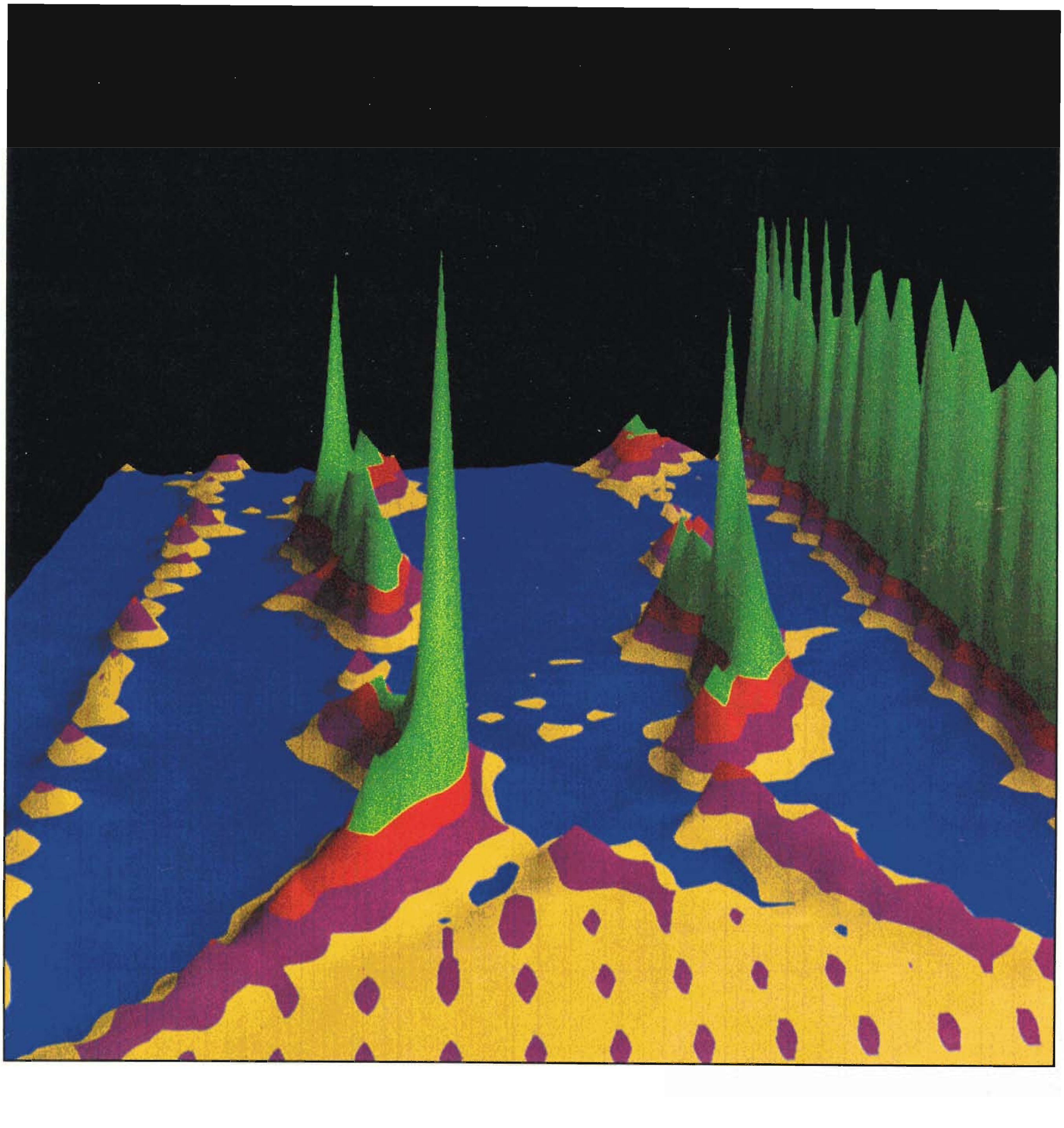
NUMBER 10 SPRING 1984

Los Alamos Science

LOS ALAMOS NATIONAL LABORATORY



EDITOR'S NOTE

The issue contains stories of three fascinating and novel explorations; first the exploration of a new idea in bioenergetics that may explain your body's remarkably efficient use of energy, second the exploration by satellite of the distant reaches of the earth's magnetosphere, and third, the exploration of surprising behavior in a tiny metallic crystal.

The new idea in bioenergetics involves solitons, an unusual kind of traveling wave. Solitons don't change their shape and they don't change their speed—even when they run into each other. They continue to pop up in one area of physics after another and now our cover article reports that these pulse-like nonlinear waves may be running around in many parts of our bodies, carrying signals and coordinating processes through membranes and across cells. This novel idea, first proposed ten years ago by the Soviet physicist Davydov, is being explored in this country by Al Scott, chairman of the Center for Nonlinear Studies, along with a number of other people at Los Alamos. In a lucid article Peter Lomdahl, Scott Layne, and Irving Bigio describe specific places in the cell where solitons might be important, outline the mathematical model and numerical results that show how energy can be trapped by proteins and self-focused into stable solitary waves, and, finally, evaluate the experimental evidence supporting the idea. A possible extension of this model, presented by Layne, an M.D., is the idea that anesthetics may work by binding to proteins and thereby interfering with soliton propagation.

Physicists agree that solitons can appear in many different contexts: in ordinary fluids, in plasmas, in lasers, in Josephson junctions, in organic polymers—technically speaking in all systems that are nonlinear and dispersive and contain little or no damping. But their existence in biological systems is still quite speculative. Of course speculation can be dangerous, but even if the Davydov model is wrong, this work still represents the beginning of a new approach to understanding macromolecules, an approach that invokes nonlinear dynamics to bridge the gap between structure and function.

The ISEE-3 spacecraft has completed its work collecting data in the earth's magnetic tail, a long tail of tenuous plasma much like the tail of a comet. The data collected during repeated crossings of the tail at distances from 400,000 to 1,500,000 kilometers from the earth have provided a remarkably detailed picture of the interaction between the solar wind and the magnetosphere, which serves as the earth's magnetic shield. After months of careful analysis, Jack Gosling, Dan Baker, and Ed Hones are able to show us the fascinating way in which ISEE-3 data demonstrate the existence of an open magnetosphere, one that is constantly being infused with plasma from the solar wind. Of course at some point plasma must also be ejected, and again the data clearly demonstrate how. The authors' comparison of ground data, data from a near earth orbit, and data from the distant tail reveals that magnetic substorms observable on the earth are local manifestations of global phenomena in which magnetic field lines reconnect in the geomagnetic tail and pinch off huge regions of plasma that go hurtling off into space. The sojourn of ISEE-3 to the deep magnetic tail was made possible by unique three-body orbits involving the earth, the moon, and the spacecraft. Now ISEE-3 has changed its name to ICE and is off chasing comets.

In a blow-by-blow description—one that captures the excitement of scientists hot on the trail of discovery—Greg Stewart, Zachary Fisk, Jeff Willis, and Jim Smith tell us how they stumbled, with skill, onto what might be the first p-state superconducting metal. This compound of uranium and platinum, UPt_3, is a nearly magnetic material, one that exhibits substantial magnetic spin fluctuations; it therefore cannot be a superconductor of the ordinary s-state variety. In a sidebar David Pines, eminent physicist, Laboratory consultant, and expert in condensed matter physics presents a cogent interpretation of the observed phenomena and its significance for research in superconductivity. Also noteworthy is a sidebar in which expert crystal grower Zachary Fisk outlines the method he used to prepare single crystals of UPt_3.

Al Scott was among the original group of scientists who, in the late fifties and early sixties, followed their intellects into the unfashionable field of nonlinear science. He is obviously in love with this subject and in his interview in this issue is able to convey to the uninitiated its elusive concepts and the revolutionary impact they are having in science. Not only are nonlinear phenomena lurking almost everywhere, but new mathematics has made it fun and productive to think about them. ■

Necia Grant Cooper

On the cover.

A computer graphic showing the formation of a self-focused soliton-like excitation in the organic polymer acetanilide. This polymer, which resembles a protein, contains parallel chains of hydrogen-bonded peptide groups. The graphic depicts the time evolution of energy on one of these chains. Time increases from foreground to background. Energy increases from blue to yellow, purple, red, and green. Initially (at the base of the figure) energy is distributed uniformly along the whole chain (yellow). As time goes on, the energy self-focuses at a few sites along the chain (green peaks). Where energy is low on the chain shown it is high on the other chains. This graphic displays results from GLOP, a computer code developed by Peter Lomdahl to calculate nonlinear dynamics for any protein geometry. The Davydov model on which the code is based is discussed in "Solitons in Biology" by Lomdahl, Layne, and Bigio. The graphic was prepared by Peter Lomdahl using Melvin Prueitt's graphics program CAMERA.

Los Alamos Science

SPRING 1984 NUMBER 10

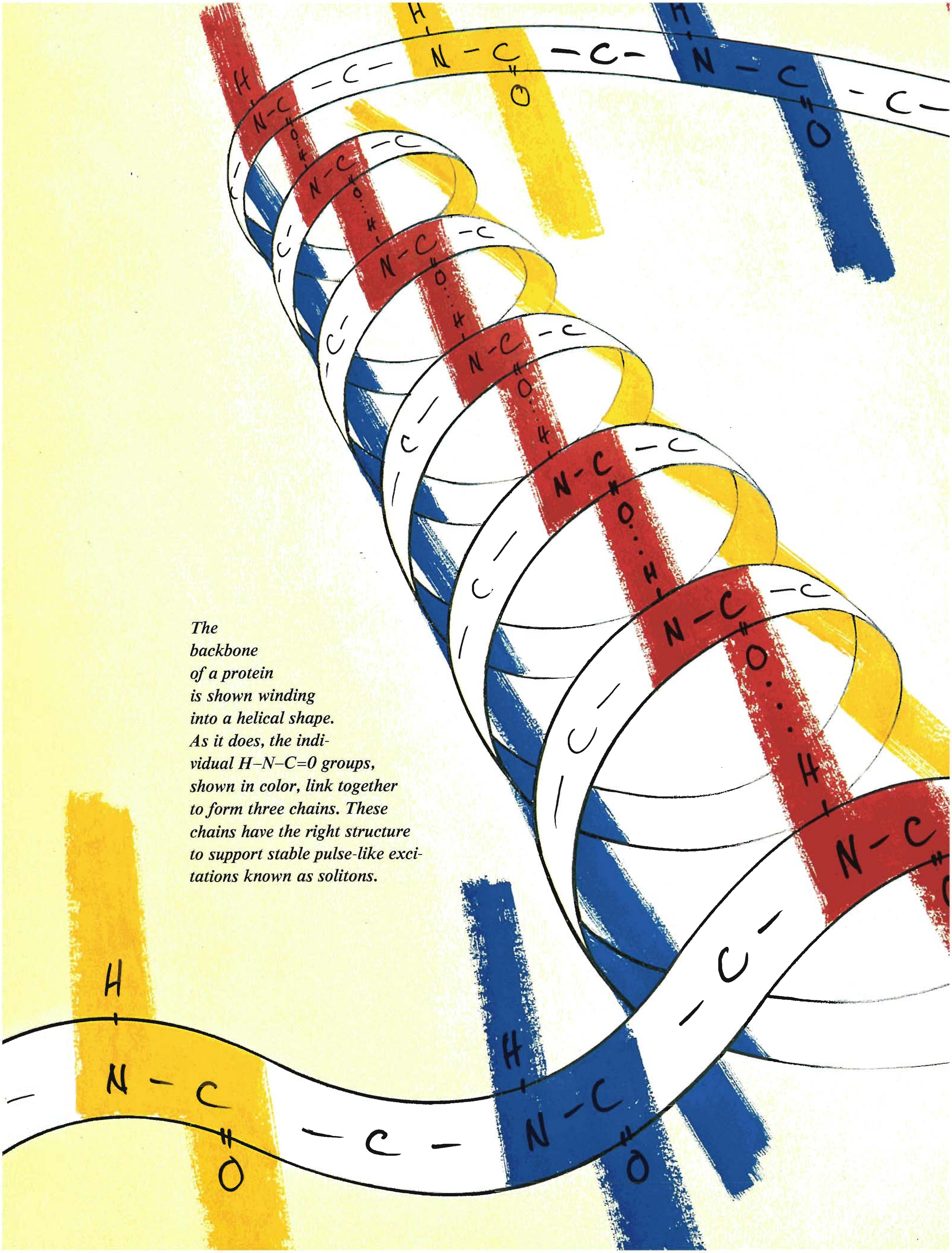
CONTENTS

Los Alamos Science is published by Los Alamos National Laboratory, an Equal Opportunity Employee operated by the University of California for the United States Department of Energy under contract W-7405-ENG-36.

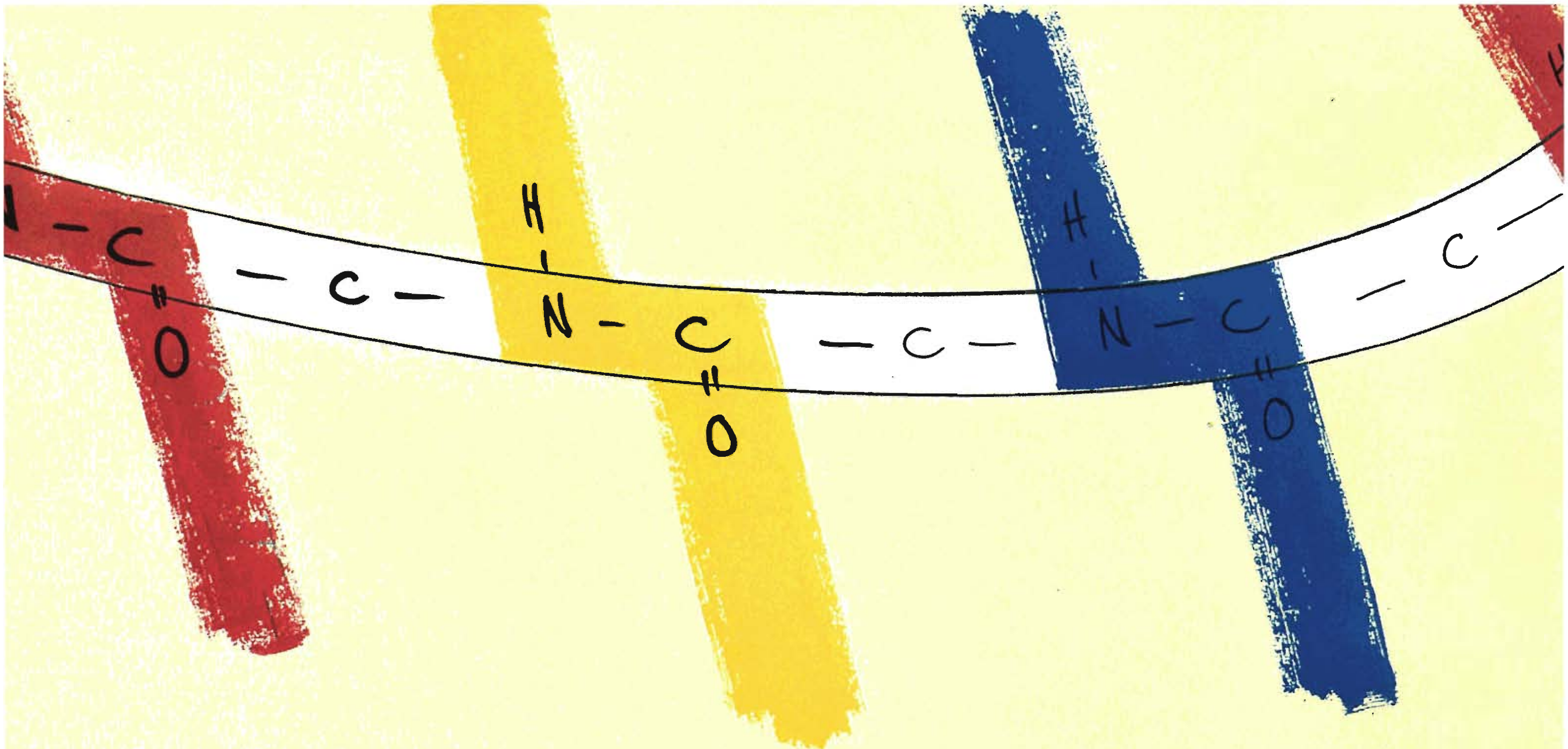
Address mail to
Los Alamos Science
Los Alamos National Laboratory
Mail Stop M708
Los Alamos, New Mexico 87545

RESEARCH AND REVIEW

PEOPLE

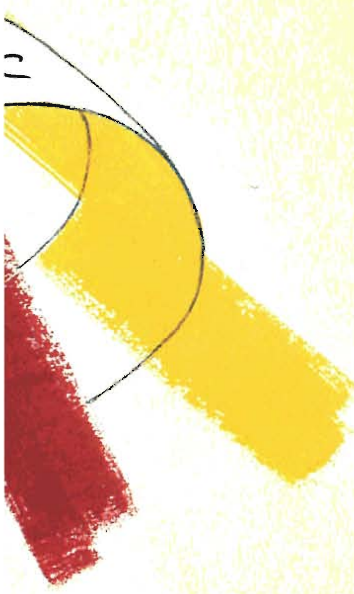


The backbone of a protein is shown winding into a helical shape. As it does, the individual H–N–C=0 groups, shown in color, link together to form three chains. These chains have the right structure to support stable pulse-like excitations known as solitons.



Solitons in Biology

by Peter S. Lomdahl, Scott P. Layne, and Irving J. Bigio



In 1973 scientists gathered at the New York Academy of Sciences to discuss an unanswered question in bioenergetics: How is chemical energy transduced and transported in biological systems? In the same year a Soviet solid-state physicist proposed a dynamic answer that was totally novel to the world of biology. Exploiting the regularity in the structure of α-helical proteins, he showed that simplified models of these proteins could self-focus, or trap, energy in stable, pulse-like waves known as solitons. If self-focusing is indeed a biological reality, it may account for many aspects of protein behavior, including the efficient transport of energy. This possibility is being studied at Los Alamos through analytical, numerical, and experimental techniques.

It is widely accepted that proteins are the principal workhorses of the cell. They are the major organizers and manipulators of biological energy and the enzymes that catalyze and maintain the life process. They are responsible for the active transport of ions into and out of the cell and for cellular and intracellular movement. Of course other macromolecules, such as DNA, polysaccharides, and lipids, have an energetic dimension, but their operation is always closely tied to that of proteins. Therefore the discipline of bioenergetics, which is the study of how cells generate and transfer their energy supply, is primarily the investigation of how proteins work. From decades of chemical analysis and x-ray crystallography and from more recent advances in spectroscopy, we know the composition and three-dimensional conformation of about two hundred proteins. Despite this extensive structural knowledge, however, there is no generally accepted model of how proteins operate dynamically. Presently, it is fair to say that the "nuts and bolts" functioning of proteins remains an outstanding question in bioenergetics.

The energy supply for most protein activities is provided by the hydrolysis of ATP (adenosine triphosphate). An ATP molecule binds to a specific site on the protein, reacts with water, and under normal physiological conditions releases 0.49 electron volt (eV) of free energy. This is about twenty times greater than the average energy available from the thermal background at 300 kelvins. The question for bioenergetics is what happens to this energy? How does it perform useful work? Is the energy used through a nonequilibrium process, or does the energy first thermalize and then work through an equilibrium process? Molecular dynamics calculations, based on ball-and-spring models of proteins, show that heat from a thermal bath induces a variety of motions in proteins. These equilibrium calculations show motions ranging from localized, high-frequency vibrations of individual bonds to collective, low-frequency motions of the entire protein. One may question, however, whether such equilibrium dynamics could account for the efficient transport and use of energy over the characteristic lengths of proteins, which range from tens to hundreds of angstroms.

An alternative hypothesis is that the energy of ATP hydrolysis is converted through resonant coupling to a particular vibrational excitation within the protein. This coupling might proceed through an intermediate vibrational excitation of water. Figure 1 shows a likely recipient in such a resonant exchange, the amide-I vibration. This vibration is primarily a stretching and contraction of the carbon-oxygen double bonds in the peptide groups of the protein (see "The Structure of Proteins"). The energy of the amide-I vibration is about 0.21 eV, which corresponds to about 1660 reciprocal centimeters (cm^{-1}).* This energy is a little less than half the energy of ATP hydrolysis and is almost equal to the energy of the H–O–H bending mode of water at about 1646 cm^{-1}. The amide-I vibration is a prominent feature in the infrared absorption and Raman spectra of proteins. Moreover, its energy remains almost constant from one protein conformation to another, indicating that it is rather isolated from other degrees of freedom. All these factors lead to the conjecture that the energy released by ATP hydrolysis might stay localized and stored in the amide-I vibration.

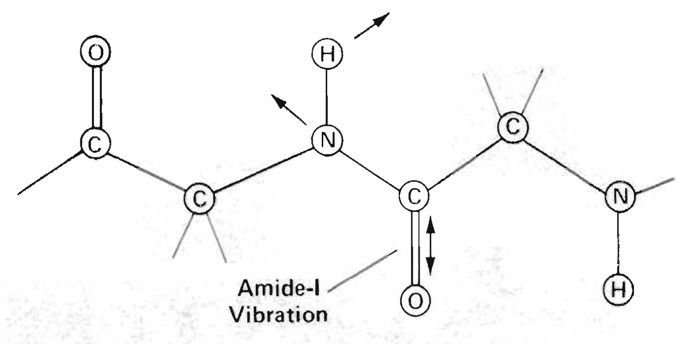


Fig. 1. Peptide group showing amide-I resonance. The amide-I resonance, which is intrinsic to every peptide bond of every protein, is one of the strongest and most characteristic spectral features of a protein. Its energy is nearly invariant from one protein conformation to another (1645 to 1660 cm^{-1} for an α helix, 1665 to 1680 cm^{-1} for a β sheet, and 1660 to 1665 cm^{-1} for a random coil). The major contribution to the amide-I resonance comes from the stretching vibration of the C=O bond, although relatively small contributions come from both the C–N in-plane stretching and N–H in-plane bending vibrations.

