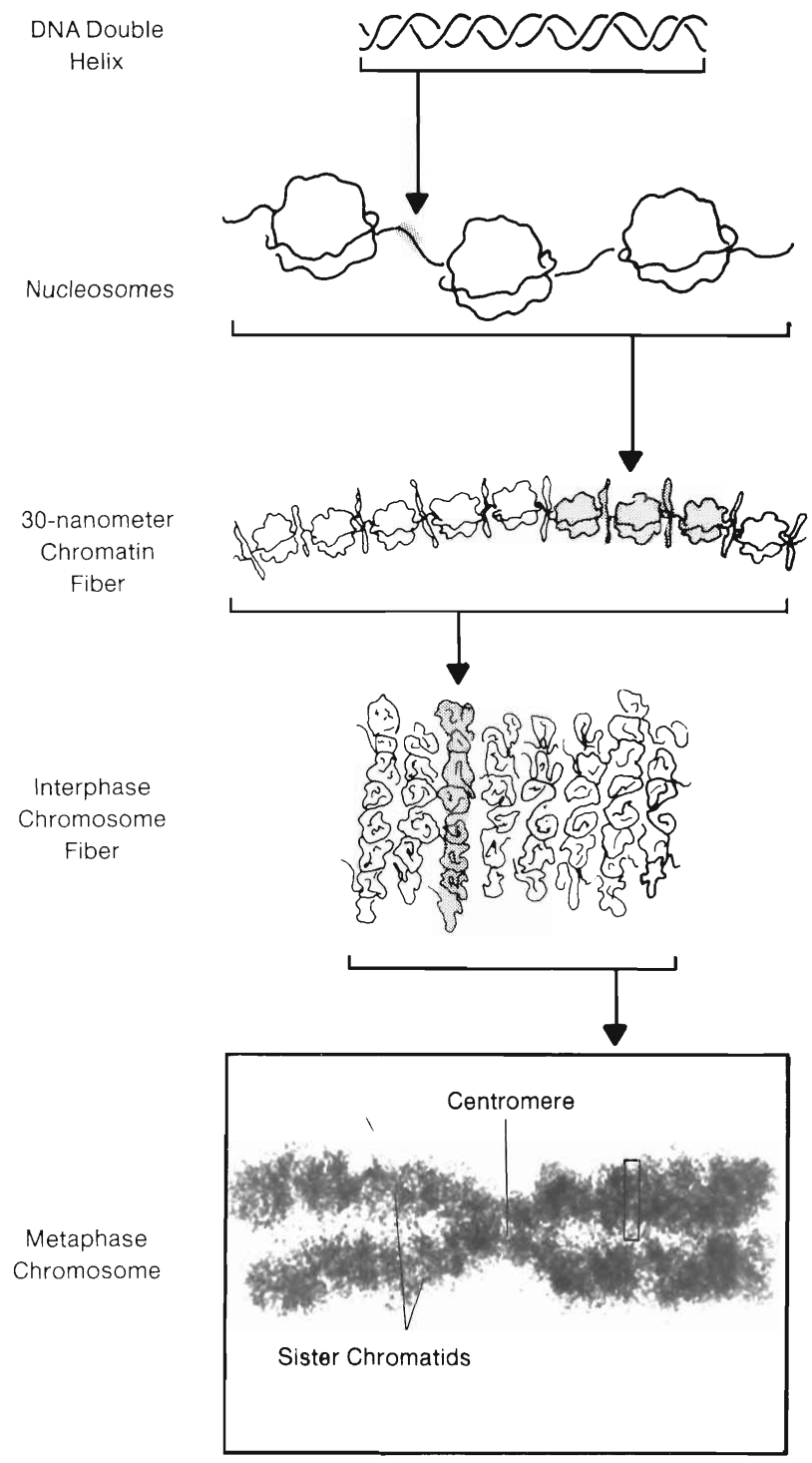


Fig. 1. Many details of the packaging of a DNA molecule and various proteins into the entity known as a chromosome are yet to be determined. These schematic illustrations are merely suggestive of the mechanisms and configurations that may be involved.

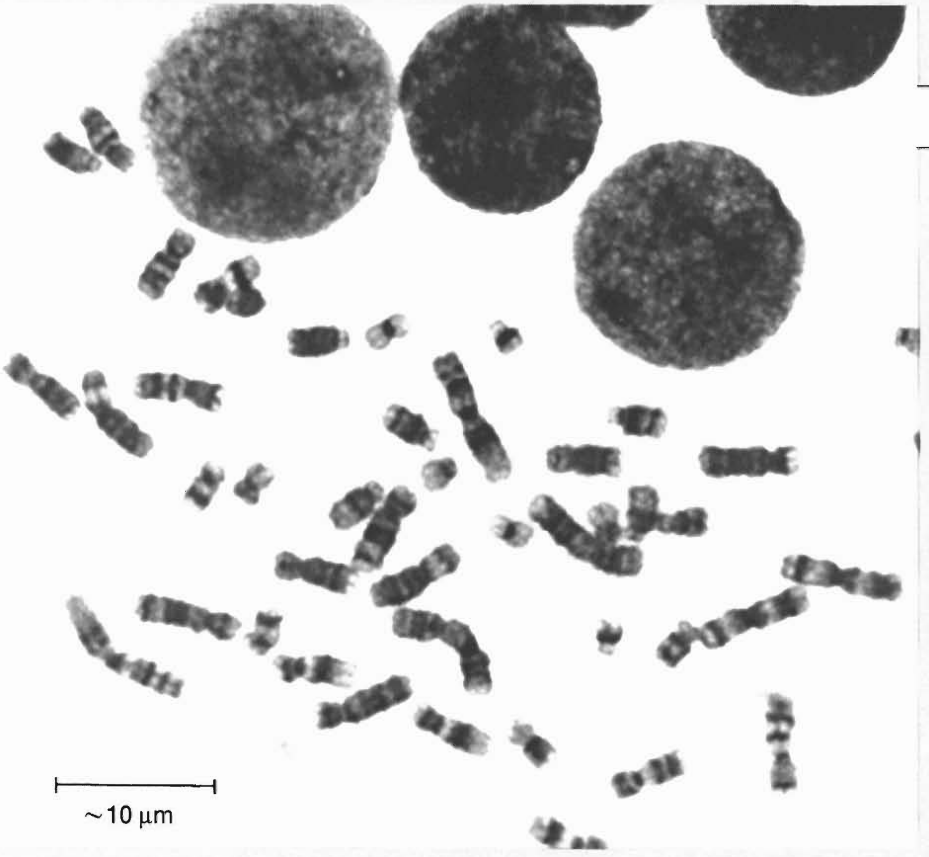


Fig. 2. Photomicrograph of metaphase chromosomes from human leukocytes (white blood cells). The cells were cultured for 48 hours, blocked at metaphase by exposure to the drug Colcemid for 2 hours, swollen in a hypotonic solution, and treated with a fixative. The chromosomes were released from the metaphase cells simply by dropping the swollen cells onto microscope slides, which breaks the cellular membranes. (The dark, round objects are nuclei released from interphase cells.) The chromosomes were then partially digested with trypsin and treated with Giemsa stain. Since trypsin preferentially digests chromosomal proteins bound to DNA rich in guanine-cytosine pairs, those portions of a chromosome containing such DNA are less readily stained. The result is a pattern of heavily stained (dark) and lightly stained bands along the length of each chromosome. The patterns of these bands permit unambiguous identification of chromosomes.

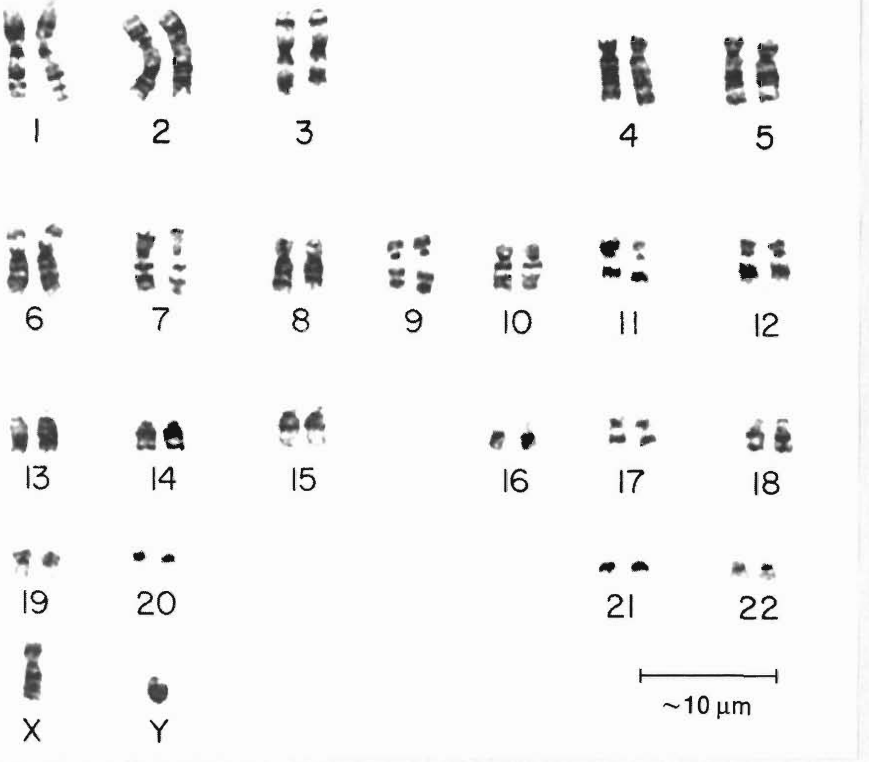


Fig. 3. A display of the metaphase chromosomes of an organism, banded as in Fig. 2 and arranged in order of decreasing size, is known as the karyotype of the organism. Shown here is the karyotype of a human male. By convention the pair of sex chromosomes is displayed separately from the homologous pairs of autosomes. Chromosomes are designated by a size-related number and, historically, by a morphology-related group: group A includes numbers 1 through 3; group B, numbers 4 and 5; group C, numbers 6 through 12; group D, numbers 13 through 15; group E, numbers 16 through 18; group F, numbers 19 and 20; group G, numbers 21 and 22. The similar morphology of the members of a group made them extremely difficult to distinguish before discovery in 1970 of the banding method described in Fig. 2.

WHAT IS A CHROMOSOME?

metric double structure with the familiar *X* shape. These "metaphase," or "mitotic," chromosomes are easily visible through an optical microscope (Fig. 2) and were first observed by German cell biologists in the early 1880s.

The metaphase chromosomes of a given species can be distinguished on the basis of a number of properties, such as overall length, DNA content, or location of the centromere. (Except in the case of an X-Y pair, the differences between the two members of a homologous pair of chromosomes are much less pronounced than the differences among homologous pairs.) The most definitive identifying property is the pattern of bands produced on metaphase chromosomes by selective digestion of the chromosomal proteins and subsequent staining. Figure 3 displays such banding patterns for the human chromosomes.

Which genes are packaged in which chromosomes, and exactly where, is known only for some 700 of the estimated 100,000 human genes. Examples of those for which this information is known include the gene for Huntingtons's chorea (chromosome 4), for interferon (chromosome 9), for β-globin (chromosome 11), and for hemophilia and red-green colorblindness (the X chromosome).

The marvelous choreographies of mitosis and meiosis normally transmit intact and unchanged the time-tested chromosomes of an organism from one cellular and organismal generation to another. But sometimes errors, such as rearrangements of chromosomes or departures from the normal number, do occur. The consequences of errors in meiosis are usually grave, if not fatal, to the organism that inherits them. For example, an abnormal number of chromosomes are found in the cells of about half of the human embryos that are spontaneously aborted, and the presence of three copies of human chromosome 21 results in Down's syndrome. Errors in mitosis also have serious consequences, as evidenced by the association of chromosome rearrangements with various types of cancer. ■

find the interesting probes, ones that reveal something unusual about the DNA, such as an alteration related to a genetic disease or a mutation.

SCIENCE: *Have libraries been produced before?*

CRAM: Whole-genome libraries, ones containing DNA fragments from all the human chromosomes, are already available, and so are a few chromosome-specific libraries, ones for the chromosomes that are easier to sort. But this project will be the very first to provide chromosome-specific libraries for all twenty-four types of human chromosomes.

VAN DILLA: With a chromosome-specific library a user can begin his research by fishing around among a pool of DNA fragments from a particular chromosome rather than the whole sea of human DNA. This is a tremendous advantage, and the libraries should increase the productivity of research in molecular genetics just as separated isotopes increase the productivity of research in nuclear physics.

DEAVEN: I have some data to support that expectation, at least as far as one application of the libraries—gene mapping—is concerned. By gene mapping is meant localizing a gene on a particular region of a particular chromosome. In 1958, when molecular genetics was still in its infancy, only about four hundred human genes had been identified. These had been classified as dominant or recessive and autosomal or X-linked but, except for the X-linked genes, could not be mapped. That had to await, first of all, the discovery some ten years later of chromosome banding, a technique for unambiguously distinguishing all twenty-four types of human chromosomes. This discovery made it possible to map a limited number of genes by associating chromosome abnormalities and genetic diseases. At about the same time human-rodent hybrid cell lines became available, and information about the proteins synthesized by various of these cell lines, each carrying only a few human chromosomes, made it possible to map more genes. Still, by 1973 only 64 genes had been assigned to particular autosomes and 155 to the X chromosome. By 1981 the total number of mapped genes had increased to only about four hundred and fifty, but in that year molecular genetics provided a powerful new tool for mapping. It was discovered that what are known as restriction fragment length polymorphisms could pinpoint the locations of genes about which practically nothing is known except the observable evidence of their expression, such as the symptoms of diseases.

A restriction enzyme cuts a DNA molecule into fragments of various lengths by breaking bonds at every occurrence of a cleavage site, a certain sequence of bases. Mutations can generate new cleavage sites and destroy existing ones, and these changes cause differences in the sets of restriction fragments from different individuals. For example, in the fragments produced by the restriction enzyme *Hpa*I from the DNA of American Blacks, the β-globin gene usually appears on a 7.6-kilobase fragment but sometimes on a 7.0- or 13.0-kilobase fragment. Differences like these are called restriction fragment length polymorphisms, or RFLPs. An RFLP can be detected by using electrophoresis to separate, according to length, the restriction fragments from, say, two individuals and seeing whether a probe that includes the RFLP sticks to different places in the two lineups.

An RFLP within a gene is obviously inherited along with the gene, but so also is an RFLP so close to a gene that it is not separated from it during meiotic recombination. So the location of such an RFLP on a chromosome is a very close approximation to that of the gene. This method of mapping, say for a gene that causes some disease, involves testing many probes to find one that reveals an RFLP unique to people with and without the disease—that's the hard part—and then mapping the probe on the chromosome—that's the easy part.

Since 1983 about three hundred RFLP probes for genes have been identified and mapped per year. Even that rate, though, is painfully slow considering how many more genes remain to be mapped. But our chromo-

some-specific libraries will make the search for probes less difficult, and we estimate that about two thousand RFLP probes for genes will be identified and mapped in the first year the libraries are available, and even more in succeeding years.

HILDEBRAND: Once a probe for, say, a defective gene has been identified, it has applications in addition to gene mapping. It can be used to detect the gene in unborn children and in unknowing carriers of the gene—unknowing because the gene may be recessive, not clearly expressed, or, like the gene for Huntington's chorea, expressed only late in life. People from families with histories of genetic disease usually want such information to help them make decisions about the course of pregnancies or of their own lives.

RFLP probes for genes are also very valuable to basic research in molecular genetics. For instance, by studying the inheritance patterns of probes for various genes, we can learn about the relative locations of the genes along the DNA molecules.

SCIENCE: *What is the history of the project?*

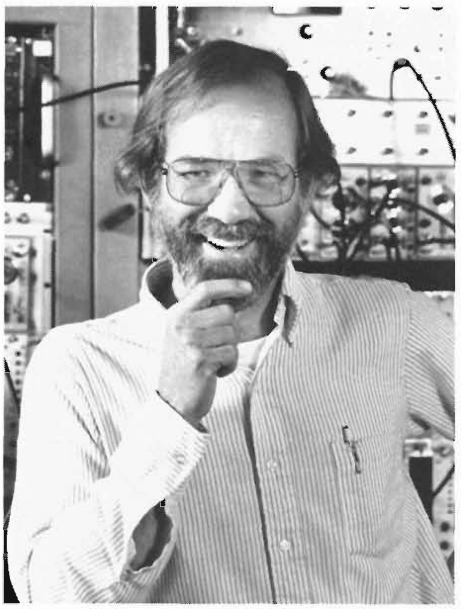
DEAVEN: About three years ago a group of us in the Life Sciences Division at Los Alamos decided to look into the possibility of establishing a new program oriented around the structure and functioning of mammalian chromosomes. We envisioned a small group of related projects to be supported by some federal agency. Among the ideas we came up with, that of producing DNA libraries for each of the human chromosome types seemed particularly worthwhile and, furthermore, related to two established strengths of the Laboratory—GenBank, which is a repository for DNA sequence data, and the National Flow Cytometry Resource, which organizes our capability in the technology required for sorting chromosomes.