When a similar idea was discussed at the 1973 meeting of the New York Academy of Sciences, the objection was raised that the lifetimes of typical vibrational excitations in complex biological molecules are too short (10^{-12} second) for them to be important in the storage and transfer of biological energy. In particular, peptide groups have large electric dipole moments; therefore, dipole-dipole interactions among peptide groups would cause the amide-I vibrational energy to spread to neighboring peptide groups. Thus the energy would not remain localized but instead would disperse throughout the protein and be lost as a source for biological processes.

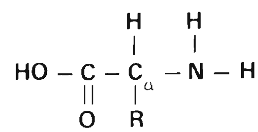
The Soviet physicist A. S. Davydov countered this objection with an argument from nonlinear physics. He suggested that the energy of ATP hydrolysis can be stored in the amide-I vibration through a nonlinear interaction that self-focuses, or traps, the energy in a soliton

**In spectroscopy one often uses the wave number* $1/\lambda = \omega/2\pi c = E/hc$, *instead of the energy* E *or frequency* ω, *to characterize vibrational states since the typical numbers are more palatable.*

Structure of Proteins

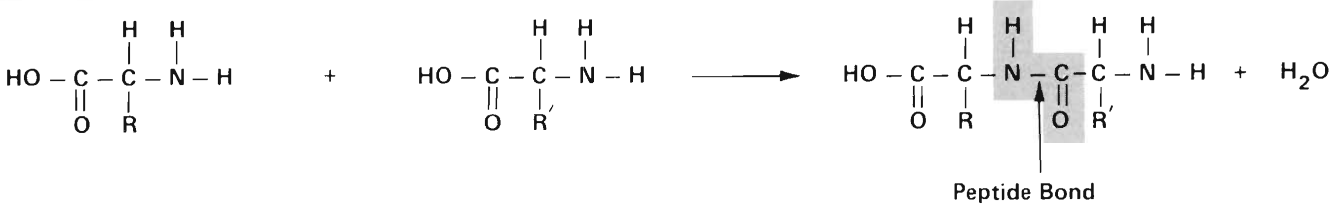
Proteins are constructed from individual building blocks called amino acids. In all about twenty different amino acids are commonly found in proteins, and about five others are found rarely. Each amino acid has an amino group (NH_2), a carboxyl group (COOH), and a side group, or radical (R), attached to the alpha carbon atom. It is the radical that distinguishes one amino acid from another.

Amino Acid



R = Side Chain on Alpha Carbon

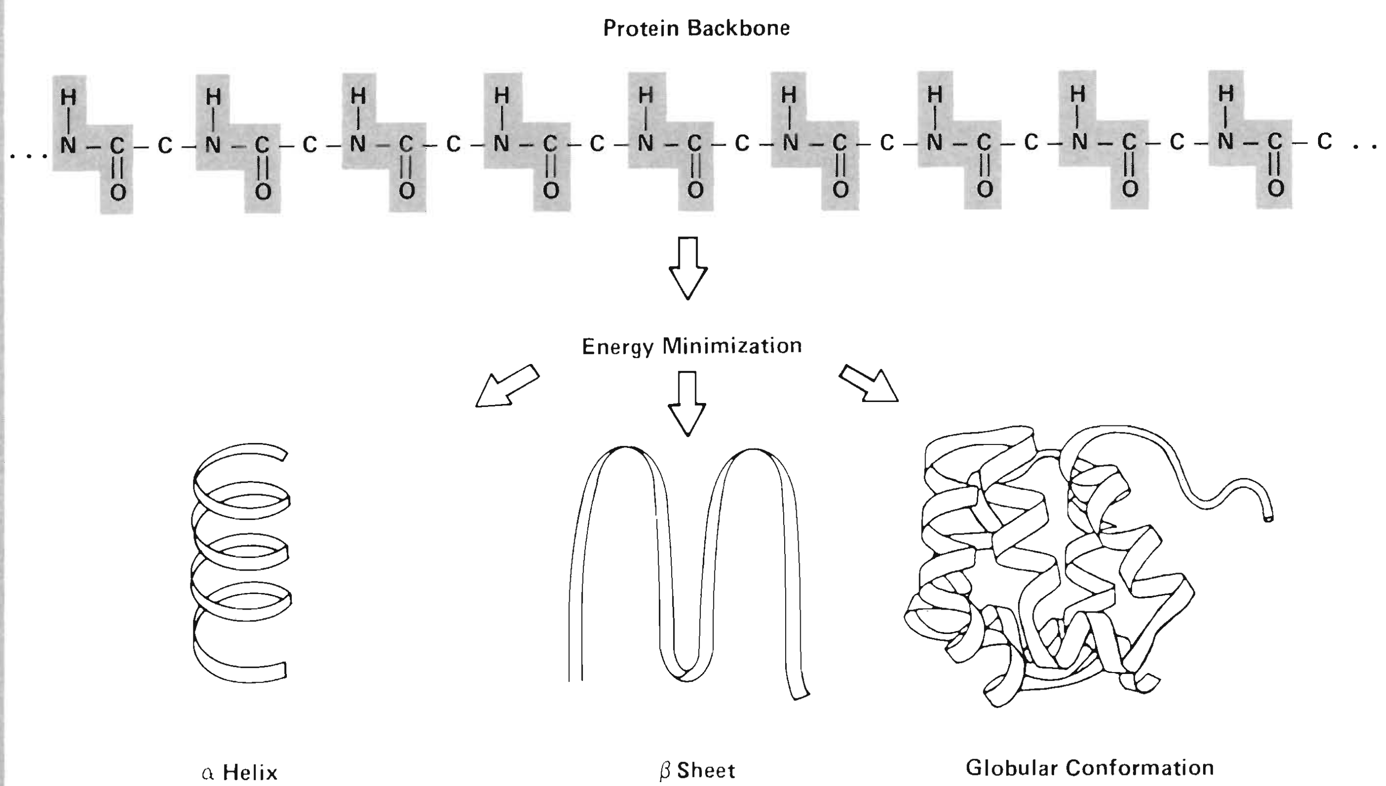
Amino acids polymerize to form long chains of residues that constitute a protein. When two amino acids join together, they liberate one molecule of water and form a peptide bond as shown below. Thus the protein is a long polypeptide chain.



Once a protein chain forms, it can fold into a variety of complex three-dimensional conformations. Among the possible conformations usually only one exhibits biological activity. This "native" conformation generally minimizes the free energy of the protein and is therefore the most stable. Many factors contribute to this stability including hydrogen bonding, disulfide bonding, Van der Waals forces, and solvent interactions.

Three common structural motifs recur over and over again in proteins: the α helix, the β sheet, and the globular conformation. In the α helix the chain is tightly coiled about its longitudinal axis. In the zigzag β sheet the chain can be visualized as pleated strands of protein. In the globular conformation, which is the most complex, the chain is irregularly and tightly folded into a compact, nearly spherical shape. Short stretches of the chain are often constructed from α helices or β sheets.

Many large proteins consist of smaller protein subunits that interlock into one macromolecule. Such complex structures operate as coordinated factories in which each subunit contributes a specialized function to the macromolecular protein. ■



(see "What Is a Soliton?"). The soliton results from a nonlinear coupling between the vibrational excitation and a deformation in the protein structure caused by the presence of the excitation. The excitation and the deformation balance each other, and the resulting excitation moves through the protein uninhibited, much the way electrons move in the superconducting state of a metal.

Davydov worked out these ideas for one particular protein conformation, the α helix pictured at the beginning of this article. He introduced a simple mathematical model to show how solitons could travel along the three spines, or hydrogen-bonded chains, of the protein.

Davydov first applied this idea to the problem of muscle contraction. He proposed that myosin, a major contractile protein in striated muscle that has an α-helical tail approximately 1500 angstroms long, propagates a soliton that squeezes and pulls on the actin filaments around it. This action serves to slide the actin and myosin filaments together and thereby results in muscle contraction. In addition, Davydov and his coworkers have considered the idea that α-helical proteins may facilitate electron transport through a soliton mechanism. In this case an extra electron causes a lattice distortion in the protein that stabilizes the electron's motion. Thus it may be reasonable to consider charge transfer across membranes, energy coupling across membranes, and energy transport along filamentous cytoskeletal proteins in terms of a soliton mechanism, since the proteins that carry out these functions contain structural units with significant α-helical character.

The soliton model is one of several concepts for protein dynamics that should attract the careful attention of biologists. Clearly, it cannot explain every aspect of protein dynamics, but it is motivating exciting questions and new experiments. In the following sections we will describe Davydov's concept in the context of the α helix and expand it to a crystalline polymer called acetanilide, which was observed by G. Careri to have an anomalous spectral line near the amide-I band that might be due to a soliton. We will discuss experimental techniques for verifying the existence of solitons in α-helical proteins and acetanilide and consider the concept of self-focusing in globular proteins. (A further application of the soliton model is discussed in "A Possible Mechanism for General Anesthesia.")

Before discussing details of the soliton model, we will try to make the relevant biological context more vivid to the reader by presenting three specific examples where soliton-like dynamics may well be operating.

Three Sites of Action at a Distance

Alpha-helical structure is quite common in proteins, and in particular it is present where energy appears to be transported from one end of a protein to the other or where two processes appear to be coupled by a protein.

Mitochondria. Mitochondria are the energy-generating stations for living cells. These organelles, which may have evolved from separate organisms that were later incorporated into the cell, occupy approximately 20 percent of the total cellular volume. Within these organelles has developed a very specialized protein unit specifically designed to synthesize ATP. It is called the tripartite repeating unit (Fig. 2). Numerous copies of this unit make up the flexible inner membrane of a mitochondrion. As shown in Fig. 2, each time three ATP molecules are synthesized in the head of the F_0–F_1 protein, a pair of electrons (which are donated by the Krebs cycle via the intermediate NADH) circulates among the membrane-bound electron-transport proteins (labeled I, II, III, and IV). The electrons ultimately combine with oxygen and protons to produce water. The movement of these electrons back and forth across the inner membrane, in turn, creates a proton gradient across the membrane that drives the synthesis of ATP in the head of the F_0–F_1 protein. At present, the nature of the driving mechanism is an open question. How do the cytochrome proteins in subunits I, II, III, and IV facilitate electron and proton transport across the thickness (about 60 angstroms) of the inner mitochondrial membrane and at the same time couple ion transport to ATP synthesis? (Semiclassical theories account for electron tunneling between the donor and acceptor heme groups that are attached to the cytochrome proteins and have thus explained oxidation-reduction rates in cytochromes. These theories, however, have not connected electron tunneling to ATP synthesis.)

The dominant configuration of the cytochromes is α-helical, and these proteins span the inner membrane. Given these facts we may ask whether a soliton-like mechanism in these proteins may have anything to do with the stabilization of electron transport and its connection to ATP synthesis.

A related question concerns the contracted configuration of a mitochondrion during ATP synthesis. When a mitochondrion is inactive, its inner membrane is relaxed and spread out, but when active, its inner membrane abruptly contracts into a more wrinkled and twisted appearance (Fig. 3). This brings the myriad tripartite repeating units in the inner membrane into closer apposition with one another. Apparently this aggregation of transmembrane proteins is a prerequisite for ATP synthesis. Whether or not this aggregation induces a change in the conformation of the individual transmembrane proteins is not clear. However, if a soliton-like mechanism were operating during ATP synthesis, it could well be affected by such changes in protein conformation.

Cytoskeleton. The cytoskeleton is a framework of interconnected proteins that literally fills and bridges the inside of a cell (Fig. 4). It provides an internal structure on which the "bag" of the cell rests

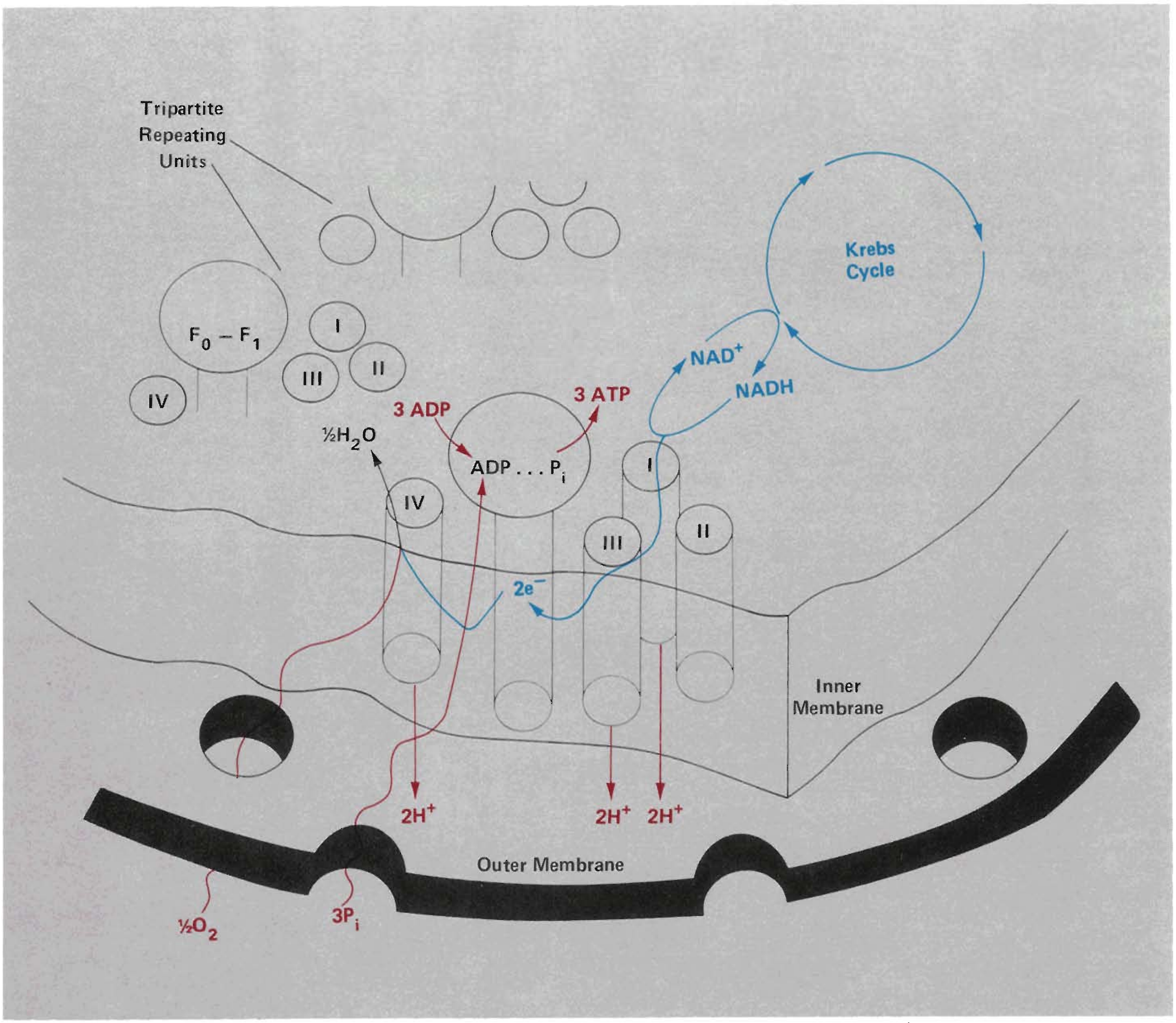


Fig. 2. Cross section of the mitochondrial membrane showing ATP synthesis (red) and electron transport (blue). These two processes both take place in the tripartite repeating units of the flexible inner membrane, and they appear to be coupled. For every three ATP molecules synthesized in the head of the mushroom-shaped central protein F_0–F_1, a pair of electrons circulates through three of the four membrane-bound protein subunits (I, III, and IV) and combines with oxygen and protons to form water. The electron pair is donated by the Krebs cycle through the intermediate NADH. A proton gradient across the inner mitochondrial membrane couples ATP synthesis to the electron transport, or respiratory, chain, but it is not clear how electron transport is related to proton pumping.

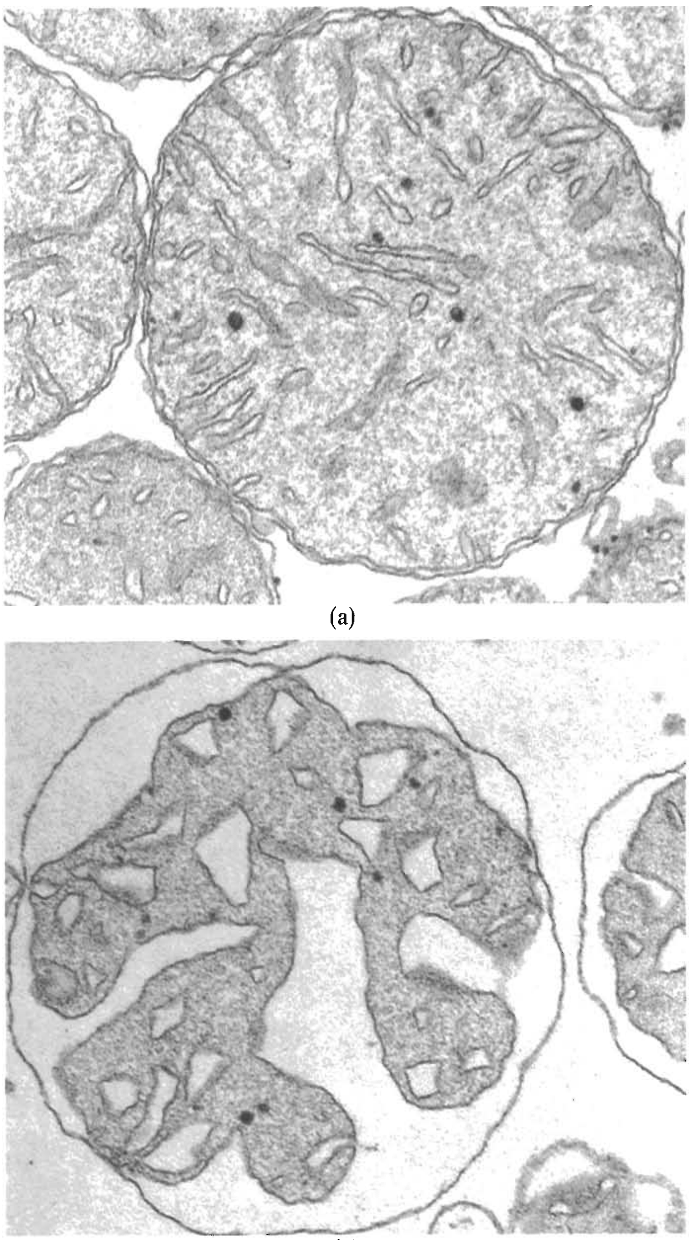


Fig. 3. Electron micrograph of isolated rat-liver mitochondria fixed in the (a) inactive and (b) active conformations. When a mitochondrion is actively synthesizing ATP, its flexible inner membrane contracts around the inner contents of the mitochondrion (darker area) and brings the tripartite repeating units of the inner membrane in closer apposition. (These micrographs, courtesy of C. Hackenbrock, appear in Mitochondria *by Alexander Tzagoloff (Plenum Press, 1982).)*

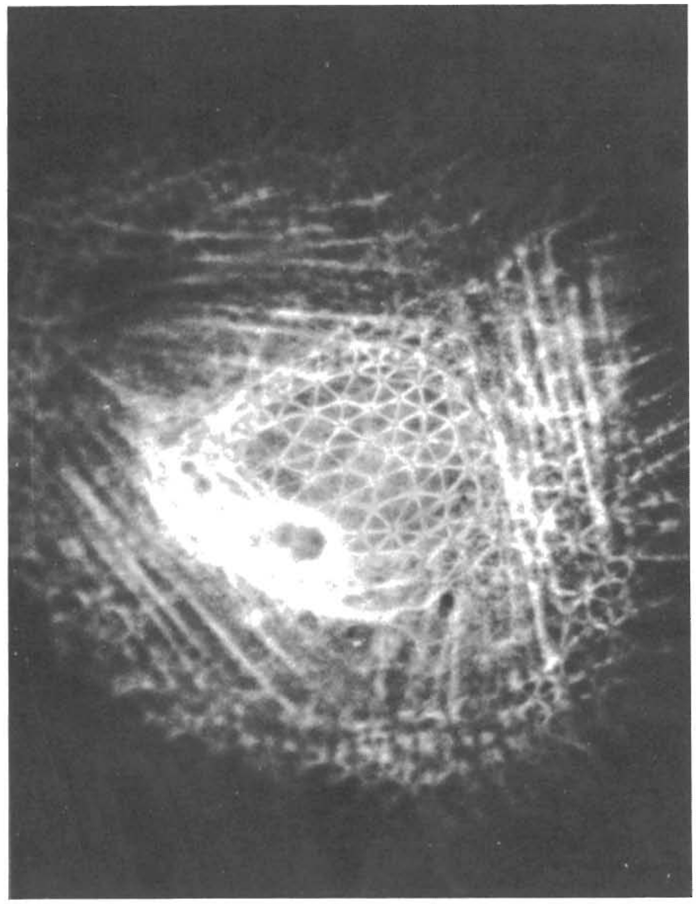


Fig. 4. Immunofluorescence micrograph showing a portion of the cytoskeleton from a cultured rat-embryo cell at a magnification of 10,000. The micrograph was prepared with anti-tropomyosin. The geodesic network of bundles visible in the micrograph forms around the cell's nucleus as the cell changes from a motile to a spread-out, immotile state. (Micrograph courtesy of E. Lazarides.)

itself and on which the cell moves, divides, and changes its outer shape. More recent discoveries suggest that it may be an internal "telegraph network" that allows for intracellular communication or an energetic "conduit" on which the soluble enzymes of the cytosol attach and coordinate their metabolic activities. In any case the cytoskeleton is constantly remodeling itself to meet the demands of the cell. This requires the expenditure of ATP, which in turn appears to be under the fine control of Ca^{2+} concentrations within the cellular juice.

In the proteins of the cytoskeleton, the α helix is a common structural motif. At least three contractile proteins are highly α helical: spectrin, tropomyosin, and myosin. There is also an extensive network of intermediate filamentous proteins within the cytoskeleton that has a predominantly helical character. Cytoskeletal proteins tend to be much longer (300-1500 angstroms) than membrance proteins, and they also tend to form very stable coiled-coil structures, which are capable of polymerizing in a head-to-tail configuration to create exceptionally long α-helical networks within the cytosol. Therefore, it is again reasonable to consider that soliton-like dynamics may be helpful for understanding energy transfer in these longer proteins, where equilibrium models of protein behavior fail to account for translational dynamics.

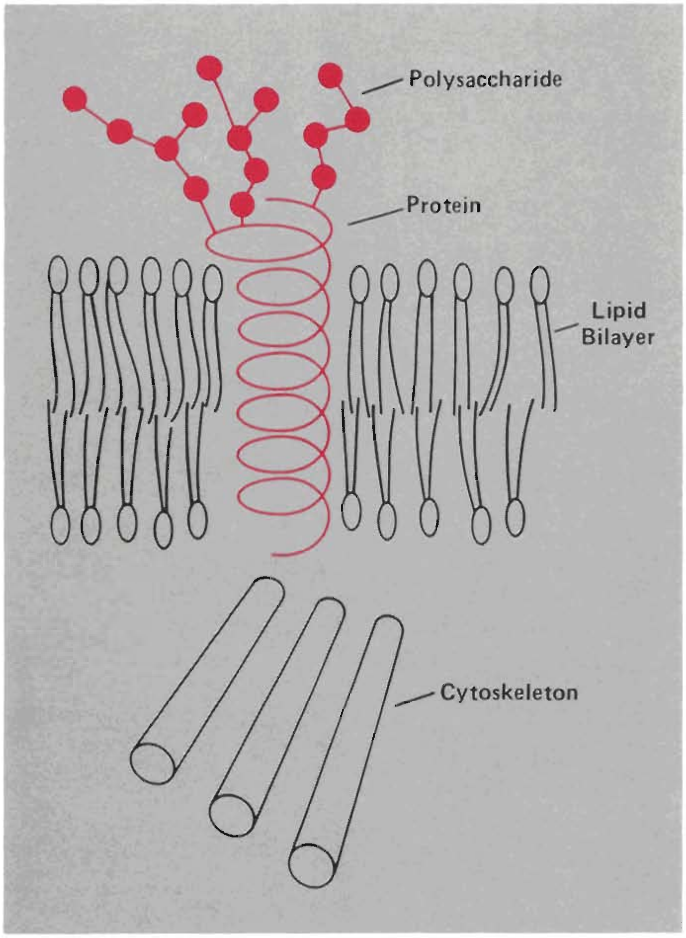


Fig. 5. Glycoproteins (red) embedded in a cell membrane physically connect the inside and the outside of the cell. The polysaccharide portion of the glycoprotein is on the cell surface, and the protein portion juts through the lipid bilayer to form the membrane channels. Some glycoproteins are closely associated with microfilaments and microtubules in the cytoskeleton.

Glycoproteins. Glycoproteins are combinations of sugars and proteins that are covalently bonded together. In the lipid bilayer, or outer membrane, of a cell are numerous glycoproteins, which vary markedly in size and chemical composition (Fig. 5). Many of these macromolecules span the entire thickness of the lipid bilayer; that is, they physically connect the inside and the outside of the cell. As a glycoprotein floats in the lipid bilayer, its polysaccharide portion is in the aqueous phase surrounding the cell, while most of its protein portion is in the lipid phase of the membrane. These transmembranous glycoproteins are crucial to the livelihood of the cell. They are implicated in cellular adhesion, cellular migration, cellular identity, intercellular communication, and transmembrane signaling. They allow a ready pathway over which signals that originate on the cellular exterior (through the binding of a hormone, neurotransmitter, or immunoglobulin) are conveyed directly to the cellular interior.

Much of the protein fraction of a glycoprotein is in the α-helical conformation. Thus information about events on the outside of a cell may be conveyed through a helical channel across the thickness of the lipid bilayer (about 60 angstroms). Therefore, nonlinear dynamics may again provide a key for understanding the mechanisms by which chemical "whispers" are detected and processed on the membrane surface.

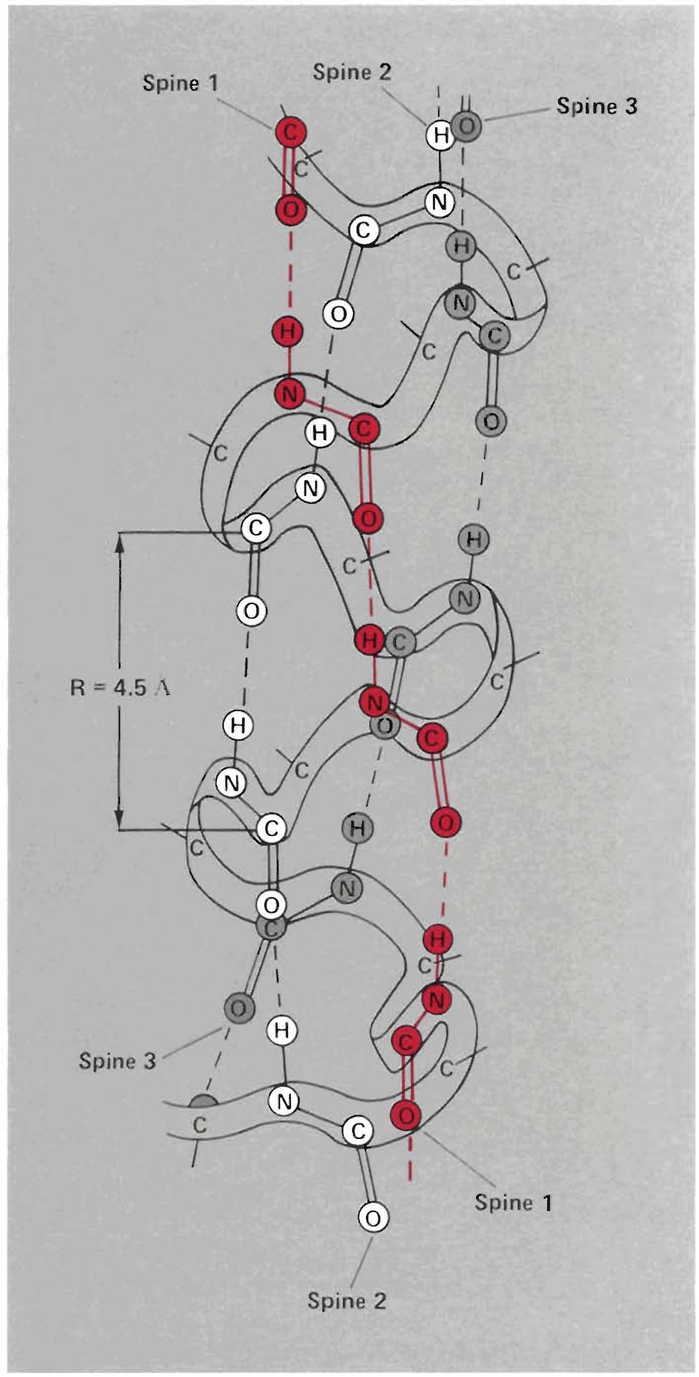


Fig. 6. Portion of an α-helical protein showing the three spines of hydrogen-bonded peptide groups. The dipole moment of each peptide group is approximately colinear with the adjacent hydrogen bond.

Solitons on the α Helix

Alpha-helical proteins, which are implicated in so many ways in energy transport and energy coupling, are the context for Davydov's theory. As the name implies, the conformation of these proteins is a helix formed by the twisting of the protein backbone. In addition, hydrogen bonds link the peptide groups together to form three spines that span the length of the helix and stabilize it. (The reader might like to make a model of the protein backbone like the one pictured in the opening figure. To form the helix, wind the backbone into a right hand spiral and attach the hydrogen of the first peptide group to the oxygen of the fourth group, the hydrogen of the second peptide group to the oxygen of the fifth, and so on. Note the formation of three spines of hydrogen-bonded peptide groups. The first spine consists of the first, fourth, seventh, tenth, etc., peptide groups. The second and third spines form similarly.) The spines of an α-helical protein are not exactly linear or parallel to the axis of the helix (Fig. 6), but nevertheless the electric dipole moments $\vec{d}$ of the amide-I vibrations are essentially in the same direction as the hydrogen bonds that define the spine. This fact, as we will see, leads to cooperative behavior along each chain of hydrogen-bonded peptide groups.

Consider a single linear chain of hydrogen-bonded peptide groups. Figure 7 shows three interactions that occur when an amide-I vibration in a particular peptide group is excited, say, by the hydrolysis of ATP. First (Fig. 7a), there will be resonant interactions with neighboring peptide groups due to electromagnetic dipole-dipole interactions, much like the interaction between transmitting and receiving antennae of a radio system. This interaction alone would lead to dispersion of amide-I energy. Second (Fig. 7b), due to changes in static forces (hydrogen bonds, Van der Waals forces, etc.), the

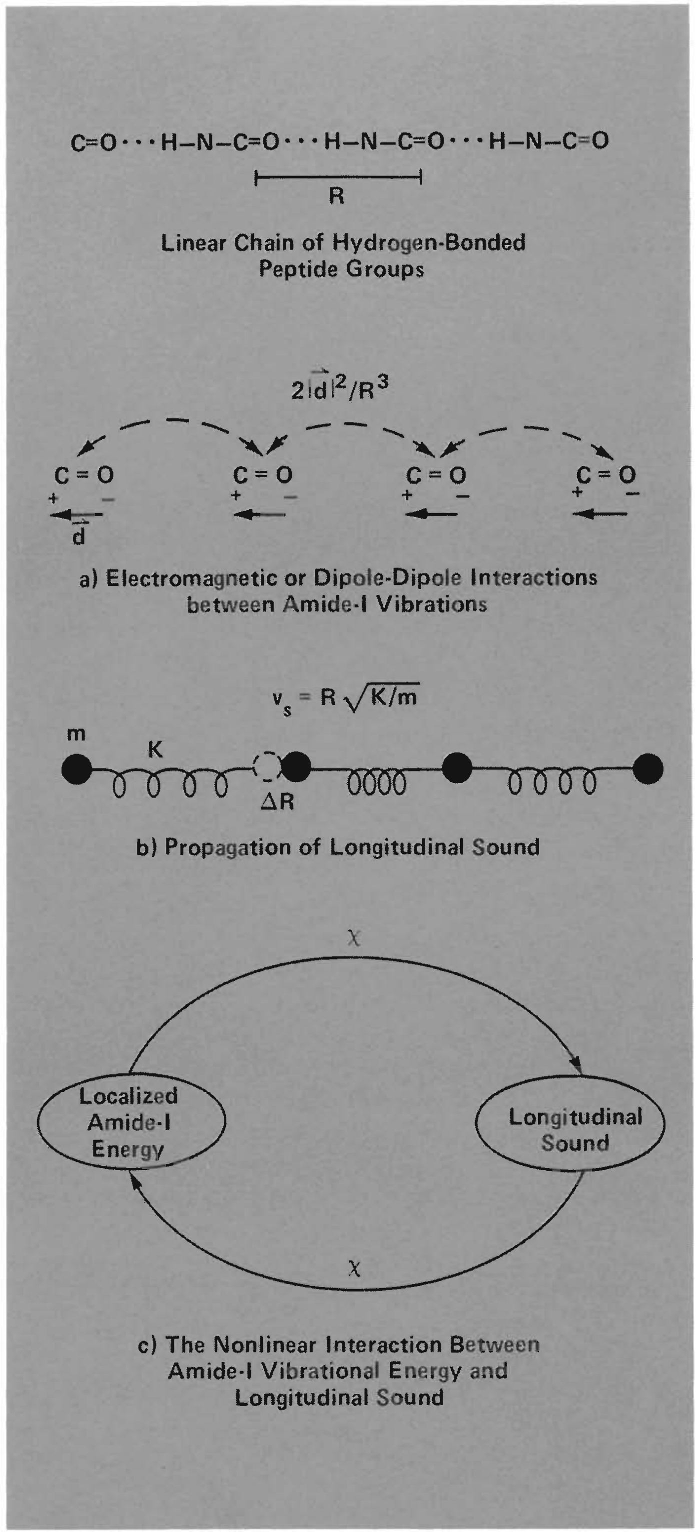


Fig. 7. Linear chain of hydrogen-bonded peptide groups showing the three interactions that combine to trap amide-I vibrational energy in a stable solitary wave, or soliton. Peptide groups with electric dipole moments $\vec{d}$ are separated by a distance R from each other. (a) The dipole-dipole interaction energy between neighbors on the chain is equal to $2|\vec{d}|^2/R^3$. (b) A peptide group of mass m displaced ΔR from its equilibrium position sets up a longitudinal sound wave along the chain. The wave travels at velocity $v_s = R(K/m)^{1/2}$, where K is the strength of the weak spring that represents the hydrogen bond between peptide groups. (c) The nonlinear interaction between amide-I vibrational energy and longitudinal sound is represented by a feedback loop in which one interaction reacts back on the other and vice versa. The strength of the interaction is proportional to χ^2 and inversely proportional to K.

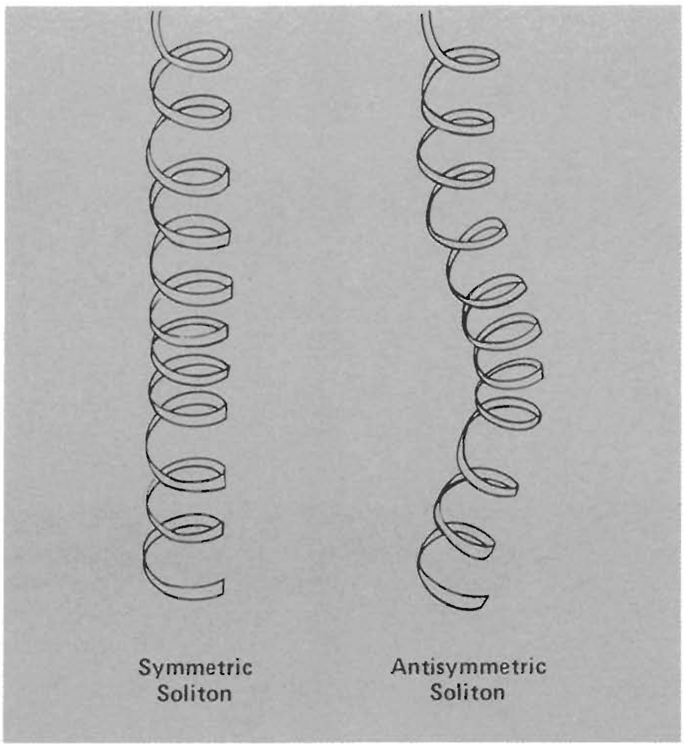


Fig. 8. Two types of solitons can form on the α helix: (1) a symmetric soliton in which amide-I vibrational energy is shared equally among the three spines, and (2) an antisymmetric soliton in which amide-I energy is shared unequally. The soliton is localized over about four turns of the helix. The symmetric soliton causes compression in the longitudinal direction, and the antisymmetric soliton causes bending of the helix.

excited peptide group will tend to move from its equilibrium position, causing a local deformation of the hydrogen bond in the region of excitation. In α-helical proteins the largest displacement will also be along the hydrogen bonds because hydrogen bonds are weaker than the covalent bonds along the helix. Since the hydrogen bond behaves like a weak spring, this movement of the peptide group away from equilibrium will set up a longitudinal sound wave, or phonon, along the chain as the peptide groups oscillate about their equilibrium positions.

These two dynamical effects are displayed in Fig. 7 as if they were uncoupled; that is, dispersion of amide-I bond energy (Fig. 7a) is independent of the propagation of longitudinal sound waves (Fig. 7b). Davydov, however, pointed out that the two effects are coupled by a *nonlinear* interaction that arises from the change in amide-I vibrational energy E caused by a change in the distance R between peptide groups along the chain (hydrogen-bond stretching). The strength of this coupling is proportional to the nonlinear parameter

$$\chi \equiv \frac{dE}{dR} ,$$

which can be expressed in units of joules per meter, or newtons. The effect of this nonlinear coupling is displayed graphically in Fig. 7c. Localized amide-I vibrational energy acts (through χ) as a source of longitudinal sound, and this longitudinal sound reacts (again through χ) as a potential well that traps the amide-I vibrational energy and prevents its dispersion. Coupled together, the localized amide-I vibrational energy and the longitudinal deformation can travel along the chain as a soliton with no energy loss.

As shown in the next section, this soliton is described by the nonlinear Schrödinger equation. Figure 7c shows that the strength of the nonlinear effect is proportional to χ^2. It is also inversely proportional to K, the spring constant of the hydrogen bonds connecting the peptide groups. If the linear chain were absolutely rigid, K would equal infinity, there would be no nonlinear interaction, and the amide-I energy would disperse.

In α-helical proteins the three spines are coupled to each other by additional transverse dipole-dipole interactions. This situation can be described by three coupled nonlinear Schrödinger equations. The solutions to these equations yield two types of solitons, a symmetric one in which the energy of the amide-I excitation is shared equally by all three chains and an antisymmetric one in which the amide-I energy and the accompanying deformation are shared unequally and the molecule bends (Fig. 8). The antisymmetric soliton is lower in energy and is therefore more likely to occur.

This collective excitation is called a soliton because it behaves in many ways like a particle. For example, its energy is the sum of its internal energy plus its kinetic energy, $1/2\, M_{\mathrm{eff}}\, v^2$, where M_{eff} is the effective mass of the soliton and v is the velocity with which the excitation and accompanying distortion travel down the length of the helix. This velocity is less than the velocity of longitudinal sound in the molecule. The soliton travels with little or no energy loss and is therefore a very efficient means of energy transport along the length of the helix. Moreover, as will be shown in the mathematical development below, the soliton has an energy less than the energy of the amide-I vibration alone. It is therefore energetically favorable for solitons to form from amide-I excitations.

The mathematical model first developed by Davydov is a semiclassical approximation in which the amide-I excitations are treated quantum mechanically, and the displacements of the peptide groups, or longitudinal sound wave, along the hydrogen-bonded chain are treated classically. We will present the Davydov model for collective excitations along a single chain of hydrogen-bonded peptide groups and show how the continuum approximation of this model leads to the nonlinear Schrödinger equation and its well-known soliton solutions. (The reader unfamiliar with the formalism of quantum mechanics may skip the next section without losing the main points of the article.)

The Davydov Model: How It Yields Soliton Solutions

The energy operator H, or Hamiltonian, for the collective excitation along the chain is a sum of three operators: $H = H_{\text{amide-I}} + H_{\text{phonon}} + H_{\text{interaction}}$, where $H_{\text{amide-I}}$ is the operator for the amide-I vibrational excitations, H_{phonon} is the operator for the displacements of the peptide groups, and $H_{\text{interaction}}$ is the operator for the interaction between the amide-I excitations and the displacements.

If E is the amide-I excitation energy and $B_n^\dagger$ is an operator for creation of this excitation on the nth peptide group, then $H_{\text{amide-I}}$ is given by

$$H_{\text{amide-I}} = \sum_n \left[EB_n^\dagger B_n - J(B_n^\dagger B_{n-1} + B_n^\dagger B_{n+1})\right] , \tag{1}$$

where the summation is carried out over all N peptide groups. The first term, $EB_n^\dagger B_n$, defines the amide-I excitation energy, and the second term describes the resonance dipole interaction between nearest neighbors. The operators $B_n^\dagger B_{n-1}$ and $B_n^\dagger B_{n+1}$ represent transfer of amide-I energy from peptide group n to $n \pm 1$ due to the dipole-dipole interaction. The dipole-dipole interaction energy J is given by $2|\vec{d}|^2/R^3$, which is the usual electrostatic energy associated with two colinear dipoles of moment $\vec{d}$ separated by the distance R.

The energy H_{phonon} associated with displacing the peptide groups away from their equilibrium positions is given in the harmonic approximation by

$$H_{\text{phonon}} = \frac{m}{2} \sum_n \left(\frac{du_n}{dt}\right)^2 + \frac{K}{2} \sum_n (u_n - u_{n+1})^2 , \tag{2}$$

where u_n is the displacement of the nth peptide group, m is the mass of the peptide group, and K is the spring constant, or elasticity coefficient, of the hydrogen bonds forming the linear chain. The first term is kinetic energy and the second potential energy.

The Hamiltonian for the interaction between the amide-I excitation and the displacements of the peptide groups takes the form

$$H_{\text{interaction}} = \chi \sum_n (u_{n+1} - u_{n-1}) B_n^\dagger B_n , \tag{3}$$

where the coupling constant χ, as mentioned earlier, represents the change in amide-I energy per unit extension of an adjacent hydrogen bond.

The total Hamiltonian $H = H_{\text{amide-I}} + H_{\text{phonon}} + H_{\text{interaction}}$ of the system must satisfy the Schrödinger equation:

$$i\hbar \frac{\partial}{\partial t} |\psi\rangle = H|\psi\rangle \quad . \tag{4}$$

The wavefunction $|\psi\rangle$, which defines the state of the system, is expressed by

$$|\psi(t)\rangle = \sum_n a_n(t) B_n^\dagger |0\rangle \tag{5}$$

and satisfies the normalization condition

$$\langle\psi|\psi\rangle = \sum_n |a_n(t)|^2 = 1 \ . \tag{6}$$

The quantity $|a_n(t)|^2$ is the probability of finding the excitation on the nth peptide group at time t. The state $|0\rangle$ represents an unexcited amide-I vibration, that is, the ground state of the system.

Substituting Eq. 5 in Eq. 4 we get, after some algebra, the following set of differential equations:

$$i\hbar \frac{da_n}{dt} = \left[E + H_{\text{phonon}} + \chi(u_{n+1} - u_{n-1})\right] a_n - J(a_{n-1} + a_{n+1}) , \tag{7}$$

and

$$m \frac{d^2u_n}{dt^2} - K(u_{n+1} - 2u_n + u_{n-1}) = \chi(|a_{n+1})|^2 - |a_{n-1}|^2) \ . \tag{8}$$

Equations 7 and 8 are the main result of Davydov's original model. They describe the time evolution of amide-I vibrational energy coupled to displacements of the hydrogen-bonded chain of peptide groups. The quantity $|a_n(t)|^2$ characterizes the distribution of the amide-I energy over the individual peptide groups of the chain.