The DOE's reaction to our idea was quite favorable. They thought that the libraries would provide the means for new research capabilities at the national laboratories and elsewhere and that production of the libraries would be a unique contribution to the biomedical community. Of course, one ultimate payoff of the project will probably be a better method of detecting mutations in humans caused by exposure to radiation or other mutagens, and this is a basic interest of the DOE's Office of Health and Environmental Research. In any case, the DOE decided to fund the project as a service to the biological and medical communities. In discussions with personnel at the DOE, the suggestion was made that we initiate the project jointly with a group of biomedical scientists at Livermore, which also was strong in flow cytometry.

People at Livermore were enthusiastic, and together we began by meeting, in October and November of 1983, with a group of molecular biologists and human geneticists who advised us about what was needed and the relative priorities to assign to those needs. This advisory group suggested that we construct, as quickly as possible, a set of libraries to be used as sources of potential probes for gene mapping and diagnosis of genetic diseases. For those applications small DNA fragments are best, since a small fragment that includes a sequence coding for a gene is not likely to include in addition one of the many noncoding repetitive sequences present in DNA. So that's what we're doing now—making libraries of relatively small fragments. When these Phase I libraries are completed, we'll go on to constructing the Phase II libraries, with larger human DNA inserts, to be used for basic research in gene structure, arrangement, and expression.

This is a good opportunity for me to acknowledge the foresight of our technical representatives at the DOE. Usually, the most important products of research are

L. Scott Cram of Los Alamos, a biophysicist, is a specialist in the application of flow cytometry to identification and sorting of chromosomes. His research interests include the role of karyotype instability in tumorigenesis.



Larry L. Deaven, a cytogeneticist, is principal investigator for the Gene Library Project at Los Alamos. His research interests include the chromosome constitution of evolving cell populations and the chromosome damage induced by agents associated with energy production.

Carl E. (Ed) Hildebrand of Los Alamos, a biophysicist turned molecular biologist, is a specialist in recombinant DNA technology. His research interests include DNA replication, chromatin structure, and gene expression.

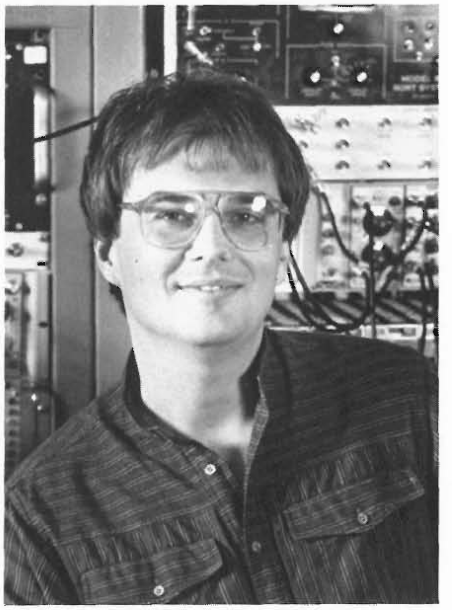
publications, and to some extent research is funded on the basis of the expected number and quality of the publications. But for this project the DOE set that criterion aside in favor of a different goal—the production of libraries. Although we expect important publications from our work on the project, our major aim is to produce high-quality libraries for use by interested scientists throughout the world.

In addition, the Gene Library Project wouldn't be as far along as it is if it hadn't been funded by the DOE, which can act very quickly when something regarded as worth supporting comes along. We have now completed approximately 75 percent of the Phase I libraries. If our technical representatives at the DOE had not understood the significance and timeliness of the project, we might well still be in the organizational stages.

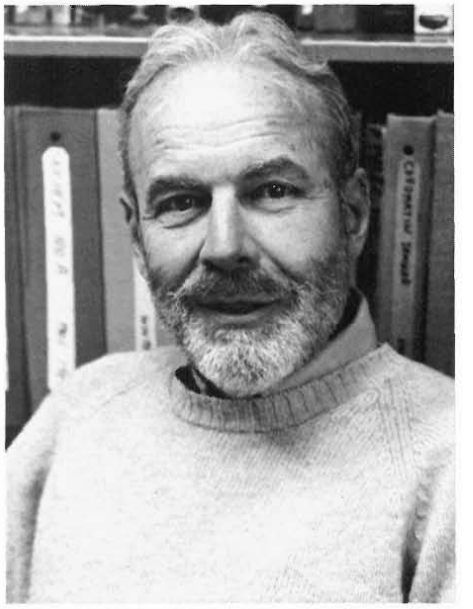
SCIENCE: *How does it happen that two weapons laboratories have the personnel and facilities required for such a project?*

DEAVEN: That's not so much of an anomaly as you might think. The AEC had a natural interest in the biological effects of radiation and realized, as have its successors, that understanding those effects in detail had to be based on knowledge of cellular processes at a molecular level. So in addition to programmatic research on the biological effects of radiation—and now of various chemicals associated with other means of producing energy—basic research in cellular biology has received strong support at the national laboratories since their beginnings.

SCIENCE: *What kind of collaboration between the two laboratories is involved, and how well is it working?*



Robert K. Moyzis, a molecular biologist, is Leader of the Genetics Group at Los Alamos. His research interests include the organization of human DNA and the regulation of gene expression.



Marvin Van Dilla, a biophysicist, is principal investigator for the Gene Library Project at Livermore. His research interests include the development of flow instrumentation and its application to biological problems.

VAN DILLA: We collaborate in the sense of drawing on each other's expertise rather than in the sense of parceling out various aspects of the project to one or the other laboratory. Both of us have gone through the whole process of making Phase I libraries, but our products are not identical—by edict rather than by chance. The Livermore libraries contain fragments of DNA cut by one restriction enzyme, and the Los Alamos libraries contain fragments cut by a different restriction enzyme. So a fragment missing from one version, because the recombinant DNA molecule containing that fragment is too large to fit into the bacteriophage, will quite likely be found in the other.

In my judgment the collaboration has been very effective, with each laboratory contributing particular strengths. For example, we at Livermore had greater experience in

sorting human chromosomes with flow cytometers and had been developing one capable of doing the job at a rate compatible with the scope of the project. We need something on the order of a million sorted chromosomes for each library. Commercially available flow cytometers can sort that many in a few tens of hours—and that's operating time, so the job actually takes much longer—but the high-speed sorter reduces the sorting time to just a few hours. We made copies of the basic components of our sorter and sent them here. I understand the Los Alamos machine is now up and running.

As for Los Alamos, they brought to the project greater experience and expertise in recombinant DNA technology and in fact had made a few chromosome-specific libraries for the Chinese hamster before the project even started. The Los Alamos group had also explored the use of hybrid cells as sources of human chromosomes.

Our first joint meeting, at Los Alamos in December of '84, was proof to me of a successful collaboration. Almost everyone involved in the project at Livermore, from Ph.D.'s to technicians, attended the meeting and took the opportunity to discuss common problems and exchange tricks of the trade with their counterparts at Los Alamos. Our chief cell farmer, for instance, had never been to Los Alamos to talk to the chief cell farmer here. They both do the same thing with somewhat different techniques and have developed somewhat different ways of getting around the many nitty-gritty problems that crop up in the busines of cell culture. Together they did two experiments during the meeting. One day they grew cells and isolated the chromosomes using the technique that's favored at Livermore, and on the next day they went through the same steps using the technique favored at Los Alamos. Each learned something from the other and seemed very happy about the whole experience.

SCIENCE: *How much interest has been shown in the libraries?*

Making the Libraries

Supplying the Chromosomes

As raw material for each chromosome-specific library, we must collect between half a million and a few million of a particular chromosome by sorting through tens to hundreds of millions with a flow cytometer. Our original hope was to obtain such large numbers of chromosomes for sorting by culturing, or multiplying *in vitro*, human cells, namely, fibroblasts from human foreskin tissue. But, as discussed below, these cells proved to be an acceptable source for only the smaller chromosomes, and another source had to be found for the larger.

We culture fibroblasts in three stages, beginning by placing small pieces of the tissue sample, together with nutrients essential for growth, in a small plastic container. (Despite the translation of *in vitro*, plastic has largely supplanted glass as a surface for cell culture.) The fibroblasts migrate from the tissue, adhere to the bottom of the container, and multiply by mitosis. (This process is similar to that by which fibroblasts *in vivo* repair a break in the skin.) Mitosis stops when the bottom of the container is covered with a monolayer of cells (Fig. 1). The monolayer is then partially digested with the enzyme trypsin to yield a suspension of single cells. This suspension is divided among several containers, fresh growth medium is added, and the culturing is repeated, again until a monolayer has formed.

Our aim is to maximize the number of fibroblasts at metaphase (on the verge of dividing), since only those cells contain the metaphase chromosomes that can be sorted. Obviously, the more rapidly a population of cells is multiplying, the larger is the fraction of that population at metaphase. But the fibroblasts begin to multiply less rapidly after about fifteen to twenty mitoses and cease multiplying altogether after about fifty. Therefore, after the second culture some of the fibroblasts are suspended in a medium containing glycerol (or some other ice-crystal inhibitor) and frozen in liquid nitrogen. These cells, which can be thawed and recultured at any time up to several years later, serve as a reservoir of fibroblasts at the height

I. SUPPLYING THE CHROMOSOMES

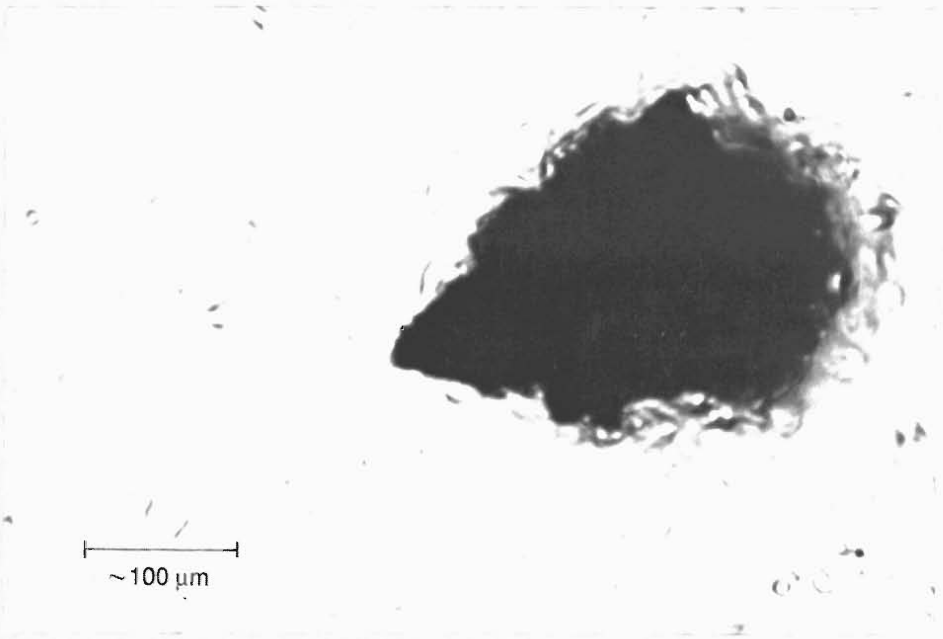


Fig. 1. Cells in culture multiply by mitosis and spread over the surface of the culture dish until a confluent monolayer of cells is formed. Shown here is an optical micrograph of a confluent monolayer of fibroblasts cultured from a piece of human skin tissue (dark area).

Fig. 2. Cells can multiply manyfold in large, slowly rotating cylinders ("roller bottles") since a large surface area is kept in contact with the culture medium.

of their multiplication rate.

The final culture of the remaining fibroblasts takes place in large "roller bottles" (Fig. 2). When a monolayer has almost formed, the drug Colcemid is added to arrest the cell cycle at metaphase. After exposure to this drug for an appropriate time, the fraction of cells at metaphase increases from the usual maximum of 0.5 percent to between 30 and 50 percent. Since the metaphase fibroblasts "round up" from the culture surface because of their more spherical shape, they are easily separated by gentle shaking of the roller bottles.

The next step is to isolate the chromosomes from the cells. first, the cells are swollen in an "isolation buffer" with an osmotic pressure lower than that of physiological saline. The buffer is then forced through a fine-gauge hypodermic needle to break the cell membranes (the nuclear membranes are already dissipated in metaphase cells) and release the chromosomes. The buffer solution also contains materials, such as proteins that coat chromosomes or dyes that intercalate with DNA, to stabilize the morphology of the chromosomes. Although this isolation step sounds simple in principle, difficulties that plague even the best methods are responsible for our yield of only 2 to 6 sorted chromosomes from every 100 cells.

The final step in preparing the chromosomes for sorting is addition to the buffer of two fluorescent dyes. One of these binds preferentially to DNA sequences rich in adenine-thymine base pairs and the other to sequences rich in guanine-cytosine base pairs. The fluorescence intensities excited in the dyes serve as the basis for flow-cytometric analysis and sorting of the chromosomes.

Analysis of chromosome preparations from human fibroblasts indicated that all the chromosomes except numbers 9 through 12 were sufficiently well resolved for sorting (see Fig. 2 in "Sorting the Chromosomes"). In practice, however, sorts for the larger chromosomes were very time-consuming, apparently because the shearing action that breaks the cell membranes also breaks up many of the larger chromosomes and greatly reduces their numbers. More important, the purities of these sorts were reduced to less than the acceptable level of 90 percent by many doublets of smaller chromosomes. Evidently the shearing action is not sufficient to break up these doublets, which form during isolation and cannot be distinguished by the flow cytometer from the larger chromosomes. Furthermore, reducing the shearing force to avoid breaking the larger chromosomes only increased the number of doublets.

As a result, we have had to turn to human-hamster hybrid cells as a source of the larger chromosomes. Such cells result from fusion of human and hamster cells and initially contain a complete set of both human and hamster chromosomes. However, after several weeks to several months in culture, the hybrid cells spontaneously and randomly lose some of their complement of human chromosomes. These losses make it possible to clone hybrid cells carrying selected human chromosomes. For our purpose an ideal hybrid cell is one with no more than three human chromosomes, each separable in a flow cytometer from each other and from the hamster chromosomes.

Since hybrid cells with a small number of human chromosomes are useful in other areas of research, various cell lines have been established and frozen at a number of laboratories. We are indeed grateful to those who have shared their cell lines with us for the Gene Library Project.

Culture of the hybrid cells is similar to, and no more difficult than, that of human fibroblasts. However, as sources of chromosomes for the libraries, hybrid cells are far from perfect. Their habit of unexpectedly losing human chromosomes, for instance, requires frequent monitoring for the continued presence of the desired chromosomes. Even more troubling is the fairly common occurrence of DNA exchange between the human and the hamster chromosomes. Such exchanges are difficult to detect but, undetected, contaminate the libraries with hamster genes. ■

DEAVEN: Well, we recently filled requests for over two hundred and fifty Phase I libraries from users in the United States and seven other countries. We expect many more requests as word of their availability gets around.

MOYZIS: I would be very surprised if every research group in the world working on human genes wouldn't want at least one of the libraries, expecially after the Phase II libraries are available.

SCIENCE: *Do you exercise any control over the uses to which the libraries are put?*

DEAVEN: No. The request form for a library includes a question about the nature of the proposed research, but we ask that primarily because we want to keep a record of the various applications. Users funded by government agencies have already agreed to the guidelines established by the National Institutes of Health for recombinant DNA research, and industry voluntarily abides by those guidelines.

Incidentally, the plan for the future is that the NIH will establish a repository for the libraries we produce and will handle the distribution. They will also collect information determined by the users about the libraries, such as purity and completeness data and characteristics of probes isolated from the libraries. This information will serve as feedback to us for improving future libraries and will of course be valuable to other users. NIH also plans to establish a repository for probes pulled from our libraries and others.

MOYZIS: One thing we do request of users is that they don't pass the libraries on to other investigators. We want the libraries to originate directly from us, or from the repository when that comes about. One reason for this is so that we can keep in touch with all the users. A more important reason for not wanting the libraries passed around is to preserve their characteristics. Each amplification of a library unavoidably introduces changes. For example, since phages carrying human DNA fragment A do not multiply at exactly the same rate as those carrying fragment B, after several amplifications the rela-

tive numbers of fragments A and B will be quite different. Some fragments may disappear altogether.

Another kind of change that occurs when a library is amplified is rearrangement of the fragments. One can start, say, with a fragment containing genes A, B, and C in that order and, because cloning is something more than stamping out pennies, end up with the genes in the order C, A, and B, which is misrepresentative of the chromosome. It was to minimize this kind of change that we chose the bacteriophage Charon 21A as the cloning vector for the Phase I libraries, since in this phage the tendency for the human DNA insert to get rearranged is minimal. When we move on to the Phase II libraries and start using other cloning vectors, we'll have to worry more about this problem.

SCIENCE: *What are some applications of the libraries?*

VAN DILLA: We've already discussed gene mapping and mentioned diagnosis of genetic diseases, detection of mutations, and basic research in molecular genetics. Some other interesting applications are comparisons of DNA sequence organization among the various human chromosomes, pedigree analysis, and comparison of human DNA with those of other species.

MOYZIS: Before we go on to other applications, I'd like to point out that gene mapping is more than just collecting dry facts—it provides basic clues about possible mechanisms for controlling gene expression. For example, some thousands of genes are expressed only in nerve cells. How can that be? If we find that these genes are scattered about all over the genome, we may conclude that some structural feature near each gene regulates its expression. But if we should find that all are located in the same region of just a few chromosomes, then a very different, global regulation mechanism may be at work.

There's also a wealth of data relating various pathological conditions, including cancer, to chromosomal abnormalities such as translocations and deletions. Why is this

Making the Libraries

Sorting the Chromosomes

Before the advent of flow cytometry, human metaphase chromosomes could be coarsely partitioned into at most ten fractions by centrifugation or sedimentation techniques. Now, a state-of-the-art flow cytometer can separate these chromosomes, with an accuracy of up to 95 percent and at a rate of up to 400 per second, into twenty-four fractions, one for each of the two sex chromosomes and the twenty-two homologous pairs of autosomes. Originally developed at Los Alamos in the late 1960s to measure the DNA content of cells,* the flow cytometer is absolutely essential to our goal of producing very pure and highly representative chromosome-specific libraries of DNA fragments.

The necessary accuracy of sorting, and hence purity of the libraries, is achieved with a dual-beam flow cytometer (Fig. 1). This instrument "recognizes" a particular human chromosome on the basis of a property of the DNA it contains, namely, the ratio of adenine-thymine to guanine-cytosine base pairs. This ratio, which varies from one chromosome to another, is translated into measurable fluorescence intensities by staining the chromosomes with two fluorescent dyes. One (Hoechst 33258) binds with high specificity to sequences rich in A-T base pairs, and the other (chromomycin A_3) to sequences rich in G-C base pairs. As the stained chromosomes intercept the two laser beams of the cytometer, the fluorescence intensities excited in both dyes are measured and used to trigger sorting if they fall within a certain range of the intensities characteristic of the desired chromosome. The data collected during this "bivariate" analysis of a chromosome preparation are monitored in real time and displayed as a "flow histogram" (Fig. 2). Such a histogram is indicative

**See "Flow Cytometry: A New Tool for Quantitative Cell Biology," Los Alamos Science, Volume 1, Number 1, Summer 1980.*

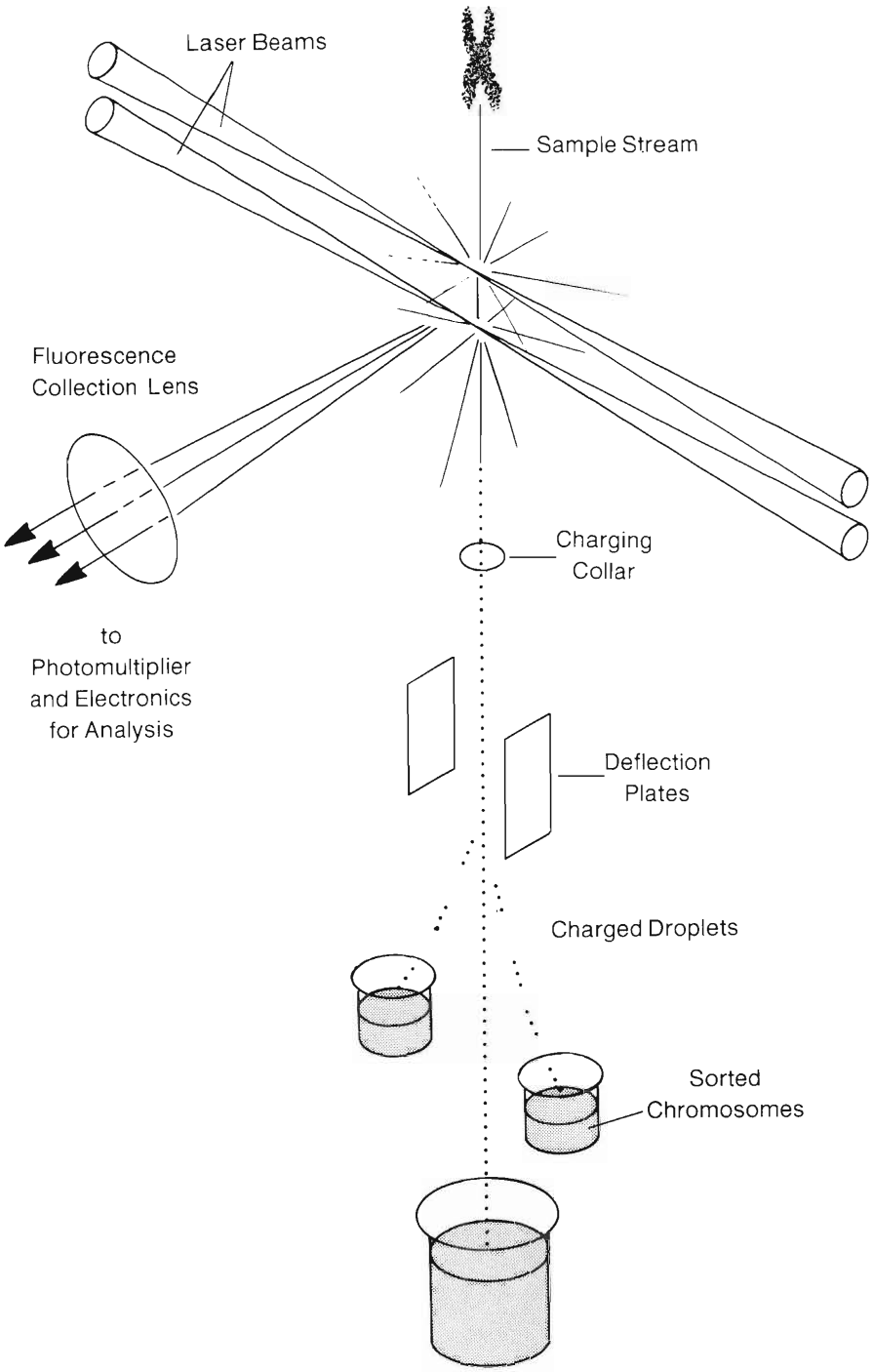


Fig. 1. Operating principles of a flow cytometer for sorting human chromosomes on the basis of the fluorescence intensities of two dyes, each of which binds preferentially to a different type of DNA sequence. A very narrow stream of the chromosome preparation flows rapidly through the focused regions of two lasers, each tuned to the excitation wavelength of one of the dyes, and is then broken into a rapid succession of tiny droplets. If the measured fluorescence intensities fall within a prescribed range of those characteristic of the desired chromosome, an electric charge is imparted to a group of droplets certain to include the chromosome. These droplets are then deflected from the main droplet stream into a collection receptacle. Two different chromosomes can be sorted from the same preparation since either a negative or a positive charge can be imparted to the droplets.

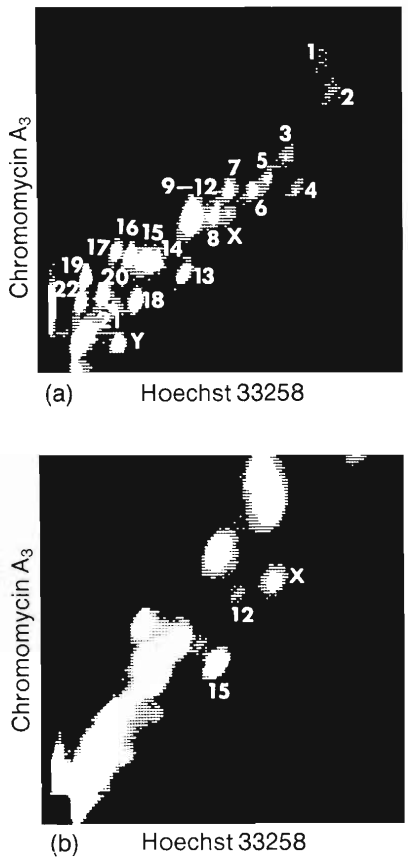


Fig. 2. (a) A flow histogram obtained from analysis of a typical chromosome preparation from human fibroblasts. The histogram is a record of the number of fluorescence events (each presumably due to a single chromosome) versus the fluorescence intensities of the dyes chromomycin A_3 and Hoechst 33258. Peaks in the histogram are attributed to different chromosomes as shown. This histogram indicates that chromosomes 9 through 12 cannot be sorted with the necessary accuracy from such a mixture of all the human chromosomes. In practice, neither can chromosomes 1 through 8, X, 14, and 15 (see "Supplying the Chromosomes"). For these chromosomes the necessary accuracy is achieved by sorting from the chromosomes in human-hamster hybrid cells rather than those in human fibroblasts. (b) A flow histogram of a chromosome preparation from human-hamster hybrid cells carrying human chromosomes X, 12, and 15. Note the excellent resolution of the human chromosomes. Hybrid cells carrying various human chromosomes are used as sources for those chromosomes not sortable at an acceptable purity from fibroblast preparations. Note that the hamster chromosomes lie on a straight line in the histogram; this is evidence of a nearly constant ratio of adenine-thymine base pairs to guanine-cytosine base pairs in the DNA molecules of these chromosomes.

II. SORTING

of the accuracy with which the chromosomes in the preparation can be distinguished.

One problem we faced in sorting the chromosomes was finding a suitable stabilizer, a material added to the isolation buffer to stabilize the morphology of the metaphase chromosomes. Some of these materials appeared to decrease the specificity of the dyes (thus decreasing the sorting accuracy) and to degrade the chromosomal DNA (thus producing fragments that cannot be incorporated in recombinant DNA molecules). These undesirable side effects have now been minimized by using the polyamines spermine and spermidine as stabilizers.

The speed of sorting is crucial to producing highly representative libraries, which requires cloning DNA fragments from very large numbers of sorted chromosomes. For example, about two million were needed for each library produced during the first phase of the project. (These Phase I libraries contain relatively small DNA fragments.) With the commercial flow cytometer then available, sorting that many chromosomes took about twenty hours, a not unreasonable time. But ten times more sorted chromosomes will be needed for each of the Phase II libraries (which will contain relatively large DNA fragments), and two hundred hours is indeed an unreasonable time. With this in mind, our collaborators at Livermore developed the high-speed sorter now in use at both laboratories. Its approximately tenfold greater sorting speed is due to more rapid formation of droplets, a modification that permits sorting from more concentrated chromosome preparations. We are also investigating the feasibility of simultaneously sorting four, rather than two, different chromosomes by charging the selected droplets to one of two levels of either polarity.

The successful application of flow cytometry to sorting chromosomes is evidence of the many advances made in its practice. But the technology still has its drawbacks. It is expensive and, more important, very labor-intensive. Computerized monitoring and control may help reduce the human labor involved. ■

so? A good first question to answer is what genes are located near those abnormalities.

HILDEBRAND: The chromosome translocation associated with Burkitt's lymphoma, a cancer of the immune system, is known to put a cancer-causing gene near a regulatory sequence for an immunoglobulin gene. But we don't know for sure whether this observation has general significance. For example, are there many such rearrangements that lead to pathological states?

DEAVEN: The same argument holds for mapping close to the translocations known to be involved in many hereditary diseases. If we had documentation of what genes were located on the chromosomes that participate in a translocation, then we would have insight into the mechanisms at work in expression of the disease.

MOYZIS: It wasn't too many years ago that the immense amount of DNA in a human cell was regarded as having been just thrown into the nucleus like a mess of spaghetti. Now more and more evidence is accumulating that this mess has a very defined order and that any perturbation of that order results in some alteration in the individual. It's now known, for instance, that approximately one-half of all pregnancies end in spontaneous abortions at a very early stage of embryonic development, and the chromosomes of about one-half of all spontaneously aborted embryos are badly scrambled. So the genetic diseases that we know about are only a very small fraction of what can possibly go wrong. Most alterations in embryonic DNA lead to embryonic death.

DEAVEN: Between 2 and 3 percent of all newborns are afflicted with one of the more than three thousand known genetic diseases. And these numbers don't include some relatively common adult disorders, such as Alzheimer's disease, that have a genetic component. So genetic disease is a major biomedical and social problem. I think that the libraries will accelerate the identification of probes for rapid, inexpensive prenatal diagnosis and genetic counseling. Biotech companies should find this a profitable field—finding and marketing batteries of such probes.

The ultimate goal, of course, is not simply to predict the occurrence of a genetic disease but to uncover the biochemical defect involved. It's amazing how little we understand about something as common as Down's syndrome. We know that it's associated with three copies of chromosome 21, but why an extra copy of that chromosome should have such devastating effects is not known. Down's syndrome may be an extremely difficult disease to tackle, though, since a whole chromosome is involved. Diseases caused by defects in single genes should be a lot easier.

HILDEBRAND: Once a disease has been traced to a defect in a gene, copies of the normal gene can be produced by recombinant DNA techniques, and this opens the possibility of gene therapy, of treating the disease by introducing the normal gene into the patient. Gene therapy is still very much in the experimental stage, and it's a sensitive issue because of the ethical and legal implications of intervention with the human genome.

MOYZIS: Yes, but probably no one is going to complain about attempts to apply gene therapy to really terrible diseases—say, Lesch-Nyhan syndrome with its symptoms of severe mental retardation, kidney damage, cerebral palsy, and self-destructive behavior. What is worrisome to many is the specter of eugenics, of using such techniques to alter mental or physical characteristics. Many people feel it's one thing to genetically engineer plants or animals for desirable traits, and quite another to genetically engineer humans. In the first place, who's to say what are desirable traits? I strongly believe, however, that the types of genetic tampering that are the most feared are exactly the ones that are the least likely to be performed in the immediate future, mainly because we don't know how. It's extremely unlikely that anyone alive today will ever see genetic engineering applied to so complex a trait as intelligence. We haven't the foggiest notion

Making the Libraries

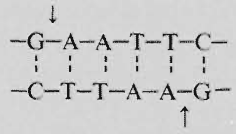
Cloning the DNA

The cells have been cultured, the chromosomes isolated and sorted. Now can begin the final act in the making of a library—that of multiplying the DNA in the sorted chromosomes by a factor of more than a million. Some tricks of recombinant DNA technology make such great amplification possible, and the magicians who perform these tricks are lowly but accommodating bacteria and viruses.

Briefly, four steps are involved in cloning the DNA: preparation of DNA fragments from the sorted chromosomes and from bacterial viruses (bacteriophages, or simply phages); formation of recombinant DNA molecules including both chromosomal and viral DNA fragments; assembly of phages containing the recombinant DNA molecules; and cloning of the phages (and hence replication of the chromosomal DNA fragments) in a bacterial host. For concreteness we describe, and illustrate in the accompanying figure, the making of what we call a Phase I library, one based on relatively small chromosomal DNA fragments. (Our Phase II libraries, which will serve a different purpose, will be based on larger fragments.)

First the DNA molecules in the sorted chromosomes must be extracted from the other chromosomal components. For reasons that will become clear later, this extraction must be performed very gently to minimize breaking the very long and very fragile DNA molecules.

The extracted DNA is then digested with the restriction enzyme *Eco*RI, which recognizes the sequence of six bases −G−A−A−T−T−C− and, as shown below, cuts the DNA at each occurrence of this sequence by breaking one covalent bond on each strand:

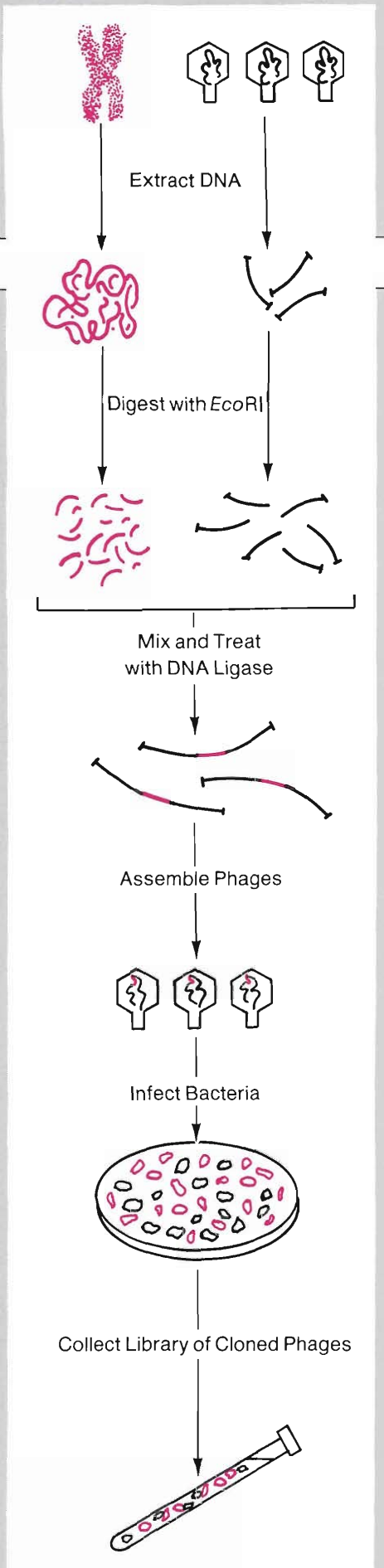


*Eco*RI is produced by certain strains of the bacterium *Escherichia coli* as a defense against infection by the DNA of bacteriophages. The enzyme severs the viral DNA and thus prevents its propagation in the bacterium. (To avoid self-destruction, the bacterium also produces another enzyme that modifies its own DNA so that it cannot be severed by *Eco*RI.) Other bacterial species produce restriction enzymes that recognize other sequences. For example, *Hin*dIII, the restriction enzyme used by our collaborators at Livermore for their Phase I libraries, is produced by the bacterium *Hemophilus influenzae*. Like *Eco*RI, *Hin*dIII recognizes a six-base sequence (−A−A−G−C−T−T−) and makes staggered cuts.