In order to demonstrate that $a_n(t)$, the probability amplitude of the excitation, does behave like a soliton, we will restrict ourselves to solutions of Eqs. 7 and 8 that vary slowly as a function of the peptide group number n. In this limit we can replace the functions $a_n(t)$ and $u_n(t)$ with continuous functions $a(x,t)$ and $u(x,t)$, thus approximating n with the dimensionless coordinate x. Equations 7 and 8 then become

$$i\hbar \frac{\partial a}{\partial t} = \left(E_0 + 2\chi \frac{\partial u}{\partial x} \right) a - J \frac{\partial^2 a}{\partial x^2} \tag{9}$$

and

$$\frac{\partial^2 u}{\partial t^2} - \frac{K}{m} \frac{\partial^2 u}{\partial x^2} = \frac{2\chi}{m} \frac{\partial (|a|^2)}{\partial x} , \tag{10}$$

where

$$E_0 = E - 2J + 1/2 \int_{-\infty}^{\infty} m \left(\frac{\partial u}{\partial t} \right)^2 + K \left(\frac{\partial u}{\partial x} \right)^2 dx .$$

The left side of Eq. 10 is essentially a wave equation for longitudinal sound in the system of coupled peptide groups; the sound velocity v_s is given by $v_s = R(K/m)^{1/2}$. The right side acts as a source term for generation of sound.

We shall seek traveling wave solutions of Eqs. 9 and 10 in the form of excitations that propagate along the chain with a velocity v; that is,

$$u(x,t) = u(x - vt) . \tag{11}$$

Inserting Eq. 11 in Eq. 10, we get

$$\frac{\partial u(x,t)}{\partial x} = \frac{-2\chi}{K(1 - s^2)} \, |a(x,t)|^2 , \tag{12}$$

where s is the ratio of the propagation velocity to the velocity of sound: $s = v/v_s < 1$.

Substituting Eq. 12 into Eq. 9, we get

$$i\hbar \frac{\partial a}{\partial t} + J \frac{\partial^2 a}{\partial x^2} - E_0 a + \kappa |a|^2 a = 0 , \tag{13}$$

where $\kappa = 4\chi^2/K(1 - s^2)$.

Equation 13 is the nonlinear Schrödinger equation, which has soliton solutions. For its general solution we refer the interested reader to "Exact Theory of Two-Dimensional Self-Focusing and One-Dimensional Self-Modulation of Waves in Nonlinear Media" by V. E. Zakharov and A. B. Shabat (*Soviet Physics JETP* 34(1972): 62-69). It is sufficient for our purpose here to look only at a stationary solution, that is, one for which $s = 0$. In this case a solution of Eq. 13 is given by

$$a(x,t) = \frac{\chi}{(2KJ)^{1/2}} \operatorname{sech} \frac{\chi^2}{KJ} (x - x_0) \exp \left[- \frac{i}{\hbar} \left(E_0 - \frac{\chi^2}{K^2 J} \right) t \right] . \tag{14}$$

The constant x_0 is the position of maximum probability of amide-I excitation along the chain, and the pulse-shaped form given by Eq. 14 falls off rapidly when one moves away from x_0 (see "What Is a Soliton?"). Equation 14 also satisfies the continuous equivalent of Eq. 6:

$$\int_{-\infty}^{\infty} |a|^2 dx = 1 ,$$

indicating that one quantum of amide-I energy is excited on the peptide chain.

The energy E_{sol} associated with the soliton is given by

$$\begin{aligned} E_{sol} &= E_0 - \frac{\chi^4}{K^2 J} \\ &= E - 2J + 1/2 \int_{-\infty}^{\infty} m \left(\frac{\partial u}{\partial t} \right)^2 + K \left(\frac{\partial u}{\partial x} \right)^2 dx - \frac{\chi^4}{K^2 J} . \end{aligned} \tag{15}$$

Inserting the solution given by Eq. 14, we get

$$E_{sol} = E - 2J - \frac{\chi^4}{3K^2 J} . \tag{16}$$

This is the energy of a stationary soliton. A similar but more complicated expression can be obtained for moving solitons.

It is instructive to see what happens in an absolutely rigid chain of peptide groups. In such a case $K = \infty$, and the multiplicative constant κ in the nonlinear term of Eq. 13 is equal to zero. Equation 13 is then a linear Schrödinger equation, which has solutions in the form of plane waves. This means that an excitation is uniformly distributed along the whole chain. In other words, the amide-I energy has dispersed and is no longer localized. It can also be seen from Eq. 16 that the energy of such an extended excitation (or exciton) equals $E - 2J$, which is larger by an amount $\chi^4/3K^2J$ than the energy E_{sol} of the spatially localized soliton excitation. It is thus more favorable for the system to localize its energy when the nonlinear coupling between amide-I energy and displacement of the associated peptide group is taken into account.

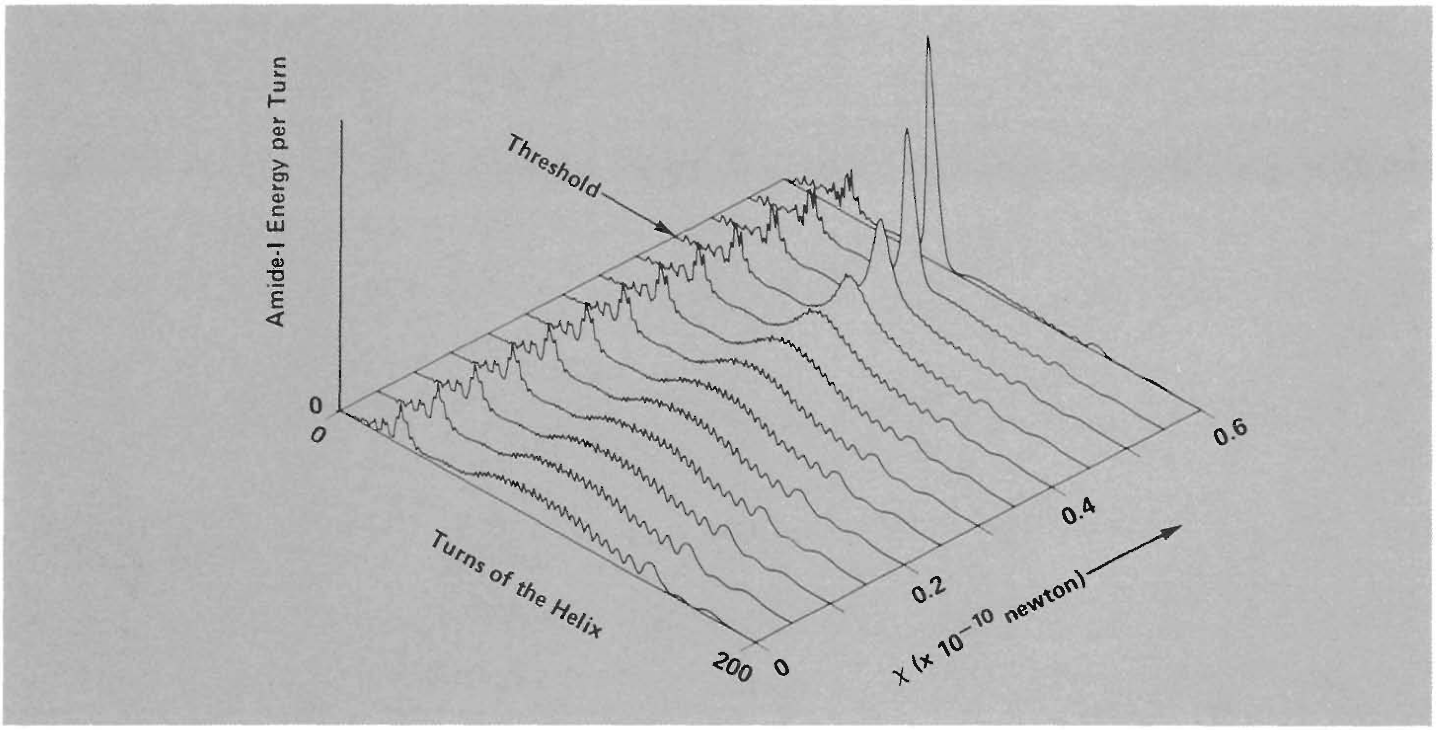


Fig. 9. Distribution of total amide-I energy (sum over all three spines) along the spines as a function of the nonlinear coupling parameter χ, as calculated numerically by L. MacNeil and A. Scott. For $\chi \geq 0.45 \times 10^{-10}$ newton, the amide-I energy self-focuses into a soliton-like excitation.

Threshold Conditions for Soliton Formation

Although the Davydov model leads to the nonlinear Schrödinger equation and soliton solutions, one can, with some justification, ask whether the nonlinear Schrödinger equation has anything to do with energy transport in a real biological system. We certainly do not expect that the smooth mathematical properties of the soliton will carry over unaltered to a real system. We do suggest, however, that the nonlinear interactions between amide-I excitations and lattice distortions will lead to stable pulse-like excitations of focused energy. Without being too dogmatic, we often find it helpful to refer to such excitations as solitons, and it is in this sense that the nonlinear Schrödinger equation serves as a useful tool in the analysis.

The strength of the coupling between amide-I vibration and lattice distortion depends on the nonlinear parameter χ. The value of this coupling constant is therefore very important when Davydov's theory is applied to real systems. In 1979 J. M. Hyman, D. W. McLaughlin, and A. C. Scott started a numerical investigation of the Davydov model applied to an α-helical protein. The equations describing this system consist of three sets of equations like Eqs. 7 and 8 but with additional terms to account for transverse dipole-dipole coupling between the three hydrogen-bonded chains of the α helix. Numerical solution of these equations yielded an important result: *The coupling between amide-I vibration and lattice distortion must be sufficiently strong for self-focusing to take place.* Below a certain threshold value a soliton cannot form and the dynamics is essentially linear. The result of a subsequent detailed investigation is summarized in Fig. 9, which shows the distribution of amide-I energy along the helix for different values of χ. For $\chi \geq 0.45 \times 10^{-10}$ newton, a soliton-like object is seen to form.

About the same time that this numerical work was being conducted at Los Alamos, an independent research effort was taking place at the Institute of Theoretical Physics in Kiev. To estimate χ from first principles, V. A. Kuprievich and Z. G. Kudritskaya were doing *ab initio* quantum-chemical calculations on the electronic structure of a dimer of formamide. This molecule consists of two peptide groups connected by a hydrogen bond and therefore serves as a tractable model for more complex protein structures. Since χ is related to the change in the C=O spring constant (K) per unit change of the hydrogen-bond length, estimates of χ can be obtained from two values of K corresponding to two hydrogen-bond lengths. Kuprievich and Kudritskaya thus estimated χ to lie between 0.3×10^{-10} and 0.5×10^{-10} newton. Careri has made an empirical estimate of χ from a comparison of amide-I energies and hydrogen-bond lengths for various polypeptide crystals. He found χ to be about 0.62×10^{-10} newton. These estimates for χ indicate that the level of nonlinearity in real systems is sufficient to allow self-focusing (soliton-like) excitations to form.

Low-Energy Spectrum of Solitons on the α Helix

Numerical studies of Davydov's model for the helix show that the collective excitations behave very much like the solitons of the nonlinear Schrödinger equation. For example, they are remarkably stable upon collision. This property is illustrated in Fig. 10, which shows the total amide-I energy (the sum over all three chains of the helix) as a function of time and peptide group number. Two solitons are launched, one from each end of the helix, by exciting two peptide groups on opposite ends of one of the three chains. The two solitons propagate in opposite directions at approximately three-eighths of the sound speed, or 17 angstroms per picosecond. They collide but pass through each other and retain their identities after the collision.

There is an internal dynamics associated with the soliton propaga-

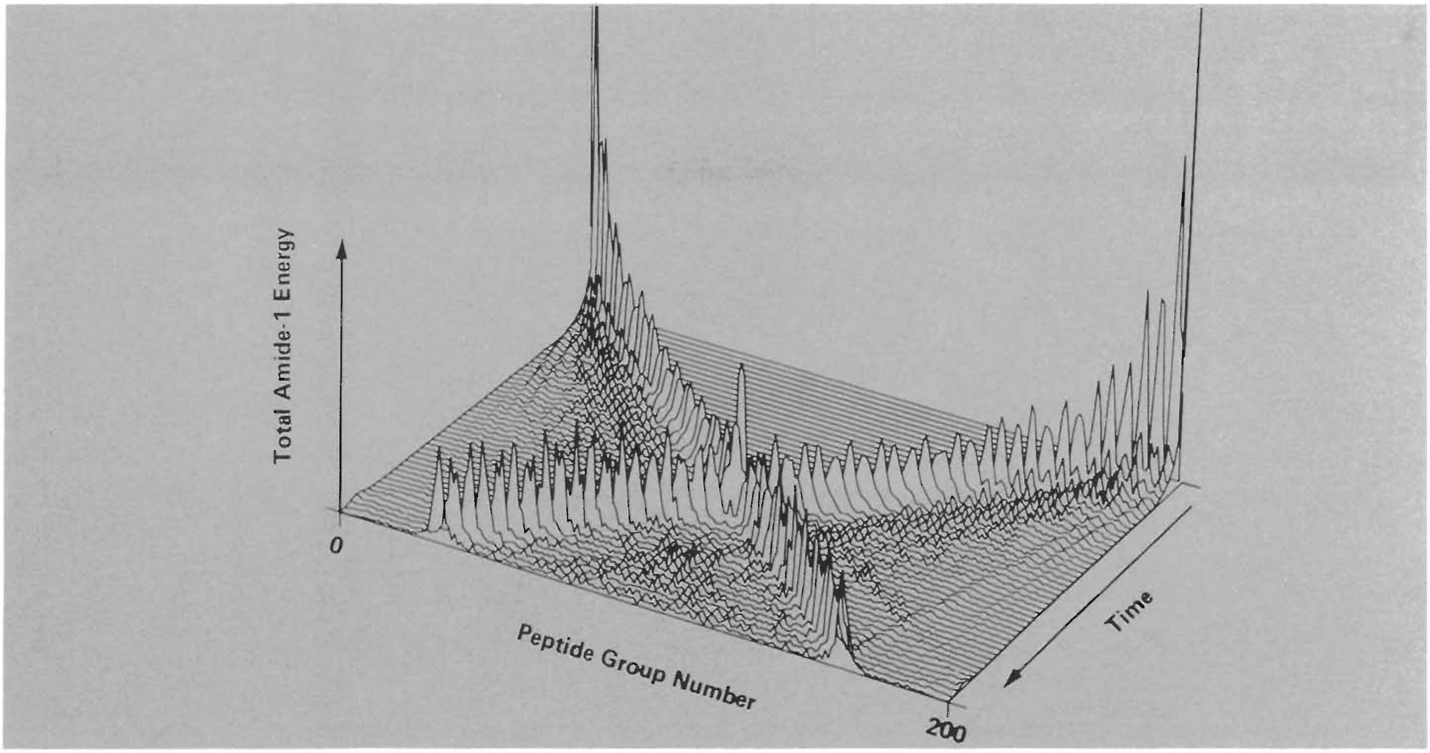


Fig. 10. Numerical calculation showing the particle-like behavior of solitons under collision. The total amide-I energy is plotted as a function of peptide group number and time. The solitons, propagating at three-eighths of the sound speed (17 angstroms per picosecond), pass through each other and maintain their identities.

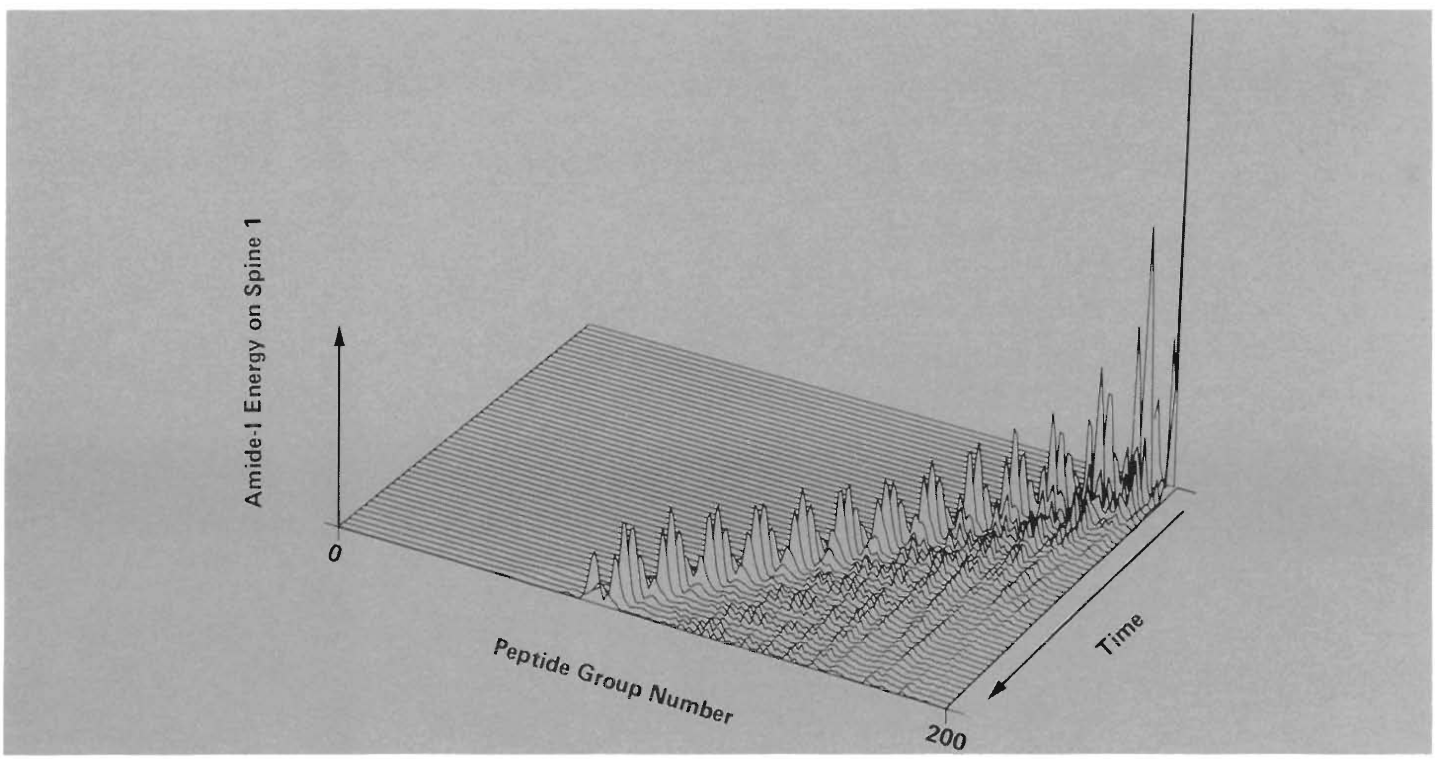


Fig. 11. Amide-I energy of spine 1 of the α helix as a function of time for a propagating soliton launched from the right end of the helix. When the energy is low on the spine shown, it is higher on the other two. The interspine oscillation of energy is caused by transverse dipole-dipole interactions among spines.

tion that cannot be appreciated from Fig. 10, namely the oscillation of soliton energy among the three spines as the soliton propagates down the length of the helix. This interspine oscillation, which is mediated by transverse dipole-dipole interactions between peptide groups on different spines, has a period of about 2 picoseconds. In Fig. 11 we show this phenomenon by depicting the amide-I energy resident on only *one* spine. When the energy is low on the spine shown, it is higher on the other two.

The internal frequencies associated with the dynamics of a propagating soliton are in a range where many proteins are known to exhibit rich dynamic behavior. The "flutter" associated with the movement of the soliton past the periodic structure of unit cells has a spectral frequency of approximately 125 cm^{-1}, and the interspine oscillation has a frequency of approximately 20 cm^{-1}. It is therefore of interest to understand what the Davydov model for an α helix predicts in terms of the low-frequency spectrum. For this purpose we calculated the "total" bending on one spine as a function of time. The bending is defined as $u_{\alpha,1} - u_{\alpha,N}$, where α is the index of one of the three spines, and 1 and N are the numbers of the first and last groups in the chain. This quantity is shown in Fig. 12a as a function of time for two values of the coupling constant χ, one above threshold (A) and one below (B). The spectrum obtained as the square of the Fourier transform of this signal is shown in Fig. 12b. We have identified the interspine oscillation as line 1. Harmonics of this basic oscillation appear as lines 2, 3, 4, and 5. The "flutter" frequency is seen as line 7, and lines 6, 8, and 10 are interpreted as the convolution of lines 1, 2, and 3 with line 7. The lines marked A and B are subharmonics of line 1 generated through the nonlinear interaction of amide-I energy with lattice distortion. This calculated spectrum shows features that correlate well with low-frequency (below 200 cm^{-1}) spectra of metabolically active cells measured by laser Raman spectroscopy. Even though the living system is immensely more complicated than what can possibly be described within the limitations of the Davydov model, the correlation is striking enough to warrant further study.