Complete digestion of the chromosomal DNA with *Eco*RI (or *Hin*dIII) produces a set of relatively small fragments. We can estimate the average length of these fragments by assuming a random distribution along the chromosomal DNA of equal numbers of the four bases A, T, G, and C. Then the probability of any six successive bases being the *Eco*RI (or *Hin*dIII) recognition sequence is $(1/4)^6$, which implies an average of about 4100 base pairs between cleavage sites. (An intact DNA molecule in the largest human chromosome (number 1) is about 2.5×10^8 base pairs in length.)

We, and our collaborators at Livermore, use DNA from the bacteriophage Charon 21A as a cloning vector, that is, as a medium capable of accommodating foreign DNA segments for replication. Charon 21A is a modified version of the λ bacteriophage; its linear DNA molecule is about 42,000 base pairs long, terminates with short stretches of single-stranded DNA, and includes a single *Eco*RI cleavage site.) The DNA molecules extracted from the phages are cut by *Eco*RI into two "arms." The cut ends of these arms are then stripped of a phosphate group (for a reason to be explained later) by treatment with the enzyme alkaline phosphatase.

The next step is fairly simple, at least in theory. We mix the two DNA preparations and let nature take its course—nature in this case meaning the tendency for single strands of DNA to "stick together" (become linked by hydrogen bonds) at regions with complementary bases. Since the cutting action of *Eco*RI has equipped both the chromosomal DNA fragments and the vector arms with just such sticky ends, aggregates consisting of two arms linked by a (single) chromosomal DNA fragment will form. (Herein lies the reason for gentle extraction of the DNA from the chromosomes: physical breaks do not result in sticky ends.) These aggregates are then converted into intact DNA molecules by treatment with the enzyme DNA ligase, which establishes a covalent bond at the break along each strand. The result is a set of recombinant DNA molecules—that is, molecules of phage DNA carrying human DNA fragments. (To increase the probability of forming recombinant aggregates, the chromosomal fragments are mixed with an excess of arms. Therefore, many of the aggregates formed will consist simply of two arms. But DNA ligase cannot repair the



III. CLONING THE DNA

DNA in such aggregates because of the missing phosphate groups. So our foresight in the previous step reduces the background of nonrecombinant DNA molecules in the final library.)

Next we assemble Charon 21A phages, drawing on proteins produced by mutant varieties of the phage for the necessary structural components and on our supply of recombinant DNA molecules for the genetic component. This step unavoidably introduces a deficiency in the resulting library: those recombinant DNA molecules containing fragments longer than about 9100 base pairs are too large to be encased in the heads of the phages, and therefore the library will lack that portion (about 33 mass percent) of the chromosomal DNA residing on such fragments. It is for this reason that we and our collaborators at Livermore use different restriction enzymes and thus produce libraries based on different sets of chromosomal DNA fragments. The expectation is that the DNA missing from one laboratory's version of a library will be found in the other.

The final step in making the libraries, that of cloning the recombinant DNA molecules, is accomplished by seeding the phages onto a "lawn," or monolayer, of an appropriate strain of *E. coli*. When a phage injects its recombinant DNA molecule into such a cell, the biological apparatus of the cell is subverted to production of new phages. The cell ruptures when between one and two hundred new phages have been produced, and these then infect surrounding cells. The process continues, resulting in formation of a plaque, or clear area, in the translucent lawn of bacteria. In each plaque are some one to ten million infective phages, each containing an exact copy of the recombinant DNA molecule carried by the original infecting phage.

These vastly multiplied recombinant DNA molecules, then, constitute a library—conveniently and safely packaged within the proteinaceous armor of phages. This construction of a library puts us at the starting line. Its application to problems of clinical and fundamental significance lies ahead. ■

what genes or combinations of genes are involved; in fact, we haven't come close to settling even the nature-versus-nurture issue. And even when a single defective gene is known to be the culprit—as it is in certain kinds of mental retardation—developing the techniques for introducing the normal gene and getting it expressed properly will take a long time.

HILDEBRAND: Yes, routine clinical application of gene therapy is a long way off, especially in light of the ethical and legal questions it raises. But in the process of ironing out those questions, society should focus on the extraordinary benefits to be derived from the technology rather than its "Brave New World" implications.

MOYZIS: I think other developments in molecular genetics pose more immediate concerns. For instance, the availability of probes specific to the X and Y chromosomes has made it very easy to determine the sex of a fetus. Since even in "sexually liberated" western countries the preference for having a male child first is still strong, routine determination of fetal sex could conceivably lead to dramatic changes in the ratio of the sexes. The ethical, social, and economic implications of this possibility need immediate consideration.

SCIENCE: *Are issues such as these heavy on your minds as you go about your research?*

CRAM: I don't think so. The goals of basic research are information and understanding. It's up to society to determine how those are applied. Take the photoelectric effect, for example. Its applications are mainly viewed as beneficial, although a few may be regarded by some as deplorable. But a scientist doesn't hide his head in the sand in the face of information that can possibly be applied in a harmful way.

MOYZIS: I would second that. You can't be paralyzed because a bit of knowledge can be used inappropriately. And I repeat that most of the worrisome aspects of gene therapy are a long way down the road. A major problem that must be solved before it can be applied to many diseases is how to introduce the good genes into only the appropriate cells of

a patient. If the defect is in a gene that is expressed, say, only in liver cells, extra copies of this gene in other cells might cause a great deal of mischief.

I am somewhat amazed, however, at the speed with which the medical community is moving toward gene therapy. I'm aware, though, that clinicians confronting patients have a different viewpoint than we do. It's no particular virtue on the part of people in basic research to be more cautious.

I think the patient and the patient's family should have a large say in deciding about experimental treatments. If a child of mine had Lesch-Nyhan syndrome, I'd be willing to take a chance on gene therapy. Terminal cancer patients are taking similar chances right now with other experimental therapies.

HILDEBRAND: An indirect form of gene therapy has already been tested clinically. Patients with a form of thalassemia were treated with a drug that would—it was hoped—reactivate the fetal hemoglobin genes. The drug worked, but since it is a potential carcinogen, the ultimate result may be initiation of tumors or other defects through inappropriate activation of other normally silent genes. Of course, all therapeutic regimes have the risks of unwanted side effects. Gene therapy presents the special risk of possibly altering the human gene pool. But so do chemotherapy and radiation therapy..

Direct gene therapy seems imminent, and public interest is high, even though the genetic diseases now being considered as candidates for this method of treatment are extremely rare. The Office of Technology Assessment has published a background paper on gene therapy, and the NIH's Recombinant DNA Advisory Committee has appointed a Working Group on Human Gene Therapy. The legal problems may take a long time to clear up, though.

DEAVEN: Yes, but that can change rapidly. About six years ago Margery Shaw, Director of the Institute of Medical Genetics at the University of Texas in Houston, gave a colloquium here at Los Alamos on the legal aspects of genetics. At that time the fetus had absolutely no legal status in this country or in other western countries. Last year she gave an update of that talk at the annual meeting of the American Society of Human Genetics. She reported that in parallel with scientific developments in recognizing forms of fetal abuse, such as overconsumption of alcohol or other drugs by the mother, principles are being established concerning the legal status of the fetus and its right to be born in a state of good health. My feeling is that a similar thing will happen with gene therapy, and that by the time gene therapy is a reality we will have arrived at legal guidelines for its application.

SCIENCE: *Do you think society turns to scientists for ethical guidance as well as technical expertise?*

CRAM: Yes, and I think most of us accept that role. But the scientific community can lose credibility because even scientists can interpret the same technical data differently. This happened when policies for radiation exposure were being established.

MOYZIS: That's because we're human beings just like everybody else. Although we may be more aware of the possible applications of our research than the average person, most of us are not particularly qualified to spout off about the ethics of those applications. Whether or not I favor abortion or genetic engineering of humans is in some ways based on my scientific expertise but in other ways is not. We're often called upon at cocktail parties to discuss these sticky issues, but I tend to feel that most scientists should be asked for nothing more than scientific input. Most scientists aren't any more objective than anybody else when it comes to judgments that involve emotions and prejudices.

DEAVEN: Yes, but I do think scientists should spend some time thinking about the implications of their research and should actively participate in the formulation of ethical and legal guidelines for society. Some have altered their careers in this direction. I'll mention Margery Shaw again. After acquiring an M.D. and having a successful career in human genetics, she went back to school and got a degree in law because she saw the legal implications of new developments in human genetics. We need more people like that.

HILDEBRAND: This is all very interesting, but perhaps we should retreat to a more direct application of the libraries—that of developing more sensitive methods for detecting mutations.

MOYZIS: Yes, we badly need a means of detecting damage to human DNA caused by low-level exposure to mutagens. It's not hard to detect the damage caused by high-level exposure. Ionizing radiation, for example, causes hundreds to thousands of double- and single-strand breaks in the DNA molecules of a single cell at exposures over 1000 rads. But very few people ever get that kind of exposure. Most of the concern, medical and legal, is about the effects of low dose rates. Is exposure to fallout from the test of a nuclear weapon in Nevada some thirty years ago, or to Agent Orange in Vietnam, the cause of some of today's cancers? Does living in Los Alamos at an altitude of 7500 feet cause a higher incidence of spontaneous abortions? Who knows? It's very difficult to make these assessments.

Currently, the incidence of mutations is determined from the observable changes within a small number of genes, most commonly a single gene. Since a single gene is only about one-millionth of a DNA molecule, the assays are not very sensitive. But with a battery of hundreds or perhaps thousands of gene probes pulled from the libraries, a much larger fraction of the DNA can be assayed for damage. A wealth of cytogenetic data indicates that some regions of certain chromosomes are more susceptible to damage. Obviously, probes for those regions would be of particular value.

DEAVEN: Los Alamos is already working on detecting mutations induced by low-level radiation by assaying what are known as tandem repeats for RFLPs.

MOYZIS: Yes, and although our results on irradiated cells are very preliminary, they're very encouraging. We're working with the repetitive sequences in human DNA, which

make up about one-quarter of the total. Many of these repetitive sequences are thought to be involved in defining chromosome structure or in switching genes on and off. They often jump to new locations in the DNA, altering the gene sequence they move into. We've cloned and characterized over 75 percent of the repetitive DNA sequences present in human DNA. Most of these sequences occur about a hundred thousand times per genome, and about a third of them occur in tandem, one right after another. With a single RFLP probe for such a tandem repeat, we can sample many orders of magnitude more DNA for mutations than with probes for a few genes. With this increase in sensitivity, Susan Atchley, a graduate student in the Laboratory's Genetics Group, has been able to detect damage to DNA at a much lower level of radiation exposure than has been reported previously. But it appears that most of the observed changes in the DNA disappear with time following irradiation. Presumably the changes are being repaired. These preliminary results have led us to facetiously throw around the idea—although maybe it's not so facetious—that a little bit of radiation may be a good thing in the sense that it keeps the repair enzymes induced to such a level that they can more readily cope with larger doses.

The repair we are seeing has relevance to the long-standing question of whether or not such a thing as a threshold for damage exists. Perhaps there is a dose rate below which the DNA repair enzymes can handle the damage, even with continuous exposure.

CRAM: If repair is blocked, what levels of damage can be detected?

MOYZIS: We haven't done that experiment yet.

DEAVEN: With the current techniques for assaying genetic damage, large groups of exposed people must be studied to get good statistics. Very few such groups are available. In fact, despite a massive effort since 1945, the data emerging for the largest group—the Hiroshima and Nagasaki populations—are questionable in terms of genetic significance. The power of the newer techniques is that they can be applied to small populations because they provide so much more information per person.

MOYZIS: The recent incident in Mexico, in which some two hundred people were exposed to cobalt-60 pellets from a junked radiation therapy machine, gave us a new population for study, but those people got horrendous doses.

DEAVEN: One problem with most of the available study populations is that the exposure is not known with certainty. We can estimate the exposure by determining levels of chromosome damage, but this method is not very accurate, especially if the chromosome analysis is not done immediately. Data from laboratory experiments, such as those Bob is talking about, should provide better correlations between damage and dose.

CRAM: Is there so much variation in restriction fragment lengths among individuals that detecting mutations requires a control sample for each individual?

MOYZIS: We've found that the variation

The success of the Los Alamos contribution to the Gene Library Project is due to the competence and enthusiasm of a large number of people. Pictured here are Kevin Albright, Marty Bartholdi, Bill Bentley, Nancy Brown, Evelyn Campbell, Doug Chritton, Scott Cram, Larry Deaven, Ed Hildebrand, Paul Jackson, Ted Lobb, Mary Luedemann, John Martin, Linda Meinke, Julie Meyne, and Bob Moyzis. Jim Jett, Jon Longmire, and Chris Munk were unavailable for the photograph.

depends on the sequence. For damage estimates we've concentrated on sequences that seem not to vary in the population at large. Preliminary variation was estimated by isolating DNA from blood samples of any individual who was unfortunate enough to walk past my lab. The Life Sciences Division here has enough people with different ethnic backgrounds to make those estimates meaningful.

Repetitive sequences that do vary in the population are also interesting. Ones that map to particular chromosome locations will make exquisite probes for those chromosomes and will be useful in prenatal diagnosis, since orders of magnitude less fetal DNA will be needed for analysis. Ed and I are currently searching for such sequences in the libraries.

HILDEBRAND: Another application of probes from the libraries is investigating permanent changes in the genome on a larger time scale, for example, in evolution and speciation.

DEAVEN: Yes. Using a battery of probes that represent an ordered arrangement of human genes, one can look for patterns of constancy and of change in the genomes of plants and animals. That sort of information should give some insight as to just what the limits are on rearrangement of the genome. For example, if a group of genes has remained linked closely together over hundreds of millions of years of evolution, that's a pretty good sign that any perturbation of that conserved linkage is deleterious or even lethal. Taxonomists might be able to use such information as the basis for a better definition of speciation.

MOYZIS: It might also help solve the dilemma of cross-species extrapolation. The relevance of evidence from an animal study to humans can always be questioned. But if it is known what genes or organizations of DNA are unique to, say, mice and to humans and common to both, estimates of risk assessment from animal studies will be on firmer ground.

DEAVEN: Pedigree analysis will also benefit from the libraries. Pedigree analysis is the tracing of a genetic trait through generations of a family to determine whether the gene for the trait is dominant or recessive and whether it is located on one of the sex chromosomes or on an automsome. For this job geneticists have been using phenotypic evidence of the presence of a gene. But some genes don't provide unambiguous phenotypic evidence. The use of RFLP probes for such genes will make pedigree analysis easier and more accurate.

MOYZIS: In the near future enough probes for RFLPs will be available that it will be possible to assign paternity with 100 percent certainty.

CRAM: I believe a test like that was done on the pandas at the Washington zoo to find out whether the female was impregnated by her mate or by artificial insemination with sperm from a panda in England.

MOYZIS: I didn't realize that, but similar techniques are used routinely to distinguish different strains of plants. DNA might come to replace fingerprints for certain identification purposes.

SCIENCE: *Has production of the libraries been fairly routine or full of surprises?*

VAN DILLA: Dealing with life processes and living things practically guarantees surprises. An example is our experience with hybrid cells. The first hybrid cell line we worked with was just a dream. The cells were easy to grow and, as advertised, contained three human chromosomes, which were easy to sort. So it came as a rude surprise that other hybrid cell lines didn't perform nearly so well. One was supposed to contain human chromosome 1, but after the cells were grown, we couldn't find any to sort. We had known that hybrid cells have a habit of losing human chromosomes as they multiply but had been lulled into a false sense of security by the behavior of the first cell line.

Even so routine a process as chopping up DNA with a restriction enzyme gave us a surprise. We were getting a cutting efficiency of only about 1 percent even though we were using a standard restriction enzyme whose cutting efficiency was well documented to be close to 100 percent. At first we blamed a bad batch of enzyme, but that idea didn't check out. We still don't know exactly what the problem is, but it's been solved by an additional purification of the DNA extracted from the chromosomes. So we assume the enzyme was being poisoned by some contaminant the chromosomes pick up during sorting.