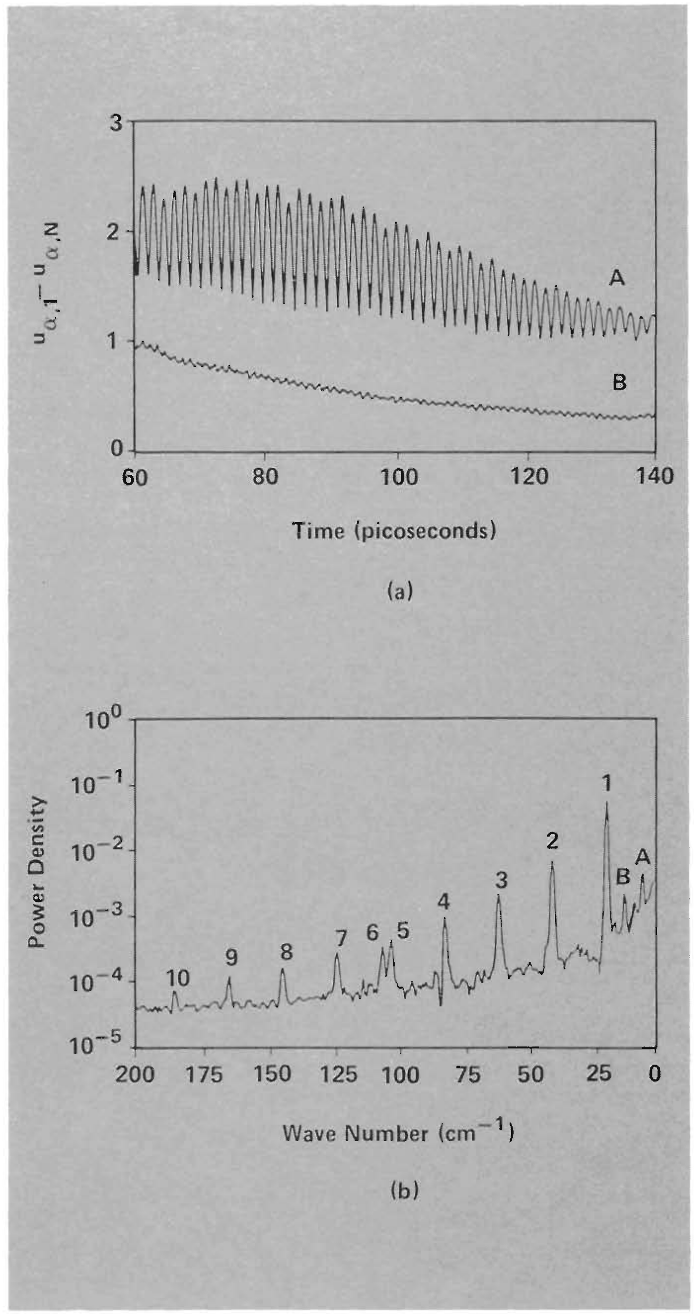


Fig. 12. (a) Time dependence of the bending of a single spine ($u_{\alpha,1} - u_{\alpha,N}$) caused by a propagating soliton for two values of χ, one above the threshold for soliton formation (A) and one below the threshold (B). (b) The spectrum obtained as the square of the Fourier transform of curve A. See the text for an interpretation of this low-frequency spectrum.

Evidence for Solitons from a "Model Protein"

The most convincing experimental evidence for the existence of solitons in proteins comes not from the spectra of proteins or living cells but from the spectrum of the crystalline polymer acetanilide ($(CH_3CONHC_6H_5)_x$), or ACN. ACN is organized into hydrogen-bonded chains (Fig. 13) that are held together transversely by Van der Waals forces. ACN was used as an analgesic in the nineteenth century (its modern chemical cousin is Tylenol), but its interest for our purposes is its remarkable similarity to the chain structure of hydrogen-bonded peptide groups in α-helical proteins (compare Figs. 13 and 6). Late in the 1960s Careri noted that the peptide bond lengths and angles in ACN are very close to those in natural proteins, and he began an experimental program at the University of Rome to see whether ACN would show any unexpected physical properties that might be of biological interest. His intuition was rewarded in 1973 by the observation of an anomalous line in the infrared absorption spectrum of ACN. This anomalous line is lower in wave number than the main amide-I peak by 15 cm^{-1}. Its intensity is low at

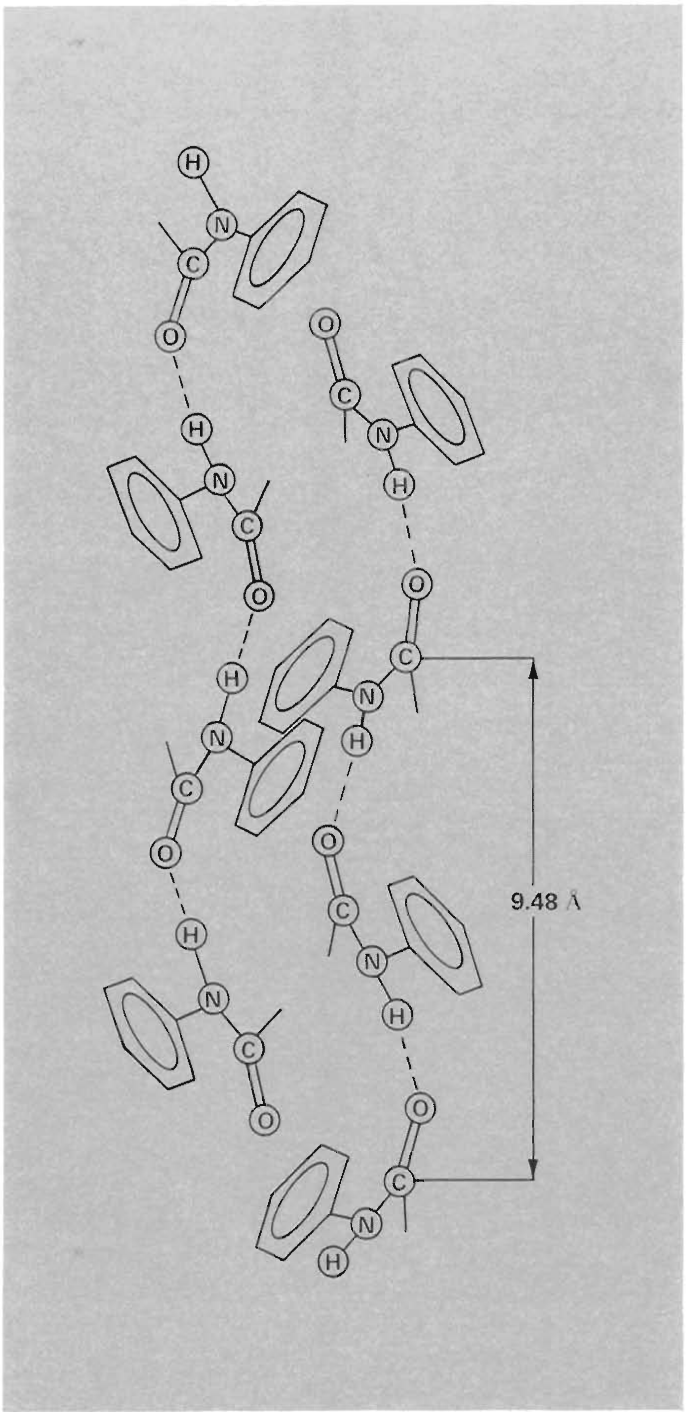


Fig. 13. A portion of two unit cells of crystalline acetanilide (ACN). The two hydrogen-bonded peptide chains resemble the spines of the α-helix shown in Fig. 6.

room temperature but increases as the temperature is lowered (Fig. 14). Numerous attempts by Careri and coworkers to find a conventional assignment for this new line were unsuccessful throughout the 1970s.

Then in 1982, when Scott became aware of Careri's data, he and Careri's group proposed that the anomalous line was due to a new type of soliton, one that results from the coupling of the amide-I vibration to an out-of-plane displacement of the hydrogen-bonding proton rather than to an in-plane displacement of the whole amide group. The soliton arising from this coupling to the lesser mass of the proton can be excited directly by electromagnetic radiation, a necessary ingredient for explaining the ACN data. The types of solitons discussed by Davydov could not be excited directly by radiation because the heavy peptide groups move too slowly about their equilibrium positions.

The mathematics of the modified theory is quite similar to the original Davydov model, and an expression similar to Eq. 16 can also be derived for the energy of this new type of soliton. Using the measured red shift of 15 cm^{-1} in this expression, Scott obtained a value for the nonlinear coupling parameter that agrees reasonably well with the estimated values of the related parameter χ.

The high-resolution ACN data of E. Gratton in Fig. 14 show the temperature dependence of the amide-I band and the anomalous band at 1650 cm^{-1}. The shoulders on the amide-I band are due to the different normal modes of amide-I excitation in the complicated unit cell of ACN. This unit cell has eight distinguishable peptide groups. The splitting of the amide-I band can thus be explained by normal-mode analysis based on group theory. The appearance of the band at 1650 cm^{-1} is consistent with the modified Davydov model and therefore may be due to the presence of solitons.

Laser Raman Spectroscopy. To further check this prediction we have repeated some of the spectroscopic measurements of Careri and coworkers and are now doing additional measurements on single crystals of ACN. (Figure 15 illustrates the principles of Raman scattering, and Fig. 16 shows the experimental setup.) Our intent is either to find alternative explanations for the 1650 cm^{-1} band or to find positive evidence for assigning it to soliton excitations.

We considered the possibility that the 1650 cm^{-1} band is due to a second-order phase transition in the crystal at low temperature. If so, the intensity of the line as a function of temperature would exhibit a threshold near some critical temperature. The measured intensity, however, shows only the smooth, gradual dependence predicted by the soliton model. Further, a second-order phase transition is expected to be accompanied by the appearance of "soft modes," which are evidenced by low-energy (less than 200 cm^{-1}) lines whose frequencies vary quadratically with temperature. No such lines were observed at temperatures ranging from 300 down to 6 kelvins.

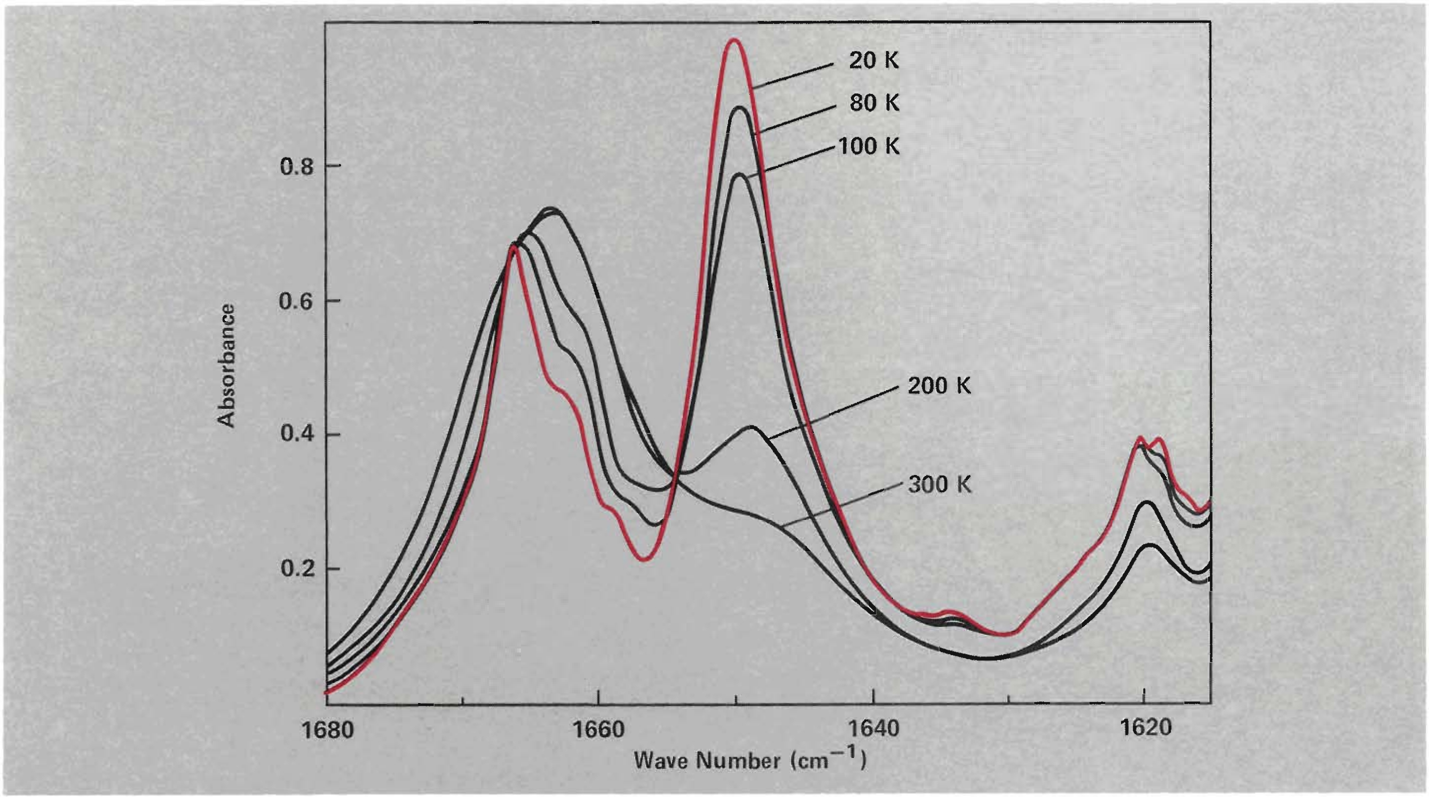


Fig. 14. Infrared spectra of ACN at different temperatures showing the amide-I band at 1665 cm^{-1} and the anomalous band at 1650 cm^{-1}. The 1650 cm^{-1} band may be caused by solitons. (Data courtesy of E. Gratton.)

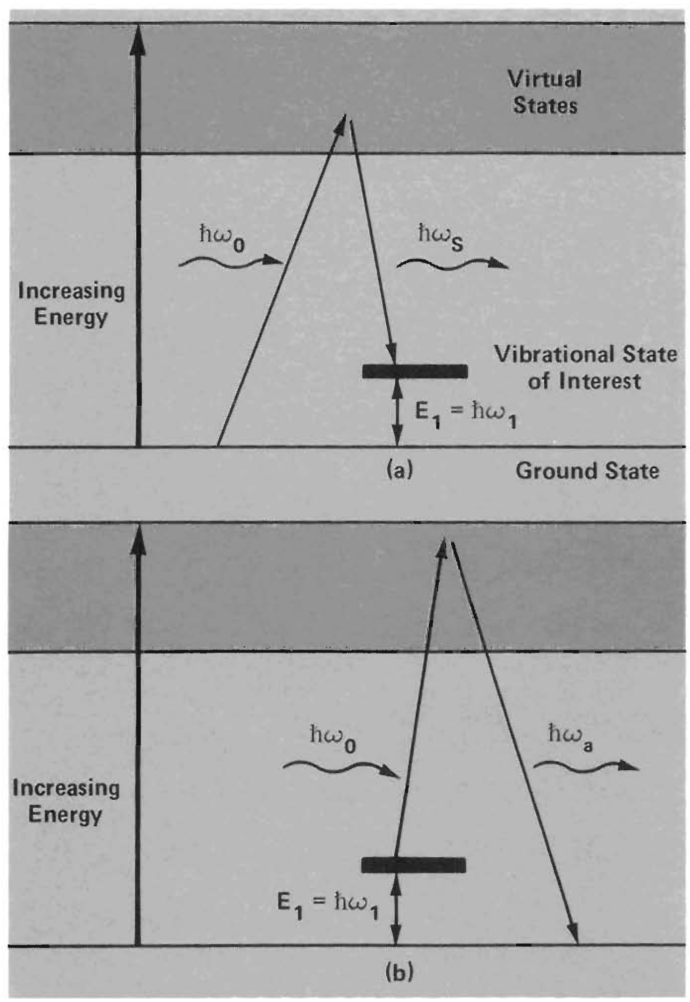
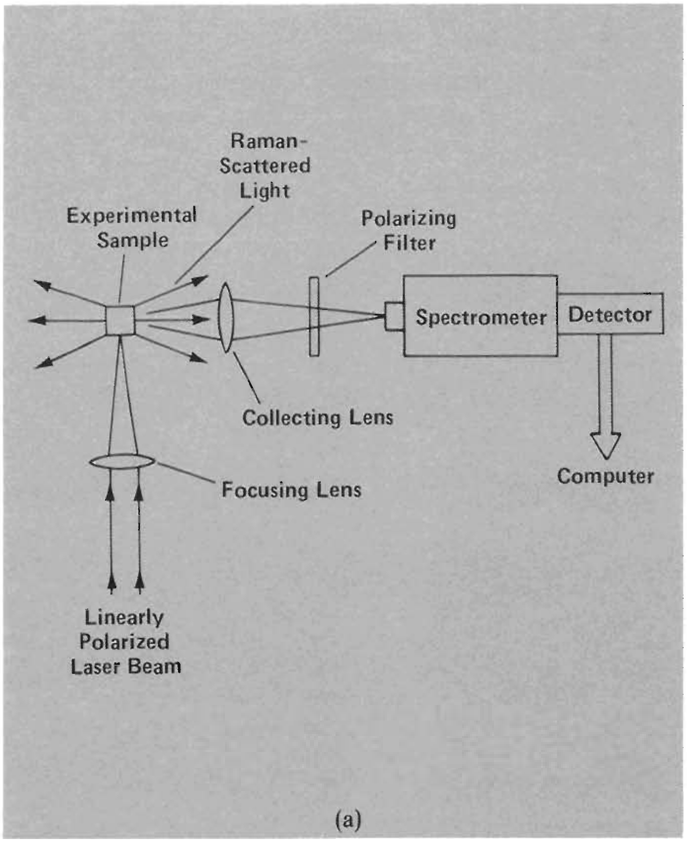


Fig. 15. Raman scattering allows one to use commonly available visible lasers and quantum detectors to analyze states whose energies correspond to wavelengths in the infrared and far-infrared spectral regions. The energy level diagrams show two types of Raman scattering, Stokes scattering (a) and anti-Stokes scattering (b). Both processes probe a particular excited state of the molecule at energy $E = \hbar\omega$. In Stokes scattering a photon from the laser field at frequency ω_0 is "absorbed" by a molecule and excites it from the ground state to a "virtual state." The molecule then instantaneously decays to the state being studied (at energy level ω_1) and emits a photon of frequency $\omega_S = \omega_0 - \omega_1$. (Note that we are using the terms "frequency" and "energy" almost synonymously since they differ only by the multiplicative constant $\hbar = h/2\pi$.) One can think of the process as one in which the laser photon scatters off the molecule, leaving the molecule in an excited state and losing an amount of energy precisely equal to the energy of that excited state. Consequently, from the measured wavelength of the scattered photon and the known wavelength of the incident laser photon, one can determine the energy $\hbar\omega_1$ of the state in question, without probing it with infrared photons of frequency ω_1. The intensity of the Stokes line at ω_1 will be proportional to the population of the ground state. On the other hand, as shown in (b), some molecules may already be in the excited state. A photon of frequency ω_0 may raise such a molecule to a different "virtual" state from which it decays to the ground state, emitting a photon of higher frequency $\omega_a = \omega_0 + \omega_1$. The intensity of the anti-Stokes line at $\omega_0 + \omega_1$ will be proportional to the population of the excited state.



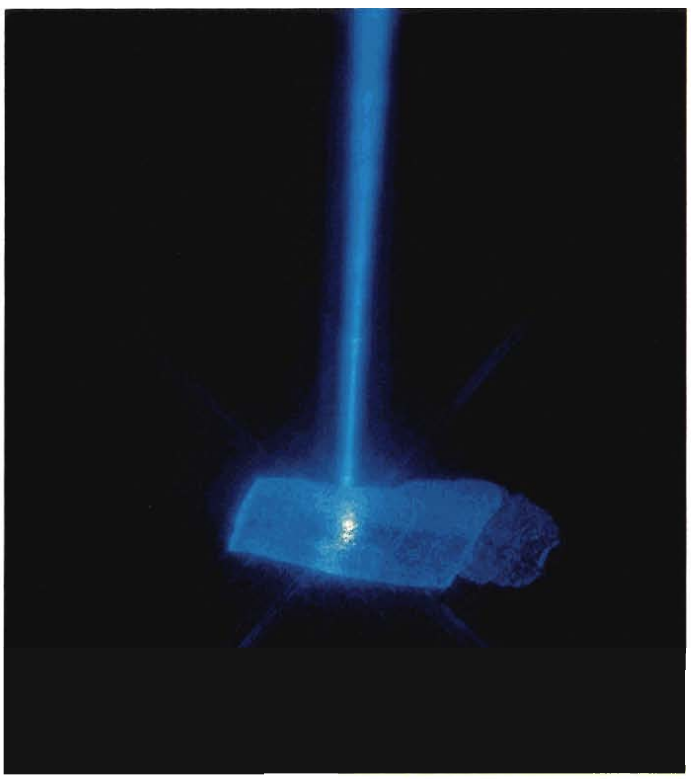


Fig. 16. (a) Experimental setup for a Raman-scattering experiment. Linearly polarized laser light is focused on a sample. Some of the scattered light is collected, usually at 90° to the incoming laser beam, by a lens or mirror. It may be passed through a polarizing filter before entering the spectrometer. (b) A single crystal of ACN scattering laser light in a laser Raman experiment at Los Alamos.

Finally, x-ray diffraction measurements of the crystal structure as a function of temperature show no evidence of structural changes.

On the positive side, if the 1650 cm^{-1} feature does correspond to a soliton, then it should mimic the behavior of the amide-I line as certain experimental parameters are varied. For example, the two lines should have the same polarization. That is, for any crystal orientation and polarization of the probing laser light, the Raman-scattered light corresponding to the line at 1650 cm^{-1} should have the same polarization as that corresponding to the 1665 cm^{-1} line. Our latest measurements (Fig. 17) show that indeed the two lines have the same polarization.

Another positive test would be to substitute carbon-13 for carbon-12 in the carbon-oxygen double bond and measure the energy shift of both lines. Almost identical energy shifts for the 1650-cm^{-1} line and the amide-I line would eliminate the possibility that the 1650-cm^{-1} line arises from some normal mode of the system other than the amide-I vibration. We are in the process of synthesizing an isotopically labeled ACN sample for this experiment.

Davydov estimated the lifetime of the soliton by taking photon-phonon interactions into account in the Hamiltonian. The radiative lifetime of the soliton is expected to be much longer than that of an exciton, or normal vibrational mode. This conjecture may also be tested with Raman spectroscopy.

To understand this, we refer to Fig. 15. In Fig. 15a it is assumed that the molecule is initially in the ground state (the most common situation), and the scattered light has a frequency $\omega_S = \omega_0 - \omega_1$. But if the molecule is initially in an excited state, then the scattering process can cause a transition down to the ground state, as shown in Fig. 15b, and the transition energy will be *added* to the laser photon, yielding a scattered photon of higher frequency $\omega_a = \omega_0 + \omega_1$. For historic reasons the former case, in which energy is lost by the light field to the molecule, is referred to as Stokes Raman scattering, whereas the latter case is called anti-Stokes Raman scattering (hence the subscripts S and a).

Since the intensity of the light scattered at frequencies ω_S and ω_a is directly proportional to the number of molecules (or population) in the ground and excited states, respectively, the ratio of anti-Stokes to Stokes scattered intensities is a direct measure of the ratio of excited to ground state populations:

$$\frac{I_a}{I_S} = \frac{P_1}{P_g} .$$

Normally (that is, under conditions of thermal equilibrium) that population ratio is given by the Boltzmann distribution formula:

$$\frac{P_1}{P_g} = \exp(-\hbar\omega_1/kT) , \qquad (17)$$

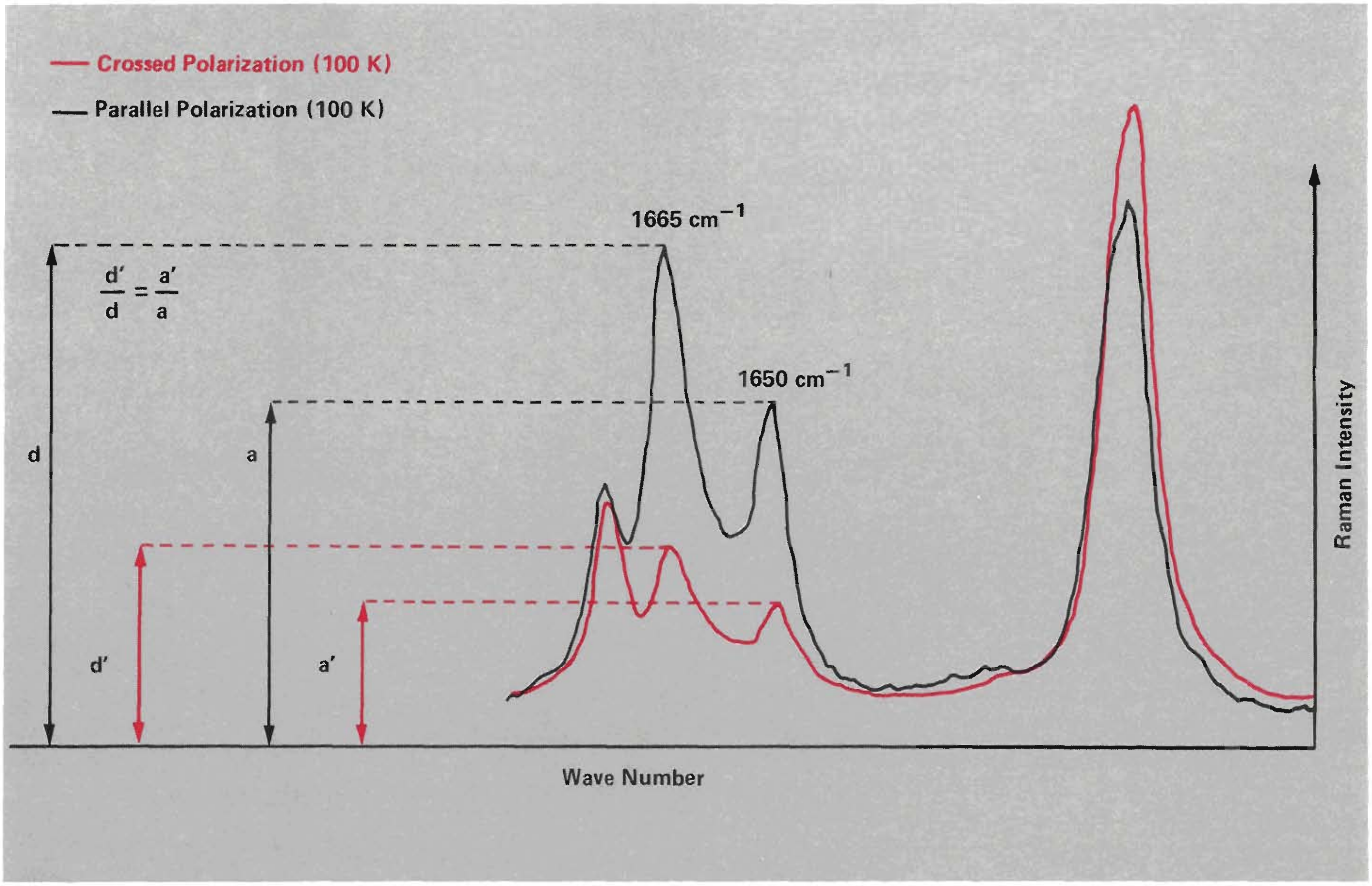


Fig. 17. Laser Raman spectra of ACN at 100 kelvins showing the 1665 cm^{-1} amide-I band and the anomalous 1650 cm^{-1} band for parallel and crossed polarizations. Since the two bands have the same ratio of parallel to crossed polarizations, they have the same polarization. This result is consistent with identification of the 1650 cm^{-1} band as a soliton arising from the self-focusing of amide-I energy.

where $\hbar\omega_1$ is the energy difference between the excited and ground states, k is the Boltzmann constant, and T is absolute temperature.

Since the soliton state is relatively stable and hence long-lived when compared with normal vibrational states of similar energy (those coupled linearly to the system), a significant excitation rate of solitons will result in a nonequilibrium, or nonthermal, population distribution. That is, the population ratio P_1/P_g will be larger than predicted by Eq. 17. Thus one would expect to see an unusually high ratio of anti-Stokes to Stokes scattering intensities when the conditions for exciting solitons exist. Those conditions could be biochemical (for example, hydrolysis of ATP) or physical (direct photoabsorption into the soliton state).

To excite the ACN soliton directly by photoabsorption, we would require a laser that emits radiation at about 1650 cm^{-1}. That wave number is available from one of the higher vibrational/rotational transitions of (what else?!) the CO molecule. We have therefore constructed a tunable electrical-discharge-excited CO laser as the excitation source. The intent is to set up a nonequilibrium population ratio by directly exciting molecules into the soliton "state." We would then expect to see unusually high anti-Stokes intensities for lines corresponding to soliton-coupled resonances.

It may also be possible to measure the lifetime of the soliton state directly. Here the techniques of ultra-fast time-resolved spectroscopy may be useful. A setup basically similar to that shown in Fig. 16 can be used but with very short-pulse excitation sources and fast detectors. The duration of the laser pulse and the temporal resolution of the detector need to be very short: between 1 and 10 picoseconds. Such measurements are being planned.

Brillouin Spectroscopy. Since the propagating soliton entails a moving density fluctuation, similar to that of a sound wave, it might be possible to scatter a photon, essentially elastically, off that fluctuation. The scattered photon would experience a Doppler shift corresponding to the speed and direction of the soliton. Then, if all the molecules of a sample were lined up, as in a crystal, all the scattered photons would experience (plus or minus) the same shift. Since the soliton speed is some fraction of the speed of sound in the crystal, corresponding to Doppler shifts between 0.01 and 0.05 cm^{-1} (300 and 1500 megahertz) the same type of equipment used to detect Brillouin (or sound wave) scattering would be applicable. This entails the use of a special type of Fabry-Perot interferometer, which is really a high-Q resonant optical cavity whose transmission is sensitive to very small changes in wavelength.

Globular Proteins

Our aim in this research is to proceed from experiments on ACN through studies of synthetic α-helical proteins to natural proteins. Although many structural proteins, such as spectrin, tropomyosin, and myosin are almost entirely α helical, there are many other important proteins that are globular. We have seen that the competition between dispersion and focusing of amide-I energy leads—in α-helical proteins or acetanilide crystals—to the formation of a soliton-like object that can travel along the chain of hydrogen-bonded peptide groups without changing its shape. This is essentially a manifestation of the fact that the system has perfect translational symmetry. A natural question to ask is: "What is the result of the competition in globular proteins?" Such proteins do not have translational invariance among the different peptide groups, and soliton formation is not to be expected. However, the mechanisms for dispersion and focusing of amide-I energy are still present.

One way to generalize Davydov's ideas to a globular protein is to take the full geometry of the molecule into account when calculating the dipole-dipole interactions. The size of the J term in the amide-I Hamiltonian (Eq. 1) will vary from peptide group to peptide group, and all possible dipole-dipole interactions have to be considered in order to account for dispersion of amide-I energy. A preliminary computer code, based on assumptions corresponding to those leading to Eq. 13, has been developed for arbitrary protein geometry. This code provides evidence of self-focusing of amide-I vibrational energy in acetanilide (see cover) and in globular proteins. As our physical experiments evolve toward biologically realistic preparations, we plan to make corresponding improvements in this code.

Our present understanding of energy migration in biological systems is very much in its infancy. Our research efforts are directed toward identifying simple but important features in this context. The key scientific question that we have raised in this article may be stated: Is self-trapping of amide-I energy important for transport phenomena in biological materials? Experimental and theoretical studies on model proteins have so far led us to expect an affirmative answer to this question. ■

Further Reading

A. S. Davydov and N. I. Kislukha. "Solitary Excitons in One-Dimensional Molecular Chains." *physica status solidi (b)* 59(1973):465-470.

A. S. Davydov. "The Theory of Contraction of Proteins under Their Excitation." *Journal of Theoretical Biology* 38(1973):559-569.

A. S. Davydov. *Biology and Quantum Mechanics.* Oxford: Pergamon Press Ltd., 1982.

G. Careri. "Search for Cooperative Phenomena in Hydrogen-Bonded Amide Structures." In *Cooperative Phenomena*, H. Haken and M. Wagner, editors (Springer-Verlag, 1973), pp. 391-394.

V. E. Zakharov and A. B. Shabat. "Exact Theory of Two-Dimensional Self-Focusing and One-Dimensional Self-Modulation of Waves in Nonlinear Media." *Soviet Physics JETP* 34(1972):62-69.

J. M. Hyman, D. W. McLaughlin, and A. C. Scott. "On Davydov's Alpha-Helix Solitons." *Physica D* 3D(1981):23-44.

Alwyn C. Scott. "Dynamics of Davydov Solitons." *Physical Review A* 26(1982):578-595.

Alwyn C. Scott. "The Vibrational Structure of Davydov Solitons." *Physica Scripta* 25(1982):651-658.

L. MacNeil and A. C. Scott. "Launching a Davydov Soliton II: Numerical Studies." *Physica Scripta.* In press.

P. S. Lomdahl, L. MacNeil, A. C. Scott, M. E. Stoneham, and S. J. Webb. "An Assignment to Internal Soliton Vibrations of Laser-Raman Lines from Living Cells." *Physics Letters* 92A(1982):207-210.

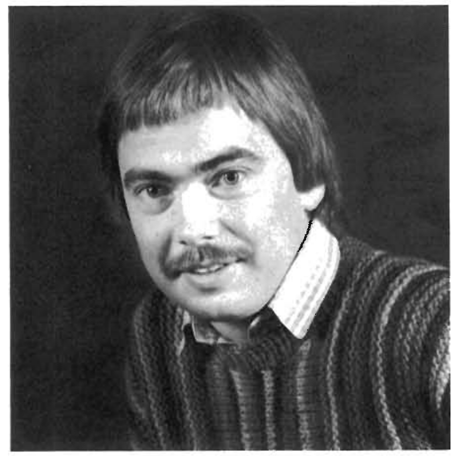
G. Careri, U. Buontempo, F. Carta, E. Gratton, and A. C. Scott. "Infrared Absorption in Acetanilide by Solitons." *Physical Review Letters* 51(1983):304-307.

G. Careri, U. Buontempo, F. Galluzzi, A. C. Scott, E. Gratton, and E. Schyamsunder. "Spectroscopic Evidence for Davydov-Like Solitons in Acetanilide." Los Alamos National Laboratory unclassified release LA-UR-84-483 and submitted to *Physical Review B.*

Peter S. Lomdahl. "Nonlinear Dynamics of. Globular Proteins." Los Alamos National Laboratory unclassified release LA-UR-83-2252 and to be published by Plenum Press in *Nonlinear Electrodynamics in Biological Systems.*

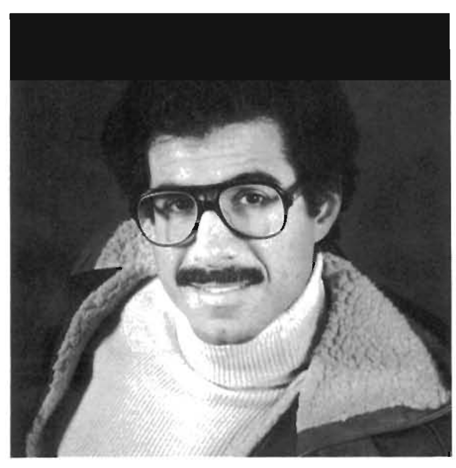
AUTHORS

Peter S. Lomdahl was born and grew up in Copenhagen, Denmark. In 1979 he received an M.S. in electrical engineering and in 1982 a Ph.D. in mathematical physics from the Technical University of Denmark. His graduate work specialized in nonlinear wave phenomena in superconducting Josephson junctions, which is an interest he has continued after joining the Laboratory in 1982 as a postdoctoral fellow with the Center for Nonlinear Studies. The main theme in most of his research has been computer studies of nonlinear dynamics in real condensed matter materials, including conducting polymers and proteins. He has also done work on chaos and has a strong interest in numerical methods for partial differential equations.



Scott P. Layne was born in Chicago, Illinois, in 1954. He received his B.A. in chemistry from DePauw University in 1976 and his M.D. from Case Western Reserve University in 1980. After completing an internship at Loma Linda University in 1981, he joined the Laboratory's Center for Nonlinear Studies as a postdoctoral fellow. Some of his interests include energy transport by biological molecules and the molecular mechanisms of general anesthesia.

Irving J. Bigio received his B.S., M.S., and Ph.D. degrees in physics from the University of Michigan in 1969, 1970, and 1974, respectively. His doctoral work under John Ward and Peter Franken dealt with nonlinear optics, and he has maintained a broad interest in the field of quantum electronics ever since. He came directly to Los Alamos in April 1974 as a staff member in the laser isotope separation program and has also worked in the laser fusion program. In 1976 he received a Fulbright Senior Scholar Award and spent the 1976-77 academic year as a visiting professor at the Weizmann Institute of Science, Rehovot, Israel, where he taught graduate courses in laser physics and nonlinear optics and helped direct graduate student research. Since returning to Los Alamos he has resumed his research and has taught courses at the University of New Mexico Graduate Center. Currently, he is working on a variety of topics in quantum electronics and has taken an interest in the application of laser techniques and nonlinear optics to the solution of biophysics problems. He was recently appointed Deputy Leader of the Discharge Lasers Group.



A Possible Mechanism for General Anesthesia

by Scott P. Layne

RELATED TOPICS

The first general anesthesia for human surgery was administered at the Massachusetts General Hospital in Boston in 1846. The patient was put to sleep by breathing diethyl ether from a glass vesicle, and the surgeon quickly dissected a tumor located under the jaw. After completing the operation the surgeon remarked to his audience, "Gentlemen, this is no humbug."

Since this first successful demonstration of diethyl ether, researchers have discovered well over twenty drugs that induce general anesthesia. These drugs have highly diverse chemical structures and physical properties and, as a whole, lend little insight into their mechanism of action. In order to overcome this perplexity, H. Meyer and E. Overton (about the year 1900) originally proposed that anesthetic potency could be related to lipid solubility. They showed that stronger anesthetic agents were more oil-soluble than weaker ones and used this relationship to argue that anesthetics insert into the lipid bilayer and thereby expand its volume. More recent theories along this line have suggested that the expanded lipid bilayer compresses intrinsic membrane proteins and thereby disturbs normal protein shape and function. These theories have suggested also that the membrane-bound anesthetic molecules "fluidize" the lipid bilayer. This increased fluidity, in turn, alters the permeability of the membrane. While these popular ideas might be applicable to agents that are both volatile and highly lipid-soluble (oil-to-gas partition coefficient $\geq$100:1), they are not particularly suitable to a large class of *intravenous* general anesthetics that are orders of magnitude less lipid-soluble and are capable of forming hydrogen bonds. For the case of hydrogen-bonding anesthetic agents, the simplest idea is that they act by binding directly to a particularly sensitive protein, which may or may not be located in a lipid membrane, and inhibiting its normal function.

In this discussion we will focus on an important class of intravenous general anesthetics that are only slightly lipid-soluble and are capable of forming hydrogen bonds. These agents are represented primarily by *barbiturates*. From Fig. 1 it is easy to see that a barbiturate contains four H–N–C=O groups in its ring. These H–N–C=O groups are very similar to the peptide groups in proteins that are important to the propagation of solitons (see "Solitons in Biology"). The other drugs shown in Fig. 1 also contain H–N–C=O groups but to a lesser extent than barbiturates. Hydantoins contain three peptide groups, glutethimides and succinimides contain two, and urethanes contain one. These drugs are not used as general anesthetics per se, but they nevertheless have a similar inhibitory effect on the central nervous system. The potency of these six drugs appears to be related directly to the number of H–N–C=O groups in the molecule. This is supported by the fact that N-methylated barbiturates (which contain two H–N–C=O groups) are shorter acting and less potent than nonmethylated barbiturates and that trimethadione (which is devoid of H–N–C=O groups) is inactive until

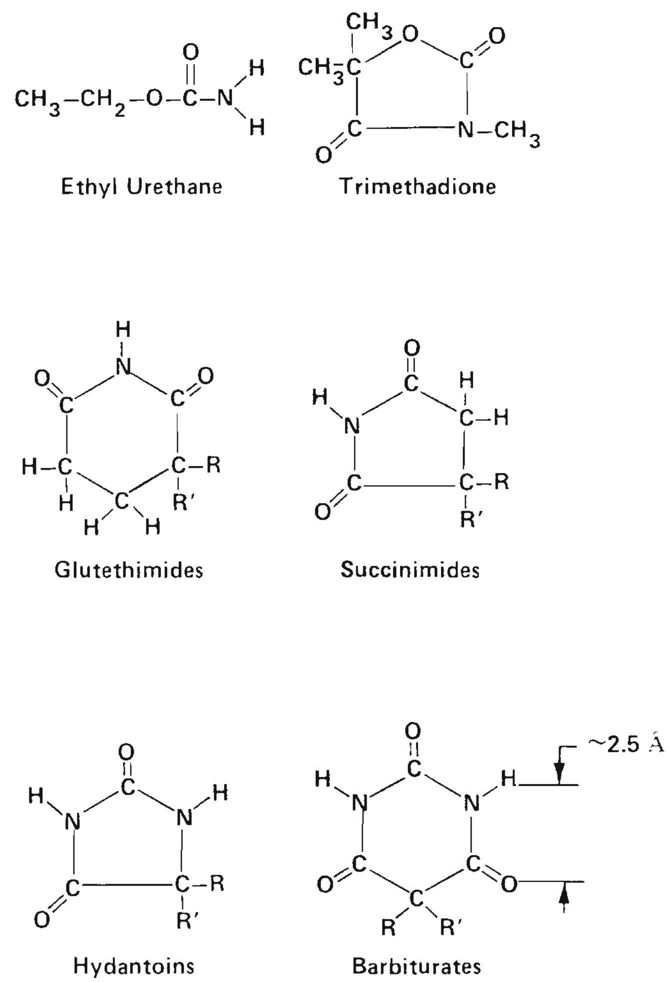


Fig. 1. The six drugs shown above, all of which contain H–N–C=O groups, inhibit the central nervous system. Hydantoins, succinimides, and trimethadione are used primarily as antiepileptic agents, whereas glutethimides are used as sedatives. Ethyl urethane is a common veterinary general anesthetic but is not used in humans because its actions are not smooth. The presence of an alkyl or aryl group at R *and* R′ *confers increasing lipid solubility, and, generally, increased lipid solubility promotes an increased drug potency.*

it is demethylated by hepatic enzymes. After demethylation, trimethadione contains two H–N–C=O groups.

C. Sandorfy and coworkers have shown by infrared spectroscopy that barbiturates are capable of dissociating hydrogen bonds in the 1-cyclohexyluracil/9-ethyladenine dimer. This dimer forms hydrogen bonds of the N–H ···· O=C type that is common to proteins. They have also shown that barbiturates form hydrogen bonds with solutions of N,N-dimethylacetamide (NNDA) and N-methylacetamide (NMA). In this instance the N–H groups of barbiturates act as proton donors, and the O=C groups on NNDA and NMA act as proton acceptors. From these data we can infer that barbiturates are capable of forming hydrogen bonds with proteins, and, for the case of α-helical proteins, this bonding might take the form shown in Fig. 2. Note that this type of two-point hydrogen bonding along a spine of the α helix has half the chance of taking place if an N–H group in the barbiturate ring is replaced by an N–CH_3 group.

How does the binding of an anesthetic molecule to a protein modify normal protein behavior? We shall answer this question using the soliton model as a paradigm for *normal* protein function. The soliton model proposes that α-helical proteins effect the transport of ATP hydrolysis energy through a coupling of vibrational excitations to displacements along the spines of the helix. This coupling leads to a self-focusing of vibrational energy that has remarkably stable qualities (see "What Is a Soliton?"). We suggest that the binding of an anesthetic molecule to a protein interferes with soliton propagation. We suggest further that this type of interference would be most important in two separate regions of a cell where soliton propagation is an attractive candidate: first, in the α-helical proteins of the inner mitochondrial membrane, which appear to participate in ATP synthesis and electron transport, and second, in the membrane proteins of neurons, which are responsible for chemical reception and signal transduction. This proposal is motivated by the fact that barbiturates are capable of binding to these sites and further by the fact that these proteins have significant α-helical character. To see whether this idea makes sense from a theoretical standpoint, we need to calculate the effect of anethestic binding on soliton propagation.