The logistics of producing the libraries gave us a surprise too. When we switched to sorting with the high-speed flow cytometer in the spring of '84, the poor cell farmers were suddenly faced with keeping up with its ten times greater appetite for chromosomes. That shouldn't have come as a surprise, but it did, probably because we were too busy with other aspects of the project to think hard about the manpower needed to keep the high-speed sorter happy.

SCIENCE: *Have you enjoyed participating in the Gene Library Project?*

HILDEBRAND: Definitely yes, but there has been an element of frustration because it is taking a much longer time than we anticipated. We're still adapting to the reality that this project is a long-term research and development effort with a defined endpoint. The Laboratory has vast experience in dealing with long-term experiments in physics. Life scientists, however, have traditionally been accustomed to experiments that can be completed in a period of days, weeks, or months. To a physicist what we call an experiment would be just one step in a much larger process.

MOYZIS: Just a coffee break. But that does bring up a good point. For the Life Sciences Division the Gene Library Project is a major undertaking in terms of the time necessary to get to the starting point of actually using the libraries. The DOE may have been right in thinking that Livermore and Los Alamos were the only places where enough biologists would talk to each other long enough to get the project going. Physicists are more accustomed to projects involving lots of people. Biology is still very much done by one guy with his test tube at his bench.

DEAVEN: Group research in biology is going to become more and more common, though. As more and more sophisticated technologies come into use, the breadth of expertise needed in a research effort gets greater. This project is certainly an example of that trend. Between twelve and fifteen people, representing a variety of specialties, are involved on a day-to-day or week-to-week basis.

CRAM: The point that Ed made about the difference between biologists and physicists is worth emphasizing. We can do an experiment sometimes in an afternoon, but physicists often spend years on an experiment. A biologist might publish six papers a year, but physicists doing an experiment at an accelerator might publish a single paper with ten authors every couple of years. I think the Gene Library Project is an example that crosses that barrier.

MOYZIS: Biology is still such a wide open field with so much left to learn that we can get significant, publishable results in a relatively short time. A number of people involved in this big project are going to have little holes in their curricula vitae because they felt it was so important. Not many of the biologists at universities would have been willing to devote two years or so to getting these libraries made.

DEAVEN: Working on something so different as this project has certainly been an interesting challenge. Sometimes though, and this is probably true of all of us, I dream of being on the other side of the fence—opening my mail, unwapping a little vial, and actually doing some research with a library. ■

Acknowledgments

The following people have made substantial contributions to the organization and administration of the National Laboratory Gene Library Project: Mark W. Bitensky, Leader of the Life Sciences Division at Los Alamos National Laboratory; Brian D. Crawford, now Director of the Molecular Biology Program at Long Island University; M. Duane Enger, now Head of the Department of Zoology at Iowa State University; Mortimer J. Mendelsohn, Associate Director of the Biomedical and Environmental Research Division at Lawrence Livermore National Laboratory; David A. Smith, Office of Health and Environmental Research, U. S. Department of Energy; Carleton C. Stewart, Leader of the Pathology Group at Los Alamos National Laboratory; Robert Wagner, Consultant at Los Alamos National Laboratory; and Ronald A. Walters, a member of the staff of the Associate Director for Chemistry, Earth and Life Sciences at Los Alamos National Laboratory.

Further Reading

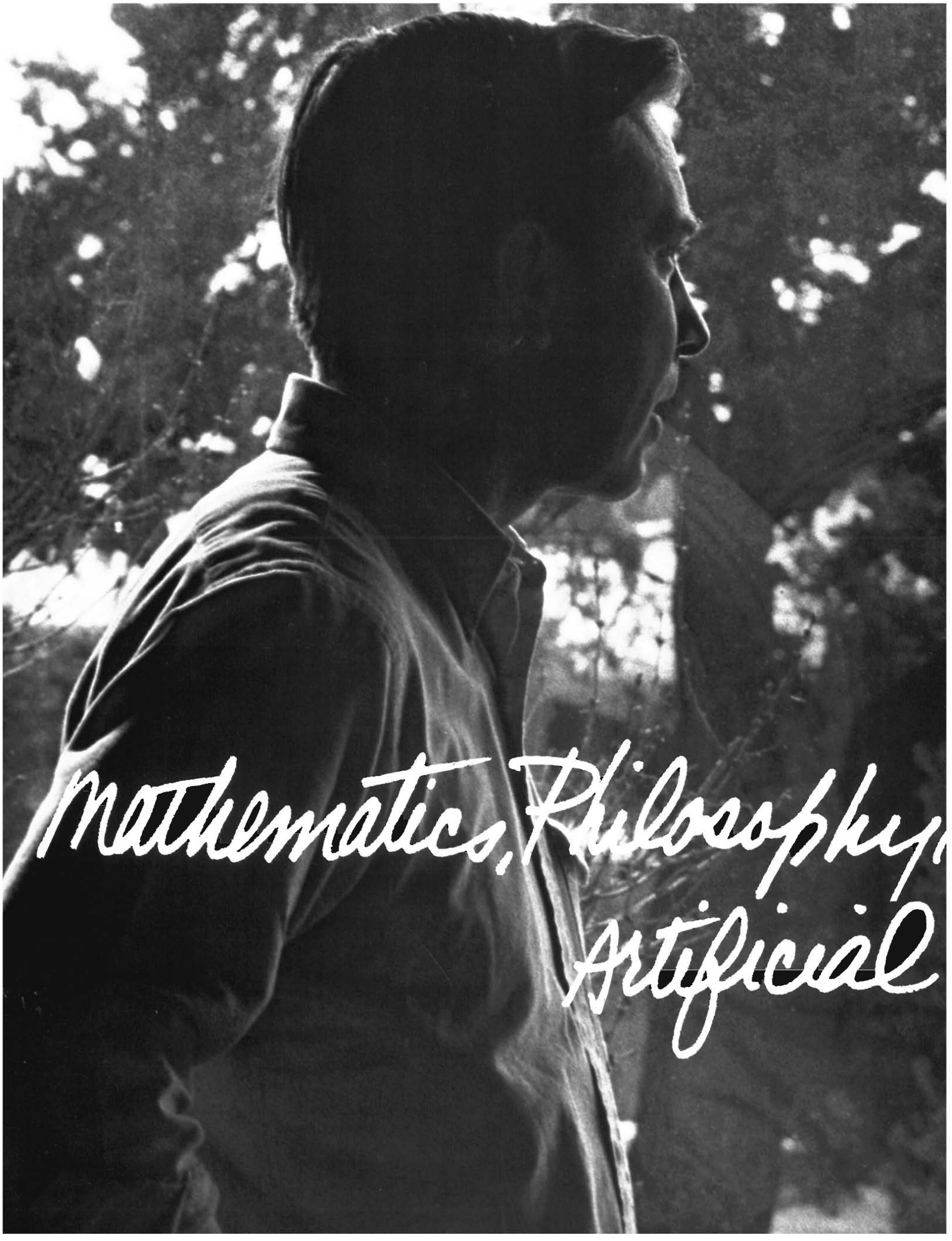
James D. Watson, John Tooze, and David T. Kurtz. *Recombinant DNA: A Short Course.* Scientific American Books, 1983; distributed by W. H. Freeman and Company.

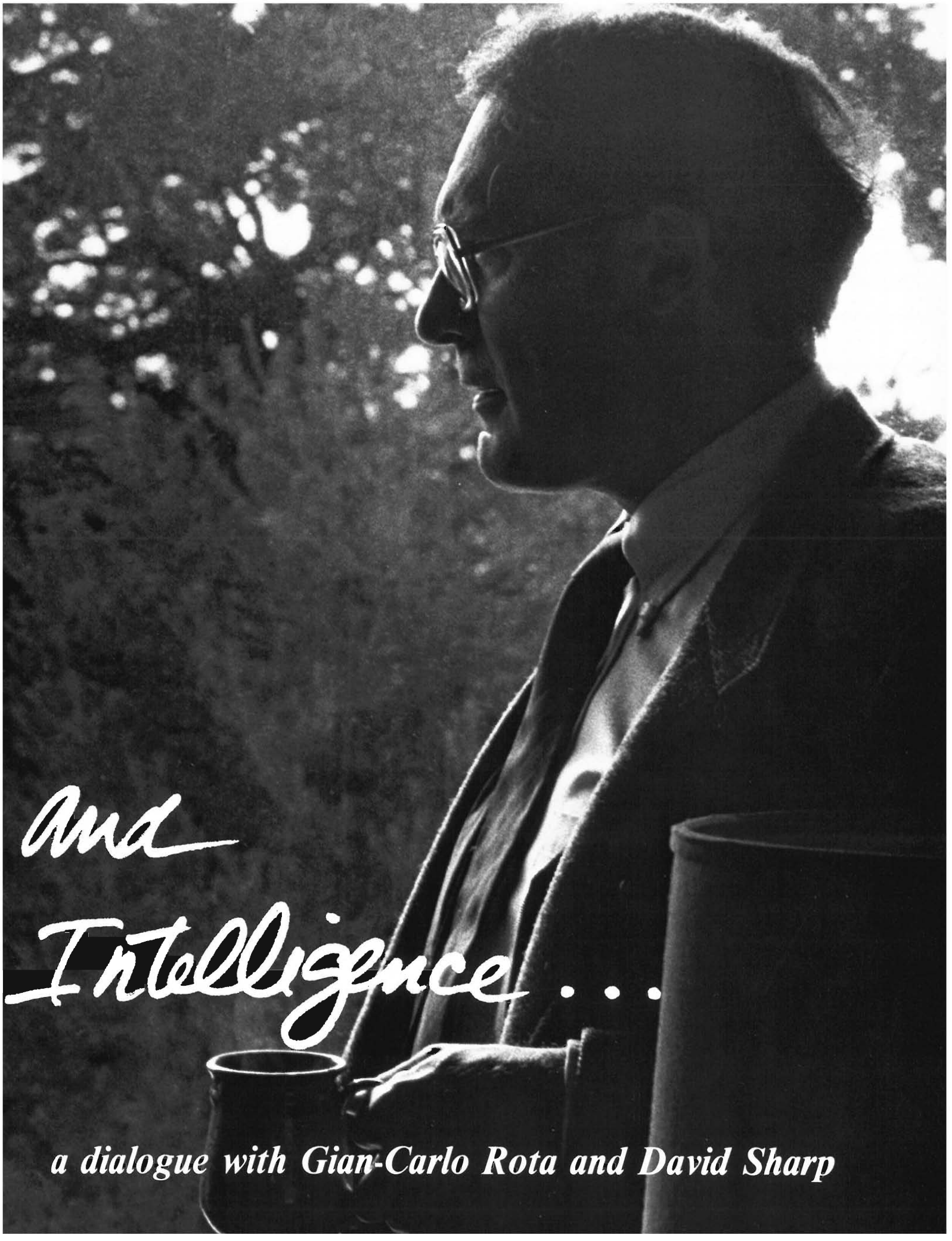
Alan E. H. Emery. *An Introduction to Recombinant DNA.* Chichester, England:John Wiley & Sons, 1984.

Maya Pines. "In the Shadow of Huntington's." *Science 84,* May 1984, pp. 32-39.

Office of Technology Assessment. *Human Gene Therapy—A Background Paper.* Washington, District of Columbia:U.S. Congress, Office of Technology Assessment, OTA-BP-BA-32, December 1984.

Yvonne Baskin. "Doctoring the Genes." *Science 1984,* December 1984, pp. 52-60.

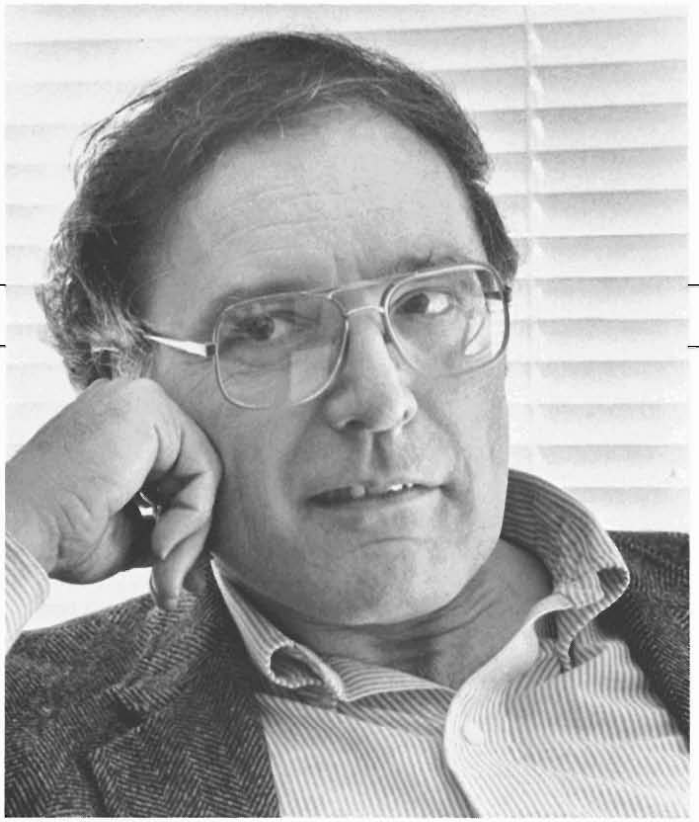

Mathematics, Philosophy,
Artificial



And Intelligence...

a dialogue with Gian-Carlo Rota and David Sharp

The Talmud and the Tao are good trainers of scientific minds.



Gian-Carlo Rota, survivor of seventeen years as a consultant to Los Alamos, is a leader in the field of combinatorics and a popular professor of applied mathematics and philosophy at MIT. He was born in Milan, raised, as he says, by the iron hand and given charge at the tender age of eight of his father's impressive library, a place where he developed a liking for the smell of printed matter. Erudite, worldly, and immensely generous to his friends and students, Gian-Carlo shares his penetrating insights with the accent on humor. Asked about his role at Los Alamos, he replied, "I wish I knew. I manage to snoop around and every once in a while pop into the Director's office and have a chat with him. I can be outspoken—no one is offended because next Sunday I leave anyway. By the time I return everything's forgotten and we can start all over again."

DIALOGUE

SCIENCE: What's going on in science? Does anyone have the big picture, or are people too stuck in their narrow specialties?

ROTA: I'm afraid I can talk only about mathematics. Until a few years ago, you found everywhere a strong tendency to specialization; now you see much straddling between branches of mathematics that used to be tight, separate compartments. People who can work in more than one field, and who are able to see analogies between seemingly disparate concepts, will be the ones who set the pace for the future. The younger generations are not wasting any time in picking up this hint.

SHARP: In physics, too, the trend toward synthesis is evident. The recent marriage of particle physics and cosmology is leading to an understanding of the early universe. Nonlinear dynamics now cuts across hydrodynamics, many-body theory, and plasma physics.

SCIENCE: What about biology?

SHARP: A good example of scientific interaction is neural networks. The theory was started by a neuroscientist, Warren McCulloch, and an applied mathematician, Walter Pitts, caught the eye of a pure mathematician, von Neumann, and was worked on by a mathematical logician, Steve Kleene, and eventually by an engineer, Claude Shannon. The subject borrows from neurobiology, artificial intelligence, and classical logic. It purports to offer a synthetic approach to the old and difficult scientific problem of the working of the mind.

ROTA: As you were saying, a new unit in science is being formed that remains to be named. It will include the best of theoretical computer science, neurophysiology, molecular biology, psychology, and the mathematical theory of information. It will be important to name it properly. As the Latin proverb says, "*Nomen est omen.*"

SHARP: The name is a presage.

ROTA: We need to name it glamorously, to make it into a respectable profession, and to bring people together to work on it. This is beginning to happen. It is the central drama of today's science.

SCIENCE: Has it something to do with understanding the idea of intelligence?

ROTA: It has to do with muscling into processes of behavior, whether human or machine.

SHARP: Including intelligence.

ROTA: But not specifically intelligence. As is true in the early stages of any science, it's hard to tell the crackpots from the geniuses. They are mingled together, as they were in the beginning days of physics when, for example, Kepler thought he could classify the distances of the planets from the sun by the lengths of the sides of the five regular solids inscribed in a sphere. Newton himself believed in magic.

SHARP: Not to mention alchemy.

ROTA: Nevertheless, he discovered the laws of mechanics. Similarly, we could view the new science as a melting pot of good guys and bad guys, of con men and serious people. It is hard to sort them out. Sometimes the same person is both.

SCIENCE: Why do some scientists remain active, while others burn out?

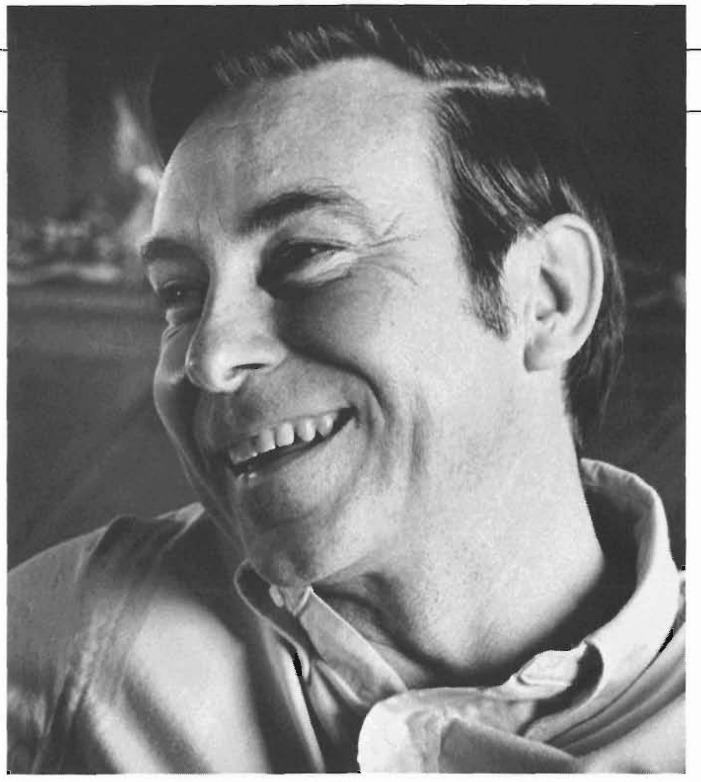
ROTA: I have a one-word answer. It is the word *Kultur*. A broad cultural background, the learning at an early age in the family of the value of things intellectual, is the main factor that keeps people from burning out. That's the reason many first-rate intellectuals come from Jewish or Chinese backgrounds, where they were exposed to intellectual values at an early age. The Talmud and the Tao are good trainers of scientific minds.

SHARP: There is more to it than that. You go into science because you like it. That liking carries you through the ups and downs of your career, through the plateaus you reach from time to time in your work. While on one of those plateaus, it is important not to be afraid to try to learn new things.

ROTA: Call it intellectual chutzpah.

SHARP: If one gets caught doing a small extrapolation of what someone else has already done, the work will be neither daring nor

The farthest reaching contribution of modern psychology . . . is the realization that intelligence is not a single monolithic faculty.



David Sharp, a Fellow at Los Alamos, shares with Gian-Carlo an undergraduate background at Princeton and a capacity for working in several fields at once. Always intense, enthusiastic, and persevering, David switches from neutral networks to Rayleigh-Taylor instabilities to quantum field theory with an agility reminiscent of his figure-skating youth. He is also working with Gian-Carlo on the theory of stochastic processes, and on weekends they moonlight on the philosophy of mathematics.

exciting. Once the excitement is gone, you can kiss good-by to creative contributions.

SCIENCE: Who are the daring scientists of our times?

SHARP: Feynman is one. Near the end of his career, what is he doing? He is not working on particle physics where he made such enormous contributions. He's working on computer science, on tiny computers based on quantum-mechanical principles; he's now helping to teach a course at Caltech on computer science.

ROTA: My Ph.D. thesis advisor, then at Yale, Jack Schwartz, is another example. He worked on functional analysis during my graduate-school days. Today he's deeply engaged in computer design and robotics—quite a switch.

SHARP: That brings to mind another heavyweight, Jim Glimm, who has made truly outstanding contributions in mathematics and mathematical physics, from quantum field theory to fluid dynamics to C*-algebras. Unfortunately, the department structure of today's universities discourages people from change.

SCIENCE: You said that a new science is emerging. Are young people being trained in it? Is something serious really happening, or is it all phony?

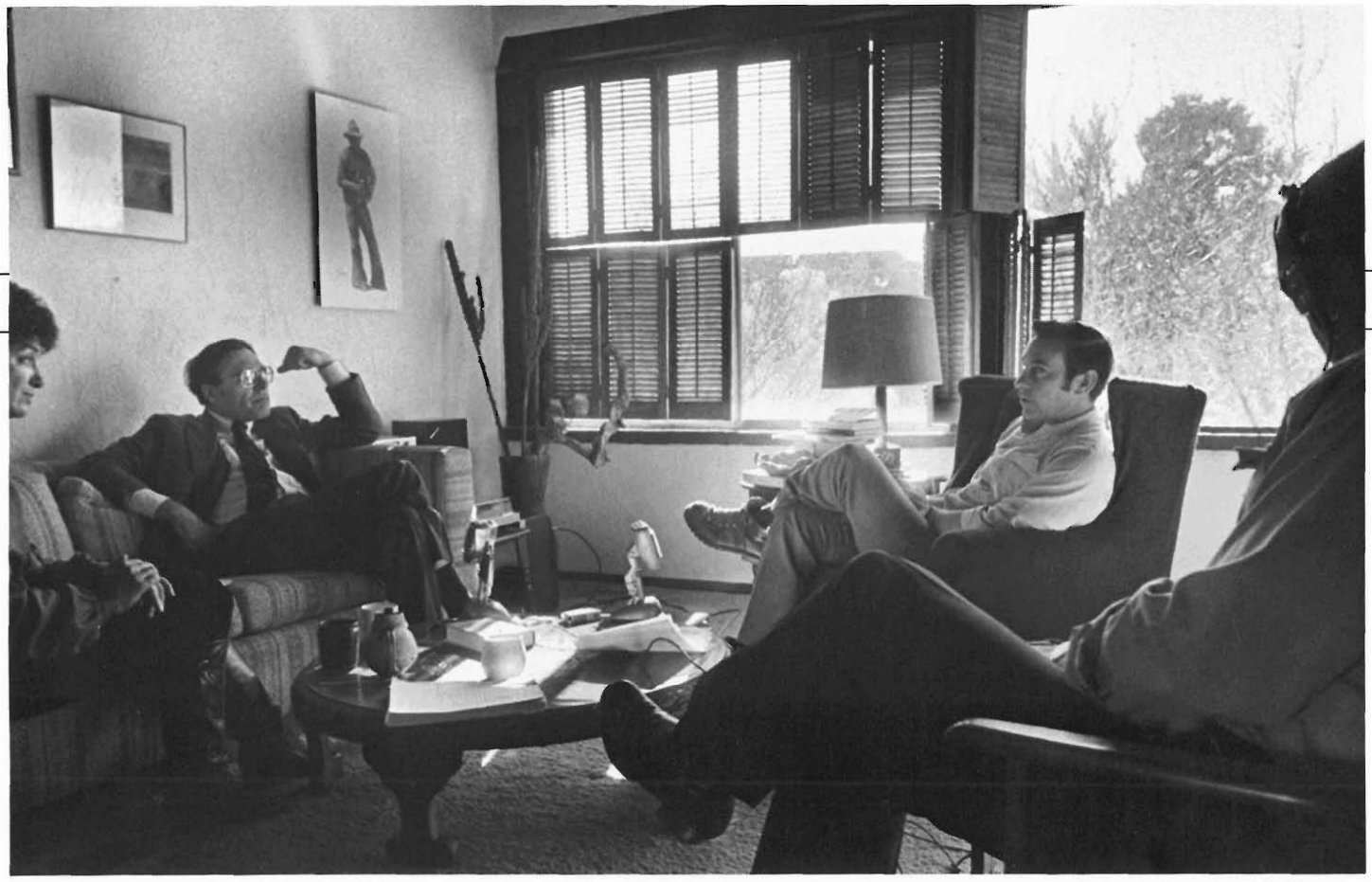
ROTA: Everyone agrees that we must reform our programs at the universities. But it is hard to change the sluggishness of an academic environment, and tension is building. We need imaginative new programs in mathematics; we need daring departures that straddle mathematics, physics, and biology.

SHARP: Mathematicians are concerned with the decline in the number of students in graduate programs, as well as with the decline in support for mathematics. For some decades there has been a split in the mathematics community. Some feel that mathematics is enriched by contact with nature, with physics and biology. Others take an aesthetic, self-contained view and argue that the internal logic of mathematics will dictate its development. This point of view is sometimes identified with the Bourbaki school. In the old days Hermann Weyl and David Hilbert were deeply involved in the development of the theoretical physics of their time. In our time computers have caught up with physics in providing the external stimulus for mathematicians.

ROTA: For mathematical logicians most of all. Witness the exodus of mathematical logicians out of their field into computer science. Thanks to logicians we have sophisticated programming languages and superior software. Computer science has become too important to be left to engineers. Fortunately, physicists and mathematicians are switching to computer and hardware design in great numbers, attracted by the higher pay. Professors of computer science, even those who do not know how to read and write, make double the salary of even the best mathematicians.

SCIENCE: Isn't it true that mathematicians switch to computer science when they're burning out? When do you know a mathematician is through?

ROTA: Von Neumann used to say that a mathematician is finished by the age of thirty. As he got older, he increased the age to thirty-five, then to forty, to forty-five, and soon to fifty. We've inherited from the 19th century a misleading notion that mathematicians have to do their work early or they're finished. That's not true. The kind of work a mathematician does as he grows older changes. An older mathematician will work on questions of wider scope, whereas a younger mathematician will choose to work on a single difficult problem. A variety of talents are required for the scientific community to thrive. What is perhaps the farthest reaching contribution of modern psychology to human welfare is the realization that intelligence is not a single monolithic faculty that can be measured on a linear scale. You may be smart at working out math problems, but stupid at doing everything else. The old IQ test was very effective at testing one kind of intelligence, what we might call quiz-kid smarts, but it did not test any of the other kinds.



SCIENCE: Are you born a mathematician, or can you be taught to be one?

ROTA: Most people have some talent or other that could make them into reasonable mathematicians, given the motivation. But there is also such a thing as raw mathematical talent, just like there is talent in music. Unfortunately, we do not know how to measure "raw talent," though we can recognize it when we see it.

SHARP: Psychology is now beginning to discover the subtleties of learning. People are either geometric and visual, or verbal and logical, or kinesthetic. Psychological studies of math teaching have shown that different teaching techniques work for different people, depending on whether they are verbal, visual, or kinesthetic. Unfortunately, these subtleties will take a long time to seep down into grade-school teaching.

SCIENCE: Most people teach themselves anyway.

ROTA: But a teacher can be effective in discovering his students' talents and in encouraging those talents. A good teacher does not teach facts; he teaches enthusiasm, open-mindedness, and values.

SHARP: Somebody once said that teaching is not efficacious except in those few cases where it is superfluous.

ROTA: A good teacher brings out the best in his students. What young people need the most is encouragement. Left to themselves, students may not know how to decide what is worthwhile. They may drop an original idea because they think someone else must have thought of it already.

SHARP: Students need to be taught to believe in themselves and not to give up easily.

SCIENCE: Is there a conflict between letting students follow their own crackpot ideas rather than learning what someone else has already done?

ROTA: There is a ratio by which you can measure how good a mathematician is, and that is how many crackpot ideas he must have in order to have one good idea. If it's ten to one then he is a genius. For the average mathematician, it may be one hundred to one. You must have the courage to discard lots of attractive ideas. This is a feature of creative thinking that the man in the street fails to realize.

SCIENCE: And you have to try out all those ideas?

SHARP: Pretty much, but you mustn't become infatuated with the sound of your own words. You have to be ruthless in throwing out your own bad ideas. You have to constantly weed your own garden.

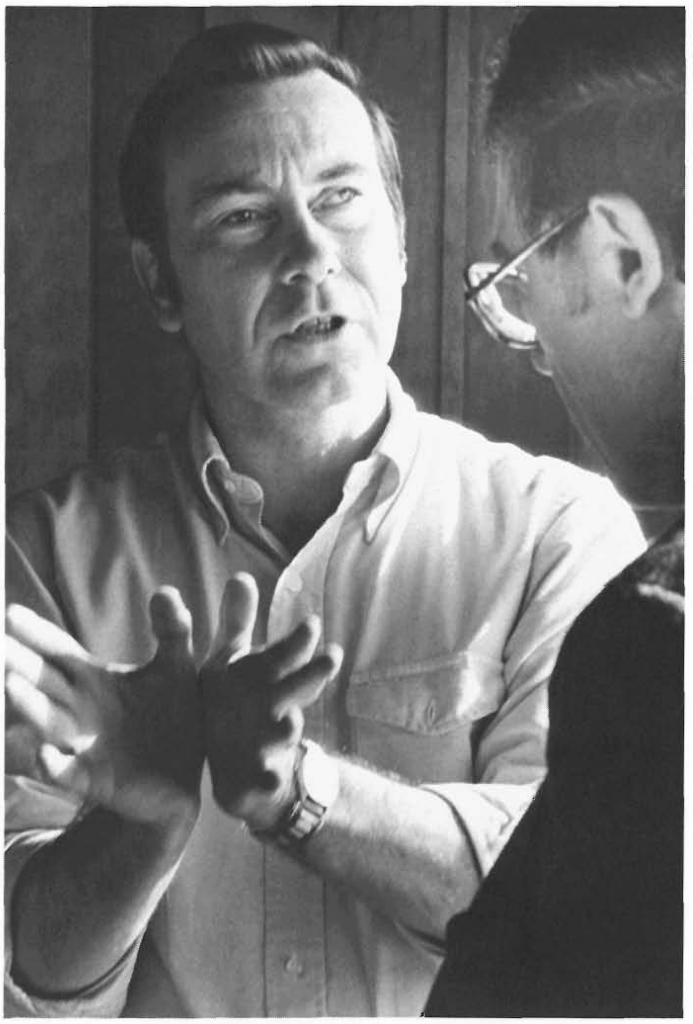
SCIENCE: Coming back to mathematical talent, how would you characterize it?

ROTA: An outstanding characteristic of mathematical talent is the ability to spot analogies. Another, one of the rarest, is the talent for applied mathematics, the talent for picking out of a maze of experimental data the two or three parameters that are relevant, and to discard irrelevant parameters. This is rare, because it is taught only at the shop level.

SCIENCE: How is mathematics applied?

SHARP: Most people, even some cultivated scientists, think that mathematics applies because you learn Theorem Three and then Theorem Three somehow mysteriously explains the laws of nature. That doesn't happen even in science fiction novels— it is pure fantasy. The results of mathematics are seldom directly applied; it is the definitions that are really useful. Once you learn the concept of a differential equation, you then see that in biology, say, there are differential equations all over. This you cannot see unless you take a course in abstract differential equations. Here, as everywhere, what applies is the cultural background you get from the course, not the specific theorems taught in the subject. If you want to learn to speak

There is nothing deadlier for a mathematician than to be . . . instructed to lay golden eggs.



French, you have to live the life of France, not just memorize thousands of words. Similarly, if you want to apply mathematics, you have to live the life of differential equations. When you live this life, you can then go back to your molecular biology with a new set of eyes that will see things that you couldn't otherwise see.

ROTA: I once naively entertained the thought that biologists could tell me what their mathematical problems were so that I could think of solving them. That's ridiculous. Biologists seldom have the mathematical view that is required to spot problems in the mathematics of biology that may be right there staring at them. A biologist will go on doing experiments all his life and never see anything deeper than binomial coefficients. It is not that the problems aren't there; rather, biologists by and large don't have the view that comes only from a solid education in mathematics.

SCIENCE: Once you have isolated a problem, how do you go about solving it?

SHARP: No mathematical problem is ever solved directly. You must manage to look at the surroundings, at the company it keeps. Eventually, with the assistance of the Holy Ghost, you might be able to see through your problem.

ROTA: All creative mathematicians and physicists I know work like this. They have constantly in their minds a list of a dozen of their pet problems. Anything they hear, they automatically test against their dozen problems. Every once in a while, there's a hit, and then people cry out, "He's a genius! How did he know that this was the right trick to apply to Problem Three?" Little do they know that the guy has been testing for years everything he hears against Problem Three. That's the way a lot of discoveries are made.

SCIENCE: What makes a creative mathematician?

SHARP: Rule One: Don't ask him to be creative. There is nothing deadlier for a mathematician than to be placed into a beautiful office and be instructed to lay golden eggs. Creativity is never directly sought after. It comes indirectly. It comes while you are complaining about too much routine work, and so you decide to spend half an hour on your secret pet project. Those are the occasions when you get good ideas. Or while getting ready to lecture your undergraduates, you realize that the textbook for the course is lousy, and that the subject has never before been properly explained. While you work on explaining some old material, lo and behold, you get a great new idea.