When a barbiturate binds to an α helix, it will form new hydrogen bonds between anesthetic and protein molecules at the expense of the protein's hydrogen bond(s). This kind of anesthetic binding will result in either broken hydrogen bonds within the protein or in weakened hydrogen bonds of increased length; we shall call this increase ΔR. We assumed for the numerical investigation that the hydrogen bonds within the protein are merely weakened and are not completely broken. We chose for ΔR a value of 0.8 angstrom, which corresponds roughly to a decrease in hydrogen-bond energy of 55 percent. It is straightforward to calculate the new dipole-dipole interaction energy J, if we assume that the two dipoles within the protein remain colinear. The decrease in J will be proportional to $(R+\Delta R)^{-3}$. However, it is not

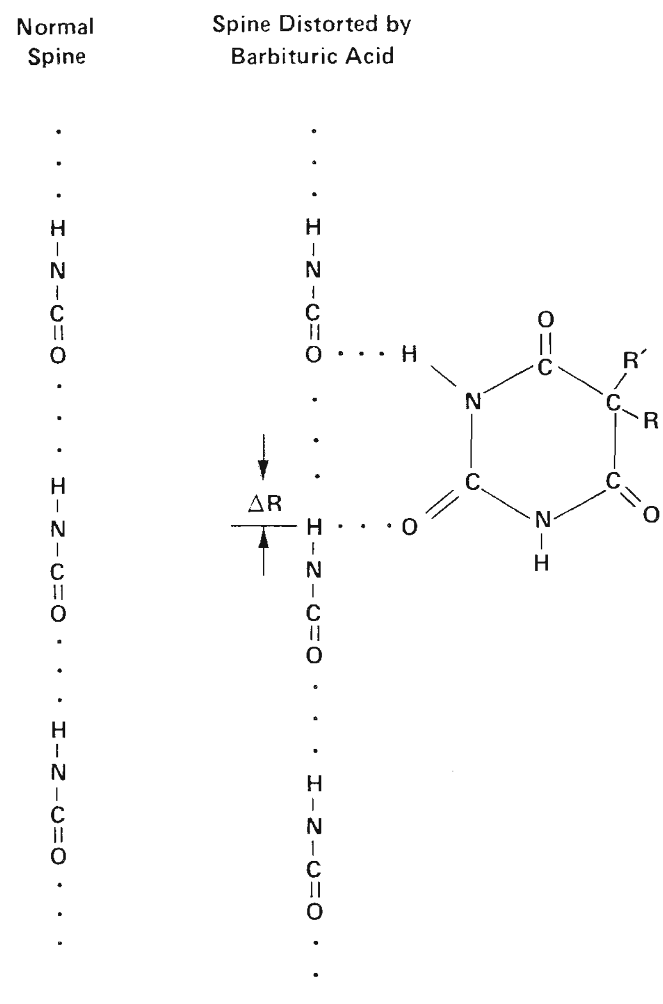


Fig. 2. A possible interaction of a barbiturate, via its H–N–C=O moieties, with one spine of an α-helical protein. The spiral configuration of the protein is stabilized by its weak hydrogen bonds, and the binding of a barbiturate changes the localized structure within the helix. In this instance, the hydrogen bond is weakened and its bond length increases by the distance ΔR.

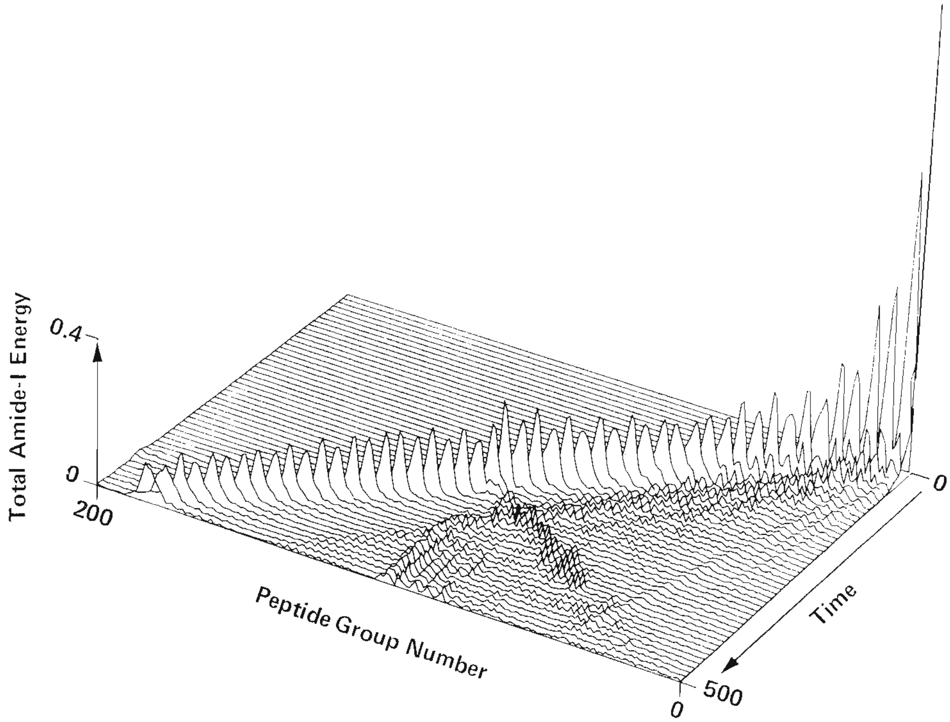


Fig. 3. Numerical calculation simulating the perturbation of a soliton by an anesthetic agent. The perturbation involves changes in the values of J, K, *and* χ *for peptide groups 100 to 103. The total amide-I energy is plotted as a function of peptide group number and time. Notice that the soliton loses amplitude and widens by the time it reaches the end of the helix.*

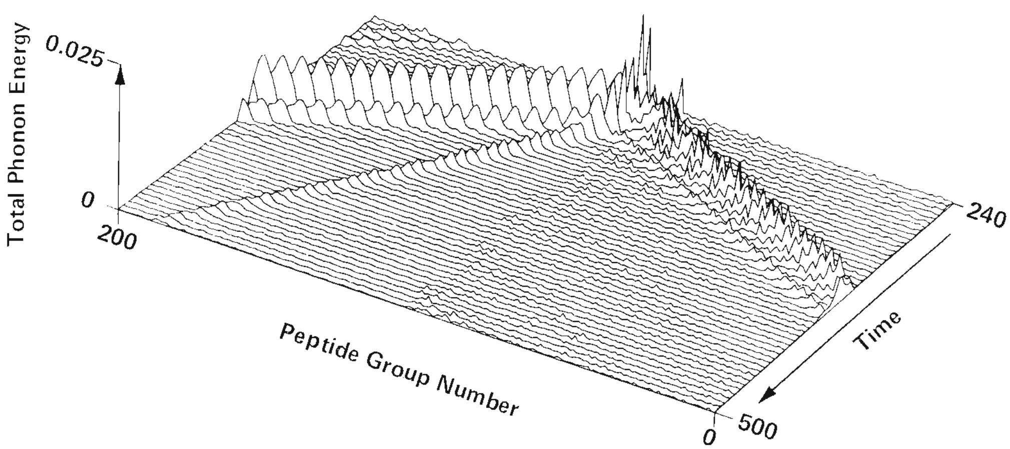


Fig. 4. This figure is the counterpart to Fig. 3. The total phonon energy is plotted as a function of peptide group number and time. In this view time has been restricted to the interval between 240 and 500 computer time units. At time 250 the soliton is just entering the region of perturbation. It radiates energy in the form of phonons as it travels through the altered peptide groups. After emerging from the region of perturbation, the soliton is seen as the low-amplitude wave, which moves at about three-eighths of the sound speed. Note that the phonon energy of the soliton is small compared to its bond energy.

as easy to calculate a new value for the hydrogen-bond spring constant K nor a new value for the coupling constant χ in weaker hydrogen bonds. As a crude estimate we assumed that K decreased proportionally to hydrogen-bond energy, and thus our new spring constant has the value of $0.45K$. We also assumed that χ is slightly decreased in weaker hydrogen bonds to the lower value that was calculated by V. Kuprievich and Z. Kudritskaya. Hence, at the point of anesthetic binding we chose $\chi = 0.3 \times 10^{-10}$ newton, which is just below threshold for soliton formation.

The results of this numerical investigation are presented in Figs. 3 and 4. The decreased values of J, K, and χ were restricted to peptide group numbers $n = 100$ to 103 on the three spines of the α helix. The perturbation was restricted to this narrow region because an anesthetic molecule is expected to weaken the hydrogen bonds in only a small region of the protein. This procedure also ensured that the soliton was well formed before entering the perturbed region. Figure 3 can be compared directly to Fig. 10 in "Solitons in Biology." It is apparent that after 500 computer time units the soliton, which traveled through the perturbation, is appreciably degraded. Figure 4 reveals that energy is radiated by the soliton in the form of phonons as it travels through the perturbation. These phonons are seen to move at the sound velocity in the α helix, which is approximately eight-thirds the soliton velocity. Up to this point we have neglected the fact that the H–N–C=O groups in the barbiturate are capable of dipole-dipole coupling to the H–N–C=O groups in the helix. Such a coupling should further degrade soliton propagation, since the interaction energy between barbiturate and α helix would be appreciable. The dipole-dipole coupling of the anesthetic molecule to the protein will depend on the number of H–N–C=O groups within it and on its spatial orientation relative to the protein.

As a final consideration of this model we pose the question: How many proteins are inhibited during general anesthesia? Barbiturates exhibit their anesthetic activity at a concentration between 200 and 1000 micromolar. At this concentration they reduce the metabolic activity of the brain by 10 to 15 percent, as measured by oxygen utilization. Taking the average membrane protein to encompass a volume of 20 angstroms × 20 angstroms × 40 angstroms = 1.6×10^4 cubic angstroms implies that about 1 percent of typical membrane proteins are associated with an anesthetic molecule. Such a small figure points out that the brain is very sensitive to alterations at the molecular level. Consciousness appears to require the coordinated effort of almost every protein.

We have presented a simplified theoretical model for anesthetic activity, taking advantage of the fact that the α helix is an important structure in membrane and cytoskeletal proteins. If the Davydov soliton finds experimental support in biology, then such a model may help to explain some of the molecular mechanisms behind general anesthesia. ■

Acknowledgment

I wish to thank Peter Lomdahl for help with the numerical code.

Further Reading

R. Buchet and C. Sandorfy. "Perturbation of the Hydrogen-Bond Equilibrium in Nucleic Bases. An Infrared Study." *Journal of Physical Chemistry* 87(1983):275-280.

N. P. Franks and W. R. Lieb. "Molecular Mechanisms of General Anaesthesia." *Nature* 300(1982):487-493.

M. Guerin, J.-M. Dumas, and C. Sandorfy. "Vibrational Spectroscopic Studies of Molecular Associations by Local Anesthetics." *Canadian Journal of Chemistry* 58(1980):2080-2088.

Scott P. Layne. "The Modification of Davydov Solitons by the Extrinsic H–N–C=O Group." Los Alamos National Laboratory unclassified release LA-UR-83-2253 and to be published by Plenum Press in *Nonlinear Electrodynamics in Biological Systems.*

RELATED TOPICS

What Is a Soliton?

by Peter S. Lomdahl

About thirty years ago a remarkable discovery was made here in Los Alamos. Enrico Fermi, John Pasta, and Stan Ulam were calculating the flow of energy in a one-dimensional lattice consisting of equal masses connected by nonlinear springs. They conjectured that energy initially put into a long-wavelength mode of the system would eventually be "thermalized," that is, be shared among all modes of the system. This conjecture was based on the expectation that the nonlinearities in the system would transfer energy into higher harmonic modes. Much to their surprise the system did not thermalize but rather exhibited energy sharing among the few lowest modes and long-time near recurrences of the initial state.

This discovery remained largely a mystery until Norman Zabusky and Martin Kruskal started to investigate the system again in the early sixties. The fact that only the lowest order (long-wavelength) modes of the discrete Fermi-Pasta-Ulam lattice were "active" led them in a continuum approximation to the study of the nonlinear partial differential equation

$$\frac{\partial u}{\partial t} + u\,\frac{\partial u}{\partial x} + \frac{\partial^3 u}{\partial x^3} = 0\ . \tag{1}$$

This equation (the KdV equation) had been derived in 1885 by Korteweg and de Vries to describe long-wave propagation on shallow water. But until recently its properties were not well understood.

From a detailed numerical study Zabusky and Kruskal found that stable pulse-like waves could exist in a system described by the KdV equation. A remarkable quality of these solitary waves was that they could collide with each other and yet preserve their shapes and speeds after the collision. This particle-like nature led Zabusky and Kruskal to name such waves *solitons*. The first success of the soliton concept was explaining the recurrence in the Fermi-Pasta-Ulam system. From numerical solution of the KdV equation with periodic boundary conditions (representing essentially a ring of coupled nonlinear springs), Zabusky and Kruskal made the following observations. An initial profile representing a long-wavelength excitation would "break up" into a number of solitons, which would propagate around the system with different speeds. The solitons would collide but preserve their individual shapes and speeds. At some instant all of the solitons would collide at the same point, and a near recurrence of the initial profile would occur.

This success was exciting, of course, but the soliton concept proved to have even greater impact. In fact, it stimulated very important progress in the analytic treatment of initial-value problems for nonlinear partial differential equations describing wave propagation. During the past fifteen years a rather complete mathematical description of solitons has been developed. The amount of information on nonlinear wave phenomena obtained through the fruitful collaboration of mathematicians and physicists using this description makes the soliton concept one of the most significant developments in modern mathematical physics.

The nondispersive nature of the soliton solutions to the KdV equation arises not because the effects of dispersion are absent but because they are balanced by nonlinearities in the system. The presence of both phenomena can be appreciated by considering simplified versions of the KdV equation.

Eliminating the nonlinear term $u(\partial u/\partial x)$ yields the linearized version

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

The most elementary wave solution of this equation is the harmonic wave

$$u(x,t) = A \exp\left[i(kx + \omega t)\right]\ , \tag{3}$$

where k is the wave number and ω is the angular frequency. In order for the displacement $u(x,t)$ given by Eq. 3 to be a solution of Eq. 2, ω and k must satisfy the relation

$$\omega = k^3 \ . \tag{4}$$

Such a "dispersion relation" is a very handy algebraic description of a linear system since it contains all the characteristics of the original differential equation. Two important concepts connected with the dispersion relation are the phase velocity $v_p = \omega/k$ and the group velocity $v_g = \partial\omega/\partial k$. (For the dispersion relation given by Eq. 4, $v_p = k^2$ and $v_g = 3k^2$). The phase velocity measures how fast a point of constant phase is moving, while the group velocity measures how fast the energy of the wave moves. The waves described by Eq. 2 are said to be dispersive because a wave with large k will have larger phase and group velocities than a wave with small k. Therefore, a wave composed of a superposition of elementary components with different wave numbers (different values of k in Eq. 3) will disperse, or change its form, as it propagates.

Now we eliminate the dispersive term $\partial^3u/\partial x^3$ and consider the equation

$$\partial u/\partial t + u\partial u/\partial x = 0 \ . \tag{5}$$

This simple nonlinear equation also admits wave solutions, but they are now of the form $u(x,t) = f(x - ut)$, where the function f is arbitrary. (That $f(x - ut)$ is a solution of Eq. 5 is easily verified by substitution.) For waves of this form, the important thing to note is that the velocity of a point of constant displacement u is equal to that displacement. As a result, the wave "breaks"; that is, portions of the wave undergoing greater displacements move faster than, and therefore overtake, those undergoing smaller displacements. This multivaluedness is a result of the nonlinearity and, like dispersion, leads to a change in form as the wave propagates.

A remarkable property of the KdV equation is that dispersion and nonlinearity balance each other and allow wave solutions that propagate without changing form (Fig. 1). An example of such a solution is

$$u(x,t) = 3c\ \mathrm{sech}^2[c^{1/2}(x - ct)/2] \ , \tag{6}$$

where the velocity c can take any positive value. This is the one-soliton solution of the KdV equation.

Although our discussion may have provided some glimpse of the interplay between dispersion and nonlinearity in the KdV equation, it has not, of course, provided any explanation of how solitons preserve their shapes and speeds after collision. This particle-like property is more than just a mere curiosity; it is of deep mathematical significance. A full understanding of this property requires an extensive mathematical discussion that we will not attempt here. We mention, however, that not all nonlinear partial differential equations have soliton solutions. Those that do are generic and belong to a class for which the general initial-value problem can be solved by a technique called the inverse scattering transform, a brilliant scheme developed by Kruskal and his coworkers in 1967. With this method, which can be viewed as a generalization of the Fourier transform to nonlinear equations, general solutions can be produced through a series of linear calculations. During the solution process it is possible to identify new nonlinear modes—generalized Fourier modes—that are the soliton components of the solution and, in addition, modes that are purely dispersive and therefore often called radiation. Equations that can be solved by the inverse scattering transform are said to be completely integrable.

The manifestation of balance between dispersion and nonlinearity can be quite different from system to system. Other equations thus have soliton solutions that are distinct from the bell-shaped solitons of the KdV equation. An example is the so-called nonlinear Schrödinger (NLS) equation. This equation is generic to all conservative systems that are weakly nonlinear but strongly dispersive. It describes the slow temporal and spatial evolution of the envelope of an almost monochromatic wave train. We present here a heuristic derivation of the NLS equation that shows how it is the natural equation for the evolution of a carrier-wave envelope. Consider a dispersion relation for a harmonic wave that is amplitude dependent:

$$\omega = \omega(k,|E|^2) \ . \tag{7}$$

Here $E = E(x,t)$ is the slowly varying envelope function of a modulated wave with carrier frequency ω and wave number k. The situation described by Eq. 7 occurs, for example, in nonlinear optical phenomena, where the dielectric constant of the medium depends on the intensity of the electric signal. Other examples include surface waves on deep water, electrostatic plasma waves, and bond-energy transport in proteins.

By expanding Eq. 7 in a Taylor's series about ω_0 and k_0, we obtain

$$\omega - \omega_0 = \left.\frac{\partial\omega}{\partial k}\right|_0 (k - k_0) + \frac{1}{2}\left.\frac{\partial^2\omega}{\partial k^2}\right|_0 (k - k_0)^2 + \left.\frac{\partial\omega}{\partial(|E|^2)}\right|_0 |E|^2 \ . \tag{8}$$

RELATED TOPICS

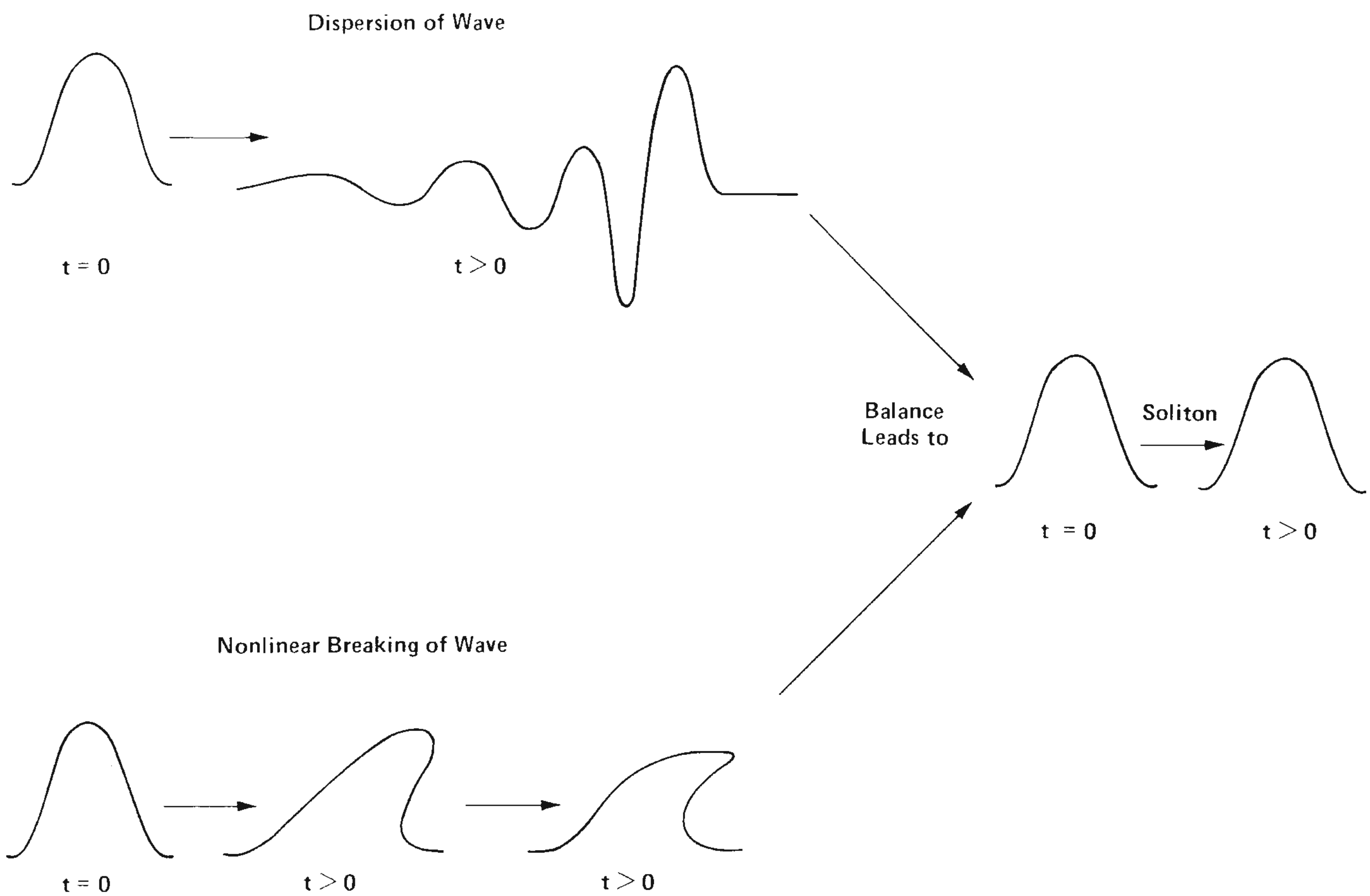


Fig. 1. Two effects, dispersion and breaking, cause the shape of a wave to change as it propagates. For a wave described by the KdV equation, these two effects balance, and the wave—a soliton—propagates without changing shape.