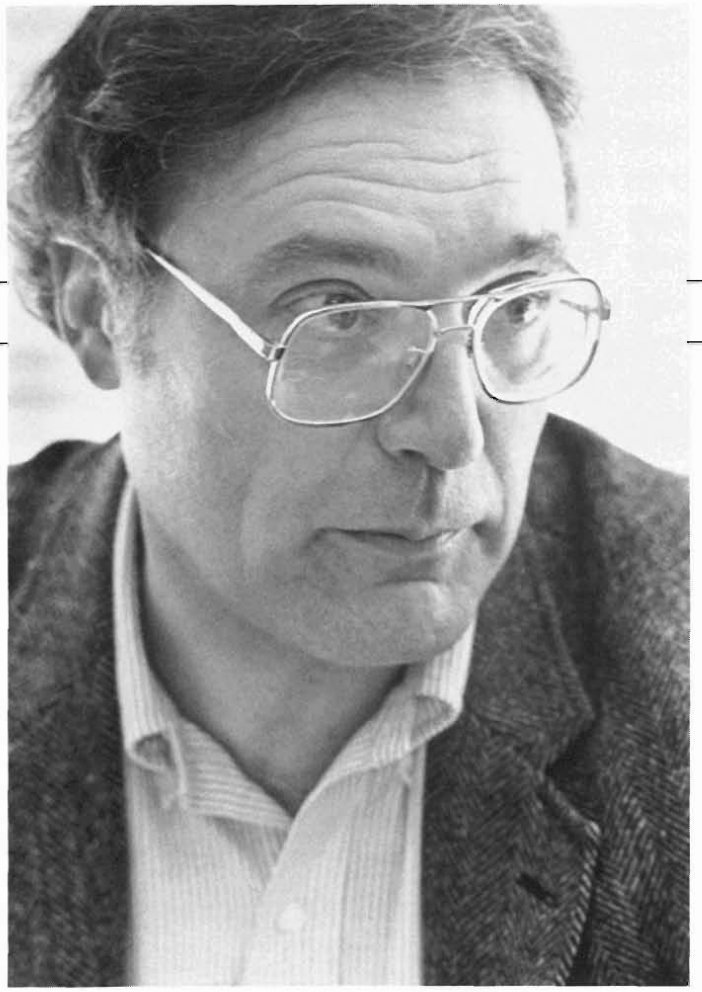
ROTA: Creativity is a bad word. Unfortunately, we must leave it in the books because the people in power believe in it with sanctimonious credulity. It is a dangerous and misleading word.

SCIENCE: We recently sponsored a conference on creativity at Los Alamos.

ROTA: Look at the list of participants. It raises your eyebrows. You cannot bunch together creativity in one field and creativity in another. It's like matching producers of shoes with producers of meat loaf, because they're both producers. It is an error of logic. A friend of mine, a well-known painter, was looking at a copy of a painting by Velázquez. I watched her reactions. She started by saying, "How funny, this stroke is going down! Normally, we brush this way, but he brushed it that way." Then, "This is a combination of colors I've never seen," and so on. She said nothing about Velázquez's creativity. It is demoralizing to children to hold up Einstein or Beethoven as examples of creativity to be imitated. The idea of genius, elaborated by German romantics, is destructive; it is a flight into fantasy. There is reason to believe we've killed classical music because of that idea. People think that they will be either geniuses like Beethoven or nothing. But look at the Baroque Age—there were hundreds of little Italians who wrote good music and didn't give a hoot about being creative.

SCIENCE: Are there fashions in mathematics as there are in the arts?

SHARP: Today, mathematics is returning to the 19th century, to concrete computations, after seventy years of very abstract



Philosophers are needed today more than ever to tell the AI engineers some unpleasant truths.

mathematics. The latest fashion is 19th-century mathematics. Some of the best work in mathematical physics is based on a constructive approach.

SCIENCE: What does constructive mean?

SHARP: Instead of proving an abstract existence theorem, you produce an algorithm that delivers the solution. It's a powerful methodology.

ROTA: Concreteness is the word of the day. Now that we have learned to be abstract we can afford again to be concrete. Today's mathematics is more concrete than the mathematics of twenty or thirty years ago.

SHARP: In the fifties at Princeton you couldn't hold your head up if you weren't working on algebraic topology. Combinatorics had never been heard of, except possibly by a couple of dazed statisticians. Now combinatorics is a flourishing field. Gian-Carlo, what gave birth to the field of combinatorics?

ROTA: The time was ripe for it. Combinatorics is an honest subject. No adèles, no sigma-algebras. You count balls in a box, and you either have the right number or you haven't. You get the feeling that the result you have discovered is forever, because it's concrete. Other branches of mathematics are not so clear-cut. Functional analysis of infinite-dimensional spaces is never fully convincing; you don't get a feeling of having done an honest day's work. Don't get the wrong idea—combinatorics is not just putting balls into boxes. Counting finite sets can be a highbrow undertaking, with sophisticated techniques.

SCIENCE: What about the practical side of combinatorics? Is it as important for computer science as people are saying?

ROTA: The area of combinatorics that has found substantial application in computer science is the invention of efficient algorithms. Algorithms are instructions for performing a task. Even the fastest computers need good algorithms. A telephone company that succeeds in implementing an even slightly more efficient algorithm in, say, its switching network may well save a billion dollars. The payoffs are staggering. Today much effort in combinatorics is going into developing a theory of algorithms that will tell the optimal speed at which problems can be worked out. Some problems, unfortunately, cannot be worked out at any reasonable speed. We want to be able to tell when a problem can be solved efficiently, or whether it must be reformulated.

SCIENCE: What sort of problems?

ROTA: Take sorting. The combinatorial problem is the following. Given a sequence of numbers, how can you rearrange them in increasing order with a minimum number of transpositions? After years of research, we now know the most efficient algorithm for rearranging numbers in increasing order. We know that we have achieved the maximum speed.

SCIENCE: What kind of a proof can one give for something like that?

ROTA: The hard part comes after someone thinks he has found the fastest algorithm and wants to prove that it is the best possible. At the beginning someone will prove that N numbers can be rearranged at a speed, say, of N^2. Then someone else will modify the procedure and show that the same task can be performed with a speed of $N \log N$, and so on, until finally someone will cleverly prove that no one will ever do any better. That is the difficult part. In the theory of algorithms, one of the unfortunate turn of events was the discovery of NP-complete problems, namely, problems that can only be solved by algorithms that grow exponentially and therefore cannot be worked out on computers in a reasonable time. How to get around NP-complete problems is a frontier of combinatorics on which the best people are working.

SCIENCE: Working on what?

ROTA: On how to change an NP-complete problem into one for which a workable computer algorithm can be invented, by leaving out irrelevant cases or by taking a carefully chosen subset of the problem.

SHARP: Some problems are NP-complete only if one asks for the exact solution. If you are content with an accuracy of 2 or 3 percent, then an NP-complete problem may become tractable.

SCIENCE: Let's turn to highbrow combinatorics.

ROTA: Much combinatorics of our day came out of an extraordinary coincidence. Disparate problems in combinatorics, ranging from problems in statistical mechanics to the problem of coloring a map, seem to bear no common features. However, they do have at least one common feature: their solution can be reduced to the problem of finding the roots of some polynomial or analytic function. The minimum number of colors required to properly color a map is given

Research is not so much discovering something new as becoming aware of prejudices that stop us from seeing what is in front of us.

by the roots of a polynomial, called the chromatic polynomial; its value at N tells you in how many ways you can color the map with N colors. Similarly, the singularities of some complicated analytic function tell you the temperature at which a phase transition occurs in matter. The great insight, which is a long way from being understood, was to realize that the roots of the polynomials and analytic functions arising in a lot of combinatorial problems are the Betti numbers of certain surfaces related to the problem. Roughly speaking, the Betti numbers of a surface describe the number of different ways you can go around it. We are now trying to understand how this extraordinary coincidence comes about. If we do, we will have found a notable unification in mathematics.

SHARP: The ultimate motivation for these developments was the Riemann hypothesis, which remains unproved. When this hypothesis is proved, it will give the best information about the distribution of prime numbers. Remarkably, this information is also coded in the zeros of an analytic function. The Weil conjectures set up an analogous function to the Riemann zeta function for a simpler case.

SCIENCE: Gian-Carlo, tell us your contribution to combinatorics.

ROTA: The one contribution of mine that I hope will be remembered has consisted in just pointing out that all sorts of problems of combinatorics can be viewed as problems of location of the zeros of certain polynomials and in giving these zeros a combinatorial interpretation. This is now called the critical problem. Over the years people have added examples of critical problems, though we still haven't gotten any idea of how to solve it in general. I'd like to see someone get such an idea before I die. The four-color conjecture—that with only four colors you can color every planar map so that no two adjacent regions have the same color—is one of these critical problems.

SHARP: I thought that had been settled by a computer proof.

ROTA: Not really. What we want is a rational proof. It doesn't help to have a brutally numerical answer spewed out by a computer. A problem is interesting only when it leads to ideas; nobody solves problems for their own sake, not even chess problems. You solve a problem because you know that by solving the problem you may be led to see new ideas that will be of independent interest. A mathematical proof should not only be correct, but insightful. Although, as Erdös says, nobody gets blamed if his first proof is messy.

SCIENCE: Will all these abstract ravings have some impact on how we view the world?

ROTA: It's a domino effect. You start with an abstract idea, and pretty soon it turns our world upside down. Leonardo da Vinci said, "Theory is the captain and application is the soldier." That's the practical side. If we take a deeper look, we see that nature imitates mathematics.

SHARP: Mathematicians look for relationships between fields of mathematics that hitherto were thought to be unrelated.

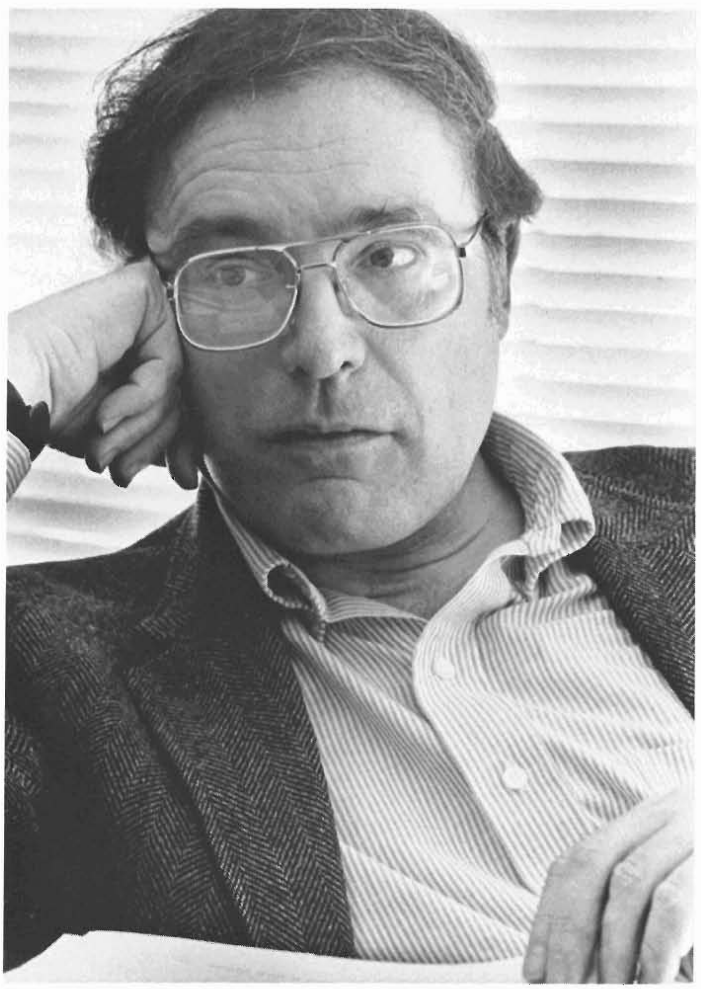
ROTA: Mathematics is the study of analogies between analogies. All science is. Scientists always want to show that things that don't look alike are really the same. That's one of their innermost Freudian motivations. In fact, that's what we mean by understanding.

SHARP: You often hear that the purpose of a scientific theory is to predict. That's not correct. The purpose is understanding. Prediction is one way to test whether our understanding is correct. Simplicity, scope, and beauty are as important as prediction in judging whether a theory leads to understanding.

ROTA: May I phrase what you just said in philosophical terms? Science turns paradoxes into trivialities.

SHARP: Gian-Carlo, there's interest, activity and, most of all, talk about artificial intelligence today. You've been following these developments. What is your candid opinion of AI?

ROTA: There is an old *New Yorker* cartoon of a wine factory in California. The director says to some visitors to the factory, "We



Any description of vision that omits the function of purpose and expectation will be ineffective for purposes of AI.

used to employ humans to press wine, but now we use up-to-date methods." In the background, you see a huge mechanical foot going up and down. To my mind that's the present state of the art in AI.

SHARP: Remind us of the objectives of AI.

ROTA: The accepted objective of AI is to build machines that perform tasks now performed by humans. Do not confuse AI with the more conservative computer science. Computer science and artificial intelligence are now distinct fields, and practitioners of either are at loggerheads with practitioners of the other.

SHARP: So AI research attempts to simulate human behavior. What progress has been made?

ROTA: We had to swallow a bitter pill and realize that tasks that were at first thought to be easy to simulate by computer turned out to be hard. Whenever a large memory is required, a computer will perform better than a human being, but any task that requires the slightest bit of recognition is awfully hard for a computer. Try to build a computer that recognizes when a man is wearing a uniform. Cognitive psychologists, computer scientists, specialists in artificial intelligence, and neurophysiologists are now figuring out a new breakdown of the basic components of human intelligence. Their research is beginning to reveal that the basic talents are not at all what we always thought they ought to be. The building blocks of the process of perception are turning out to be completely unexpected. Their discovery is bringing about an enormous advance in our thinking, such as has not happened since Plato.

SHARP: The enemy, here as elsewhere, is wishful thinking; it is the strangling power of prejudice. Much research in AI today consists in making scientists aware of unverbalized prejudices about thinking and speaking processes.

SCIENCE: Can you comment about the expert systems?

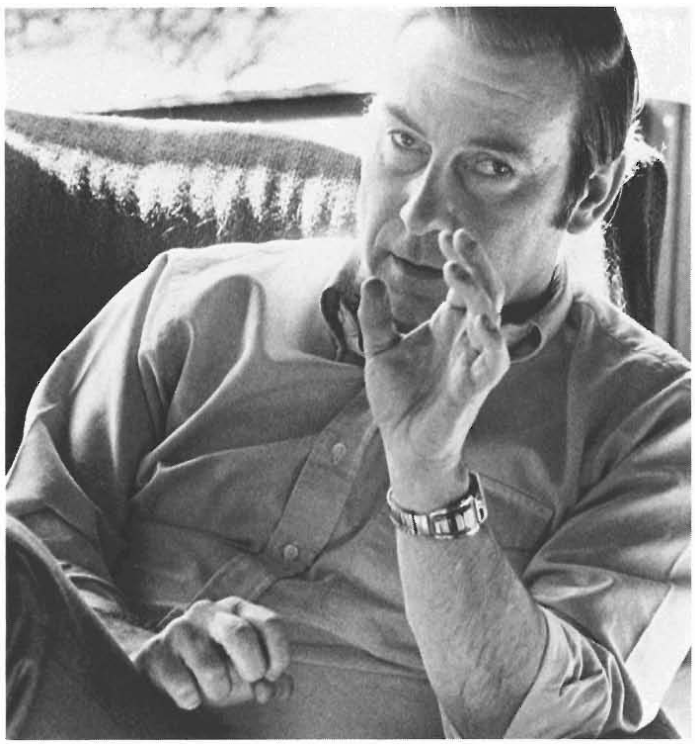
SHARP: The excessive claims made about these systems are creating a delusion.

SCIENCE: What claims?

SHARP: For example, the claim that expert systems will do away with physicians because computers will better diagnose disease.

ROTA: The first expert system to gain wide acceptance, in my opinion, will be the one designed for lawyers to look up cases. An expert system is basically the exploitation of the concept of a questionnaire, brought to its ultimate and gory conclusion by the computer.

SHARP: Let's consider how an expert system for medical diagnosis is constructed. First you sit down with a bunch of doctors and get them to tell you how they arrive at their diagnoses. You find that the better the doctor, the more difficult it is to characterize his diagnoses with an algorithm. But a basic element of their trade is collecting symptoms and looking up the possible diseases associated with those symptoms. If the symptoms don't uniquely pinpoint one disease, the next step is to get more information from the patient. The combinatorial model for this process is a tree-search algorithm supplemented by a set of rules for producing if-then statements. Although an expert system constructed in this way works pretty well, it has a built-in flaw because a patient can seldom cough up a complete and accurate list of his symptoms. A good diagnostician has a feeling for a patient's condition that is difficult to implement algorithmically.



SCIENCE: Is that what you call context?

SHARP: Yes, human beings have the ability to make use of context-dependent features. We do not know at present how to program context-dependent behavior in computers. The issue of context came up with a vengeance in computer programs for chess. In the fifties computer scientists thought that they could write powerful chess-playing programs. They coded text-book openings, and they wrote programs that would look ahead a couple of moves. The results were mediocre. The programs never learned from experience, and once a human opponent found out their weaknesses, he could consistently beat the computer. Now, faster computers look six or seven moves ahead, and programs play almost at the master level. But they still don't learn, and a very good player can eventually beat them. The issue here is context. A chess player has a feel for a strong position. This contextual feel has not been implemented in chess programs.

SCIENCE: Might it be that the algorithm will be too complicated?

SHARP: Worse. We don't know the principles. In physics we know the principles, and therefore we can write down equations that describe complicated situations, sometimes even too complicated to compute. For instance, we believe that the Navier-Stokes equations

The element of anticipation is essential to all grasp of meaning.

describe turbulence, even though we're not able to compute turbulent flow. In the problem of context-dependence, we are still missing a clear formulation of the basic principles.

ROTA: That's where philosophy comes in.

SHARP: That's radical!

SCIENCE: What's radical?

SHARP: To say that philosophy should become an active partner in the creation of the new science of artificial intelligence. Usually philosophy is a Johnny-come-lately that gives perspective on a developed science. It is usually not a leader.

ROTA: Philosophers are needed today more than ever to tell the AI engineers some unpleasant truths. The philosopher's role has always been that of stating facts that may have been on everybody's mind but that no one dared state clearly. Eventually, engineers will reluctantly acknowledge that what the philosopher says is the truth, but they will then get rid of the philosopher.

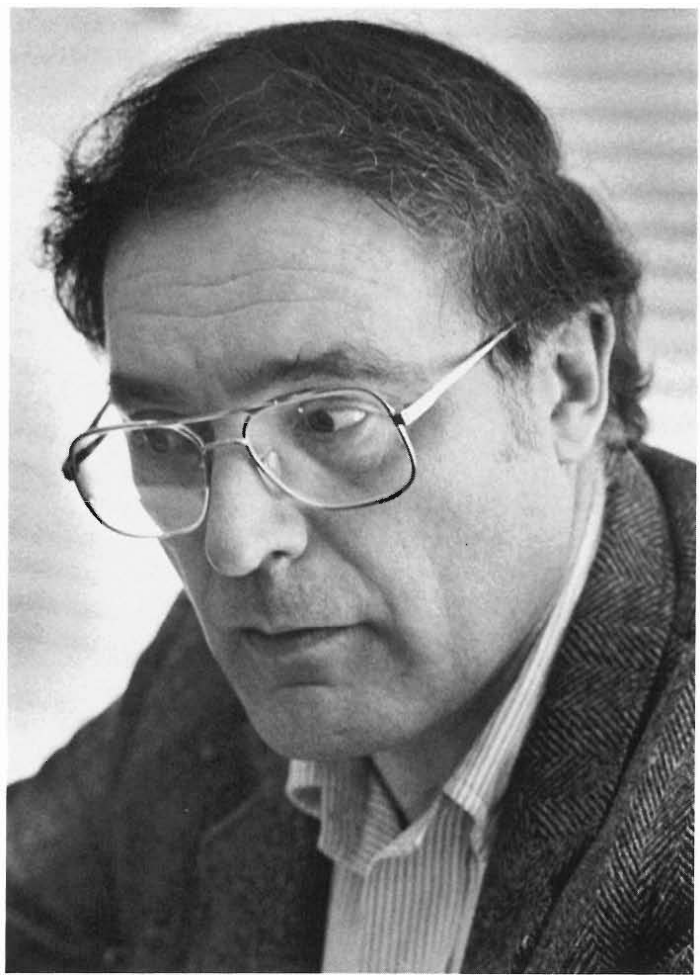
Let me give you an example. In the early fifties engineers attempted to build a machine that translated from one language to another. They fell flat on their faces because they had unclear ideas on language. A good philosopher would have said, "You must begin by realizing that language is not what you thought it to be. You must bring out your unverbalized prejudices and observe language objectively without a screen of preconceived ideas on what it ought to be like." Research is sometimes not so much discovering something new as becoming aware of prejudices that stop us from seeing what is in front of us. For example, a naive view of words states that, by and large, they have a fixed meaning. Contemporary philosophy stresses instead the variety of possible contextual senses. The problem of meaning is the problem of describing the nature of the interaction between the inherited meaning of a word and the variable contextual senses it may have. For example, when someone utters a sentence, you can understand it, because of your anticipation of what comes next. This element of anticipation is essential in all grasp of meaning. It's easy to write poems about it, but try to write down the formal rules! This is precisely the task contemporary cognitive philosophy has set itself.

SHARP: The beginnings of this formalization can be traced to Chomsky and his context-free grammars. The almost mechanical rules of these grammars turn out to capture more of the structure of natural language than anyone thought possible. Formal grammars were effective for the inventing of computer languages. Eventually, they were enriched with context-sensitive grammars. But language depends on context in a way that we don't know how to express with formal grammars. We need a new idea that is yet to come.

ROTA: The problem of context-dependence is not limited to language. It has to do with sense-making generally. Consider the following example. You are at the airport. A gate opens and people come out. What are they doing? They are arriving. Think of a robot that can tell the act of walking from the act of arriving. A child can tell, but not a machine. The difficulty is that the act of arriving is purely context-dependent. The same difficulty occurs in all AI problems. Here are a few other examples that display the difficulty of formalizing context-dependence. How can you tell birds flying from birds migrating? Or someone goes to Oxford, visits the colleges, is shown the classes, and has dinner at High Table. Then at the end of the day, he asks his guide, "I've seen all these wonderful things, but where is the university?" Why is it difficult to tell where the university is? Because a university is a contextual construct. Phenomena that we believed to be physical are revealed to be actually contextual.

SCIENCE: Are you saying that all sense is context-dependent?

SHARP: Meaning is inextricably context-dependent. Context-dependence displays a variety of different layers of description that seem dependent on each other, and yet cannot be reduced to one another. For example, one can give a detailed neurophysiological description of the brain, but that does not describe the mind, although the mind depends on the brain's physiology to function. The reductionist mistake is to think that context A is reducible to context



When you search deeply into . . . the Western mind, you discover the craving that all things should be reduced to one.

B, just because A cannot exist without B. Here is a naive example of a reductionist explanation. Johnny, having been asked to explain his presence in the girls' dormitory after curfew, answered, "I obeyed Newton's equations of motion." A neurophysiological description of the mind is a similar reductionist mistake.

ROTA: The laws of biology ultimately depend on the laws of physics. But from a knowledge of the laws of physics, you cannot infer the behavior of living organisms. Let me give a simple example of context-dependence, due to Ulam. You want to define a key. No amount of staring will be able to identify this object as a key unless you already have an unverbalized grasp of the contextual function of keys.

SHARP: No molecular analysis of this piece of metal will lead you to grasp the contextual function of keys and locks.

SCIENCE: Can you break down the notion of context into its fundamental elements?

ROTA: The act of perceiving a key requires an implicit background and an implicit foreground. These two elements are part and parcel of any perception. If you want to build a machine that sees, then this machine must have, programmed into it, some sense of purpose. Every act of seeing is inextricably bound with an unexpressed foreground of anticipation.

SHARP: We recognize visual scenes by virtue of our expectations of seeing certain events depending on particular contexts. The notion of expectation is the most fundamental in solving the problems of pattern recognition. If we are to build machines that see, we must find a way to encode in a machine the expectation of certain patterns. Without built-in expectation, there is no recognition. Any description of vision that omits the function of purpose and expectation will be ineffective for purposes of AI.

SCIENCE: That seems like a very tall order—to build a machine that has expectations!

ROTA: It hasn't been done, but it may not be as mysterious as it sounds. One cannot perceive anything, whether it is seeing the blue sky or grasping another person's mood, outside of a context where what is perceived plays a role as part of an organized project. What you perceive when you perceive is a function, not a thing. Being a key, a cup, a book are functions, not things. You may think that this is obvious, but people who are trying to build machines that see are still to some extent wedded to the old theory of perception, the one that pretended that the act of perceiving an object, a key for instance, was some sort of comparison with a little key inside the brain. The mortal blow was dealt to this simplistic theory by AI. As soon as scientists conceived the project of building a machine that perceives, they realized that the little-picture-in-the-brain theory did not make sense, and they became aware that an essential component of all perception is an act of choice that must to some extent remain arbitrary. Such a choice is not determined by physical data, but by your expectations of what the function you are perceiving is meant for.

SHARP: What you are saying is that all sense is functional, and

function is the interplay among expectation, physical presence, and purposefulness.

SCIENCE: How does this philosophical sermon affect research in AI?

ROTA: Let's take a simple example, the act of following a rule. If you reflect on the word "following," with the idea of simulating it on a machine, you realize that you do not know the functional sense of "follow." You try various thought experiments: A follows B in a dance, or a blind man touches someone who sees and follows him. Do these examples lead us to an understanding of following in general? We are not able to pin down the meaning of the word "following" in a general framework. This problem keeps coming up in AI. AI is relying upon the received meanings of words. This is valueless for scientific purposes, as if we were trying to discover chemistry by taking earth, air, fire, and water as our basic concepts. Nowadays, the need to write computer programs that work forces us to perform investigations in philosophy that philosophers were formerly loath to do. For example, I now read the word "horse" on this page. What's really there is some ink. The word "horse" is, strictly speaking, not on the page. It's intentional. How is it that I decide to see the word "horse" rather than some black dots on a white sheet? At present, we are barely beginning to understand, thanks to the possibility of computer simulation that forces us to face these philosophical puzzles.

SCIENCE: What do you mean by "intentional," and how has the computer helped you understand this notion?

ROTA: Intentionality is the key word used to denote the new theory of perception. When you look at a printed page and see the word "horse" instead of seeing a collection of meaningless ink marks, you are selecting one function of what you see against another. This selectivity is the basic component of perception. It is now realized

Einstein was the last genius of oneness.

that intentionality is the notion to be formalized in AI. The practical needs of AI act as a powerful stimulus to this formalization.

SCIENCE: It looks like AI is helping us replace the old philosophical subject/object question with the notion of intentionality. So what?

SHARP: You see, now that we begin to see the importance of intentionality, we will analyze it in limited and rigorous environments. Computer programs performing limited cognitive functions will help us to make a science out of AI. Progress in science relies on using simple cases to test basic principles of wide applicability. To give you an analogy, solid-state physics is based on Coulomb's law of electrical interaction between charges. Coulomb's law itself is simple, but the law manifests itself in condensed matter in extraordinarily complicated ways. However, Coulomb's law operates in a simple fashion in the hydrogen atom. Similarly, by considering limited instances of perceptual problems, by developing computer programs that can solve simple problems, we are beginning to have examples that will eventually reveal the basic principles of perception. We can now test our ideas on relatively simple computer examples without grappling with the full complexity of neurobiology or human experience.

If you want to build an airplane, it might seem strange to begin with the study of the flight of birds. Boeing Aircraft won't gain much from numerical studies of feathers. Nevertheless, there are ideas common both to the flight of birds and to the flight of aircraft. These common ideas are the laws of aerodynamics. Similarly, finding the common laws for both human and machine perception is an exciting task of our time for which neurophysiology and AI are supporting each other.

ROTA: In both contexts the problem is to understand how sense-making arises out of staring at some physical data that, taken in isolation, are devoid of any sense. This is the major problem of philosophy. How does this thing called sense happen? How can you make sense while physically looking at an object or hearing a spoken word?

SHARP: In formal systems we haven't found a set of syntactic rules strong enough to do away with the need of a previous tacit knowledge of semantic sense.

ROTA: Dave is putting his finger on the problem of the inadequacy of formal logic to deal with contextual matters. Present-day logic cannot even distinguish between "and" and "but." Present-day logic claims that "and" and "but" have the same meaning, but anybody who talks uses "and" and "but" in a quite different sense.

SHARP: The actual meaning of "and" and "but" is not captured at present by their syntactic description.

ROTA: The notion of axiomatic description that we have today is inadequate to render this difference. All axiomatic descriptions of "but" have failed, not to speak of those of "nevertheless."

SHARP: Or "meanwhile."

ROTA: The language we speak is at odds with logic. It used to be thought that formal logic is a rendering of our reasoning process, but after the AI experience we have realized that formal logic is just another flight from reality.

SHARP: Formalization hasn't made a dent in our understanding of context. Someday maybe, the notion of context will be incorporated in a greatly extended logic that will delve deeper into the foundations of our thinking.

ROTA: Our ideas of foundations are changing. What we mean by foundations today is quite different from what Hilbert or Russell or Whitehead meant. Old Bertrand Russell would take it for granted that there is such a thing as foundations in mathematics and that such foundations are needed. "Mathematics shall be derived from its foundations," ruled Bertie; it must follow from a simple and consistent set of axioms.

SHARP: That's an idea with a long and tortuous history. Such a simple set of axioms has been found for simple theories like the first-order predicate calculus but not for richer branches of mathematics, such as set theory.

ROTA: There is an old anecdote about Frank Ramsey asking Wittgenstein in Cambridge sometime in the late twenties, "Look at the bridge over there. Now suppose a contradiction were to be found in set theory. Do you think that bridge would fall?"

SHARP: Gödel's incompleteness theorem did away with a lot of those pretensions of logicians, by showing that in some systems there are true statements that can neither be proven nor disproven. This was a blow to Hilbert's vision of mathematics following from a

Truth is not to be held onto like a teddy bear.

simple set of consistent axioms. Gödel's theorem is upsetting because we want definitive truth, with a capital T.

ROTA: We must first look for the unstated wishes that we have in the back of our minds when we ask for foundations of mathematics or science. When you search deeply into the cravings of the Western mind, you discover the craving that all things should be reduced to one, that the laws of nature should all be consequences of one simple one, that all principles should eventually be reduced to one principle. It's a great Jewish idea. One God, one this, one that, one everything. We want foundations because we want oneness.

SHARP: Isn't this another instance of reductionism?

ROTA: This craving for reducing all physical laws to one law may be a delusion. Einstein was the last genius of oneness. Maybe he is right. But more probably, we will have to get used to several sets of laws of nature, existing together and irreducible to one another. The laws that describe living systems, if any, will not be reduced to the laws of physics or to the laws of cognitive behavior, if any.

SHARP: Physics is based on the paradigm that one analyzes physical processes in terms of the concepts of space, time, energy, velocity, and so on, not by metaphysical intervention. In the 17th century the understanding of what were the right questions to ask caused a fundamental reorientation in the way we looked at the world. Nowadays, in artificial intelligence, one needs a similar reorientation, one that points to the right questions to ask. The questions will be suggested by an enlightened analysis of behavior.

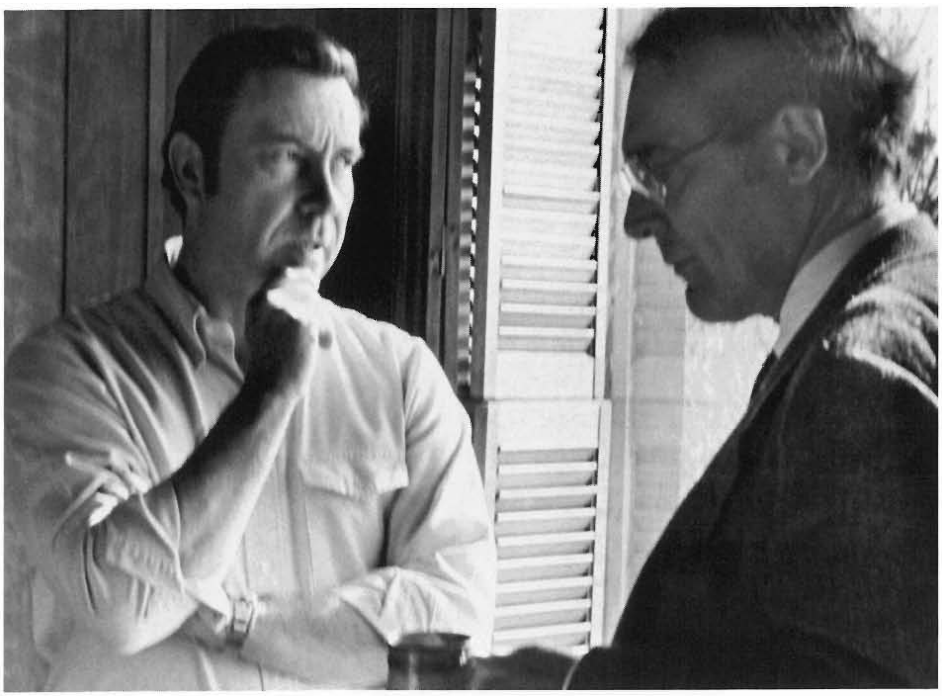
SCIENCE: How would you summarize the direction of present-day philosophy in relation to AI?

ROTA: We need to understand how sense-making arises out of staring at physical data that by themselves are meaningless. How can you get sense by merely looking? The philosopher's role is to tell the AI specialist that he doesn't know what he's talking about, to put it bluntly.

SHARP: From 1650 well into the 20th century there was a hue and cry about individual sciences dropping out of philosophy and becoming independent—first physics, then biology, and now AI. Some of these sciences might well gain from an excursion on the shores of philosophy.

ROTA: There is danger that a new profession will come into existence, the AI specialist. People who work with computers will have to hire an AI expert, like they keep a physician, a lawyer. It will be a self-perpetuating profession lacking an adequate scientific base. These people might come to control a great deal of power under the aegis of the computer. Beware!

SHARP: You see, that's what philosophers are for. ■



Gian-Carlo Rota *is Professor of Applied Mathematics and Philosophy at the Massachusetts Institute of Technology and Director's Office Fellow. He has been a consultant to Los Alamos for seventeen years.*

David H. Sharp*, a Fellow in the Laboratory's Theoretical Division, is known for his work in quantum field theory, hydrodynamics, and the theory of gravitation.*

Photos by LeRoy N. Sanchez