We have expanded only to first order in the nonlinearity but to second order in the dispersion because the first-order dispersion term, as we shall see, only represents undistorted propagation of the wave with the group velocity $v_g = [\partial\omega/\partial k]_0$. If we now substitute the operators $i(\partial/\partial t)$ for $\omega - \omega_0$ and $-i(\partial/\partial x)$ for $k - k_0$ in Eq. 8 and let the resulting expression operate on E, we get

$$i\left[\frac{\partial E}{\partial t} + \left.\frac{\partial\omega}{\partial k}\right|_0 \frac{\partial E}{\partial x}\right] + \frac{1}{2}\left.\frac{\partial^2\omega}{\partial k^2}\right|_0 \frac{\partial^2 E}{\partial x^2} - \left.\frac{\partial\omega}{\partial(|E|^2)}\right|_0 |E|^2 E = 0\ . \tag{9}$$

This is the nonlinear Schrödinger equation, so called because of its resemblance to the Schrödinger equation even though its derivation often has nothing to do with quantum mechanics. The first term of Eq. 9 represents undistorted propagation of the wave at the group velocity, and the second and third terms represent its linear and nonlinear distortion, respectively. This crude derivation of the NLS equation shows how it arises in systems with amplitude-dependent dispersion relations, but more formal methods are necessary if detail about the coefficients, such as $[\partial\omega/\partial(|E|^2)]_0$, is required.

It is often preferable to express Eq. 9 in a neater form. For this purpose we transform the variables x and t into z and τ, where $z = x - [\partial\omega/\partial k]_0 t$ is a coordinate moving with the group velocity and $\tau = 1/2[\partial^2\omega/\partial k^2]_0 t$ is the normalized time. Equation 9 then reduces to

$$i\frac{\partial E}{\partial \tau} + \frac{\partial^2 E}{\partial z^2} + 2\kappa|E|^2 E = 0 , \qquad (10)$$

where

$$\kappa = -[\partial\omega/\partial(|E|^2)]_0/[\partial^2\omega/\partial k^2]_0 . \qquad (11)$$

The NLS equation—like the KdV equation—is completely integrable and has soliton solutions. The analytic form for a single-soliton solution is given by

$$E(z,\tau) = 2\eta\ \mathrm{sech}[2\eta(\theta_0 - \eta z - 4\xi\eta\tau)]$$
$$\times \exp\{-2i[\phi_0 + 2(\xi^2 - \eta^2)t + \xi z]\} , \qquad (12)$$

where ξ, η, θ_0, and ϕ_0 are free parameters determining the speed, amplitude, initial position, and initial phase, respectively, of the soliton. Figure 2 shows the profile of this soliton.

Any initial excitation for the NLS equation will decompose into solitons and/or dispersive radiation. A monochromatic wave train solution $E(z,\tau) = E(\tau)$ is thus unstable to any z-dependent perturbation and breaks up into separate and localized solitons. This phenomenon is called the Benjamin-Feir instability and is well known to any surfer on the beach who has noticed that every, say, seventh wave is the largest. The NLS equation is in a way more universal than the KdV equation since an almost monochromatic, small-amplitude solution of the KdV equation will evolve according to the NLS equation.

The last type of soliton we mention, which is distinctly different from the KdV or NLS solitons, is one that represents topologically invariant quantities in a system. Such an invariant can be a domain wall or a dislocation in a magnetic crystal or a shift in the bond-alternation pattern in a polymer. The prototype of equations for such solitons is the sine-Gordon equation,

$$\frac{\partial^2 u}{\partial x^2} - \frac{\partial^2 u}{\partial t^2} = \sin u . \qquad (13)$$

Notice that this equation allows for an infinite number of trivial

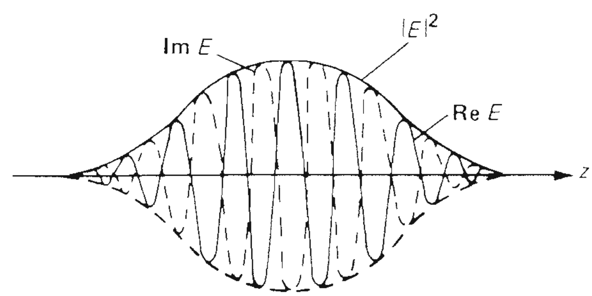


Fig. 2. Profile of a single-soliton solution of the NLS equation.

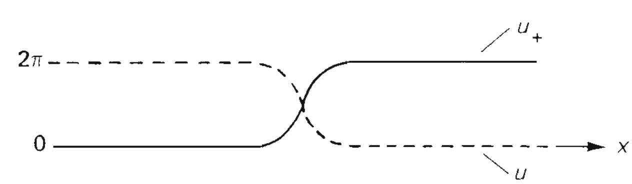


Fig. 3. Profiles of soliton solutions of the sine-Gordon equation.

solutions, namely $u = 0, \pm 2\pi, \pm 4\pi, \ldots$. Systems with a multitude of such "degenerate ground states" also allow solutions that connect two neighboring ground states. Solutions of this type are often called kinks, and for the sine-Gordon equation they are exact solitons; that is, they collide elastically without generation of dispersive radiation. The analytic form, whose profile is shown in Fig. 3, is given by

$$u_{\pm}(x,t) = 4 \tan^{-1}\{\exp[\pm(x - x_0 - ct)/(1 - c^2)^{1/2}]\} , \qquad (14)$$

where the solution u_- is often called an antikink. The parameter c $(-1 < c < 1)$ determines the velocity and x_0 the initial position. Other equations with degenerate ground states also have kink and antikink solutions, but they are not exact solitons like those of the sine-Gordon equation. It is interesting to note that small-amplitude solutions of the sine-Gordon equation also can be shown to evolve according to the NLS equation.

Equations with soliton solutions are generic, and, although real systems often contain mechanisms (impurities, dissipative forces, and multidimensionality) that destroy exact soliton behavior, they are very useful as a starting point for analysis. In fact, perturbation methods—with the perturbation taking place around the soliton—have been developed to compute the response of the soliton to external forces, damping, etc. Often the result is that the parameters characterizing the soliton (such as velocity and amplitude) are now time dependent, with the time dependence governed by simple ordinary differential equations. The original equations are therefore still very useful. Because the mechanisms that give rise to soliton equations are so prevalent, the suggestion that solitons might arise in biology is not so surprising. The question to be asked is how well a particular biological system satisfies the criteria underlying the soliton equation. ■

Further Reading

The classic paper where the word "soliton" was introduced is "Interaction of 'Solitons' in a Collisionless Plasma and the Recurrence of Initial States" by N. J. Zabusky and M. D. Kruskal in *Physical Review Letters* 15(1965):240. For many references see also "Computational Synergetics and Mathematical Innovation" by Norman J. Zabusky in *Journal of Computational Physics* 43(1981):95.

There are an increasing number of papers on solitons; a good review paper covering the subject to 1973 is "The Soliton: A New Concept in Applied Science" by Alwyn C. Scott, F. Y. F. Chu, and David W. McLaughlin in *Proceedings of the IEEE* 61(1973):1443.

Good accounts of the subject, together with up-to-date lists of references, can also be found in many textbooks, including the following.

G. L. Lamb, Jr. *Elements of Soliton Theory*. New York: John Wiley & Sons, 1980.

Mark J. Ablowitz and Harvey Segur. *Solitons and the Inverse Scattering Transform*. Philadelphia: Society for Industrial and Applied Mathematics, 1981.

R. K. Dodd, J. C. Eilbeck, J. D. Gibbon, and H. C. Morris. *Solitons and Nonlinear Wave Equations*. New York: Academic Press, 1982.

Photography by Malcolm Lockwood

Journeys of a Spacecraft

by John T. Gosling, Daniel N. Baker, and Edward W. Hones, Jr.

With an assist from the moon, a spacecraft called ISEE-3 traveled far out along the nightside tail of our earth's magnetic field. There it discovered, among other things, huge ionic plasmoids breaking off from the earth's magnetosphere and hurtling out into space.

The earth's magnetic field acts in many ways like a buffer between us and space. Over and around this field flows the solar wind, a dilute but persistent stream of protons, electrons, and other ions. This flow of charge, with its associated weak magnetic field, distorts the earth's own field, compressing it on the dayside and stretching it out on the nightside of the planet into a long tail much like that of a comet. The resulting field is called the magnetosphere.

How static is this shield? Does the solar wind flow by passively or does it deposit mass and energy within the magnetosphere that must eventually be released in sudden, dramatic bursts? The auroral disturbances known as the northern and southern lights have always been evidence for the latter point of view. But how extensive are the auroral disturbances? Does our magnetosphere dance with invisible storms that reach far out into space?

This aurora, the result of an injection of solar wind plasma into the upper atmosphere, displays the greenish color of an emission from excited oxygen at a wavelength of 5577 angstroms.

One model pictures the magnetosphere to be something like a drippy faucet (Fig. 1). Solar wind plasma seeps into the magnetosphere all along its boundary, accumulating in the tail until a portion breaks off like a swollen drop of water from a faucet. This model requires the operation of a process called *magnetic reconnection* in which different sets of field lines come together, breaking and reconnecting in new configurations. Reconnection permits the transfer of mass and energy from the solar wind into the magnetosphere, causing it to swell and distort, and is also responsible for the pinching off of a portion of the distended geomagnetic field.

How could this or other models be tested? Satellites in the near-earth tail have collected many useful data concerning magnetospheric processes, but, in one sense, the instruments aboard these satellites may have been blinded by the very processes they were trying to measure. It is as if the firing of a cannon was being studied by sitting at the point of the initial explosion! Data needed to be collected further out—where one could actually see the "cannon balls" hurtling by.

The possibility of obtaining such data was realized when orbits were discovered for the ISEE-3 spacecraft (the third in a series of International Sun-Earth Explorers) that used the moon's gravity to take it on several journeys down the geomagnetic tail. These orbits carried the spacecraft to distances far beyond the orbit of the moon into previously unexplored regions of the magnetosphere. Because ISEE-3 carried a Los Alamos instrument designed to measure electron velocity distributions in space, the data collected on these journeys were able to show that magnetic reconnection plays a critical role in the behavior of our magnetosphere.

It is now felt that the interaction between the solar wind and the magnetosphere does indeed resemble a drippy faucet. The solar wind injects plasma into our magnetosphere along reconnected field lines. When a critical amount of plasma and energy have built up within the tail, field lines in the central portion of the tail pinch off, again by the process of magnetic reconnection, and a bundle of plasma and field lines called a plasmoid suddenly forms and shoots out into interplanetary space. Plasma and field lines inside the breakage point snap back toward the earth, injecting charged particles into our atmosphere and causing an auroral disturbance. Thus, auroral disturbances are only one aspect of a much more widespread disturbance involving the entire magnetosphere.

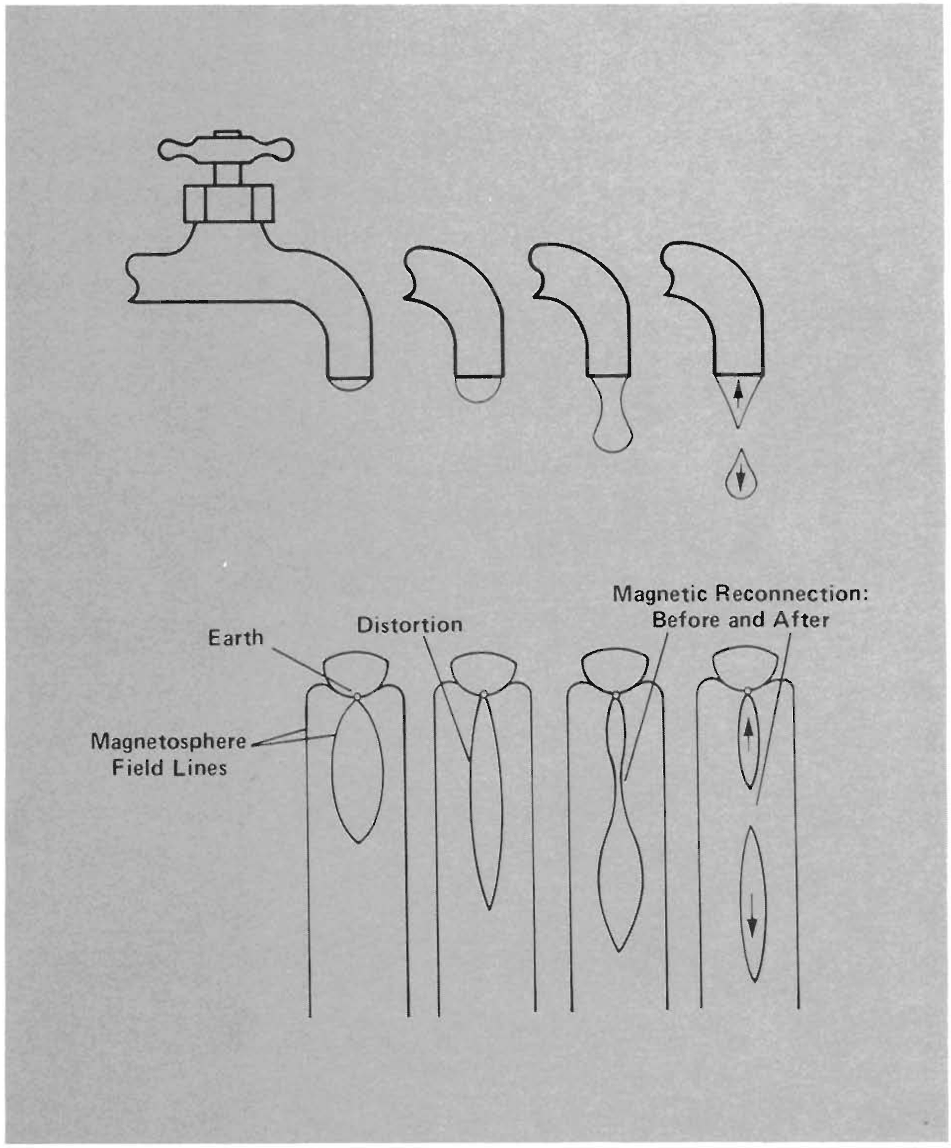


Fig. 1. The drippy-faucet model of the magnetosphere in which (lower series) solar wind plasma is added to the magnetosphere, causing part of it to swell and distort until magnetic reconnection of field lines allows one portion to break away and another portion to spring back toward the earth.

Models of the Magnetosphere

The earth's magnetic field originates from electric currents flowing within the liquid metallic core of the earth. Before the solar wind was taken into account, this field was pictured as an undistorted magnetic dipole (part (a) of Fig. 2).

The first modification to the dipole model occurred after the existence of the solar wind was suggested (see "The Solar Wind") but before satellites had begun to explore the distant reaches of space around the earth and record the properties of the solar wind. Many scientists thought that the continuous flow of the solar wind plasma might force the earth's magnetic field into the shape depicted in part (b) of Fig. 2. Both this model, proposed by Francis Johnson in 1960, and the dipole model are *closed* magnetospheres, that is, ones in which all the field lines leaving one hemisphere of the earth return to the other hemisphere.

Early in the sixties satellites measured for the first time the magnetic field of the solar wind. In 1961 James Dungey proposed that the interplanetary field lines carried by the solar wind might interconnect with the terrestrial field lines, creating the field geometry sketched in part (c) of Fig. 2. This is an *open* magnetosphere because field lines from high latitudes on the earth do not return to the other hemisphere. Instead, they are connected to the field of the solar wind.

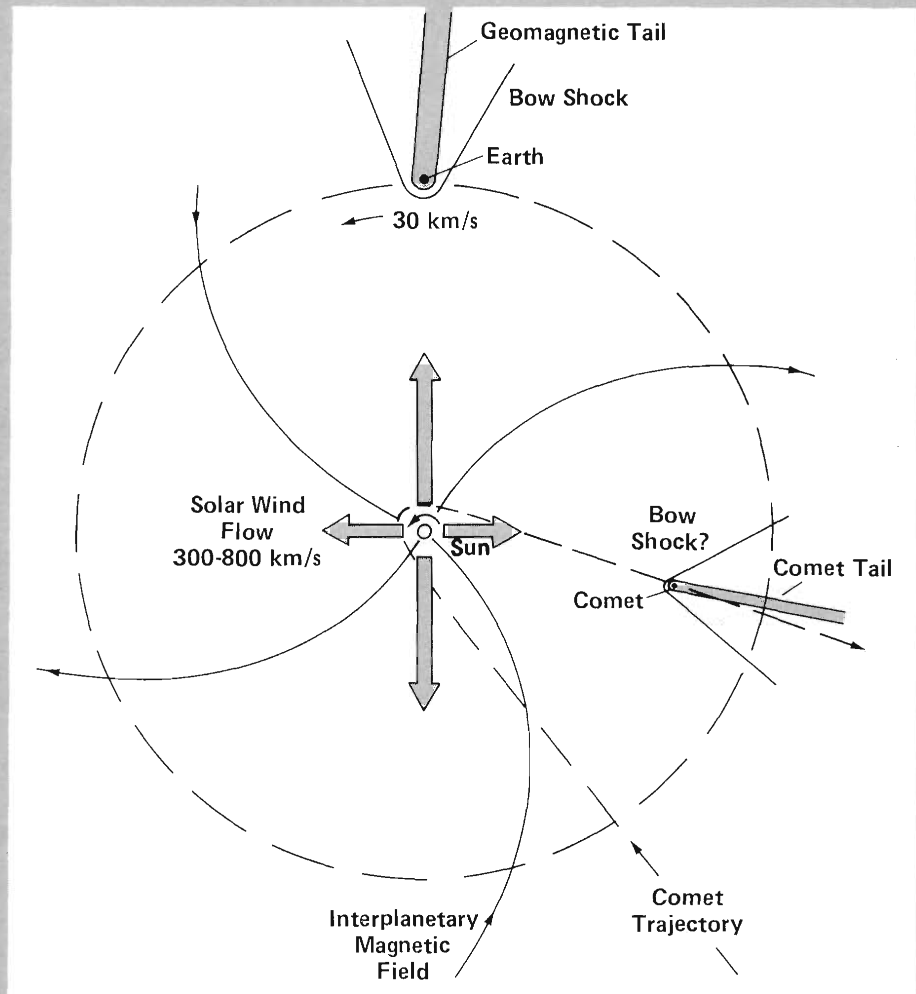
Magnetic Reconnection. The two highlighted regions in part (c) of Fig. 2 are of great interest because they both have fields of opposite polarity that are pushed together. On the dayside (left) the interplanetary and the terrestrial field lines are oppositely directed; on the nightside (right) it is the terrestrial field lines above and below the

The Solar Wind

The sun's outer atmosphere, the corona, has been familiar to man for many centuries as the faint silvery glow surrounding the black disk of the moon during a total eclipse of the sun. Only in recent years, however, has man learned that the corona pervades the entire solar system as a wind of ionized gas, gusting outward from the sun at speeds that vary from 300 to 800 kilometers per second (see figure). This wind is a consequence of the million-degree temperature and the high thermal conductivity of the ionized coronal gas. These conditions produce such a high and extensive thermal pressure that even the enormous gravity of the sun is insufficient to contain the corona as a static, bound atmosphere.

As the solar wind rushes outward, it carries frozen within it a remnant of the sun's magnetic field. If the sun did not rotate, the resulting interplanetary magnetic field would be nearly radial. Solar rotation (once about every twenty-seven days as viewed from the earth) forces the interplanetary field into an Archimedean spiral (when viewed from above as shown in the figure). The polarity of the field, that is, whether it is directed away from or toward the sun, depends on the polarity of the field at the sun where the flow of plasma originates. Because the solar magnetic dipole is generally inclined significantly to the solar equator, the polarity of the solar wind field at earth tends to reverse sign two or more times per solar rotation.

The major constituents of the solar wind plasma are protons and electrons. Typical solar wind densities measured at the earth are about 10 particles per cubic centimeter, whereas typical field strengths are 5×10^{-5} gauss. By way of comparison, the particle density of the earth's atmosphere at sea level is about 3×10^{19} per cubic centimeter, and the earth's magnetic field strength at the poles is 0.6 gauss. Despite the dilute nature of the solar wind plasma and the weakness of the interplanetary field, the flow of the solar wind determines the overall shape of the earth's magnetosphere. Further, as observations from satellites have shown, the orientation of the interplanetary magnetic field controls the transfer of mass and energy from the solar wind to the magnetosphere, and this transfer is the cause of auroral disturbances and geomagnetic activity. ■



The solar system is filled with a supersonic solar wind blowing nearly radially outward from the sun. Embedded in the flow is a remnant of the solar magnetic field, which, however, is not radial but is bent into an Archimedean spiral by the rotation of the sun. The flow of the solar wind past the earth produces a stretching of the earth's magnetic field into a long, tail-like structure on the nightside and causes a detached bow shock to form on the dayside. This geomagnetic tail is similar in some respects to the ionic tail of a comet, which results from the interaction of the solar wind with gases emitted from the head of the comet.

midplane that are oppositely directed. Plasma regions containing such opposed fields become separated by thin layers of intense electric currents called current sheets. Slow quasi-steady flow of plasma toward the current sheet occurs on both sides (large arrows that point at each other in part (c) of Fig. 2 and in Fig. 3). Such flow carries with it opposing magnetic fields.

To balance the inward flow of plasma there must also be an outward flow. Ejection of energized plasma takes place within the current sheet in two narrow wedge-shaped jets (arrows that point away from each other in part (c) of Fig. 2 and in Fig. 3). The vertices of these jets are located at the x-shaped magnetic neutral point. In part (c) of Fig. 2 we see that in the open model of the magnetosphere, outward jetting of plasma at the front of the magnetosphere can link up with inward flow of plasma at the rear. Some terrestrial field lines are broken at the front of the magnetosphere, become connected to interplanetary field lines, are dragged into the tail by the flow of the solar wind, and finally reconnect far down the tail at a distant neutral line. Thus, there is a net flow of plasma and field over both poles of the earth. Reconnection at the rear of the magnetosphere either directs plasma back toward the earth or out into space.

If we watch the fate of the magnetic field lines being carried in by the flow of plasma in Fig. 3, we can understand the origin of the term magnetic reconnection. Field lines such as C_1-C_2 and D_1-D_2 are carried toward the current sheet. When they touch at the magnetic neutral point (the primed lines), they are severed and reconnected. Subsequently, they leave the system in the exit jets as field lines C_1''-D_1'' and C_2''-D_2''.

The plasma motion leads to an electric field that is perpendicular to the plane of the magnetic field lines shown in Fig. 3. The finite resistivity of the plasma near the neutral point prevents the short circuiting of this field.

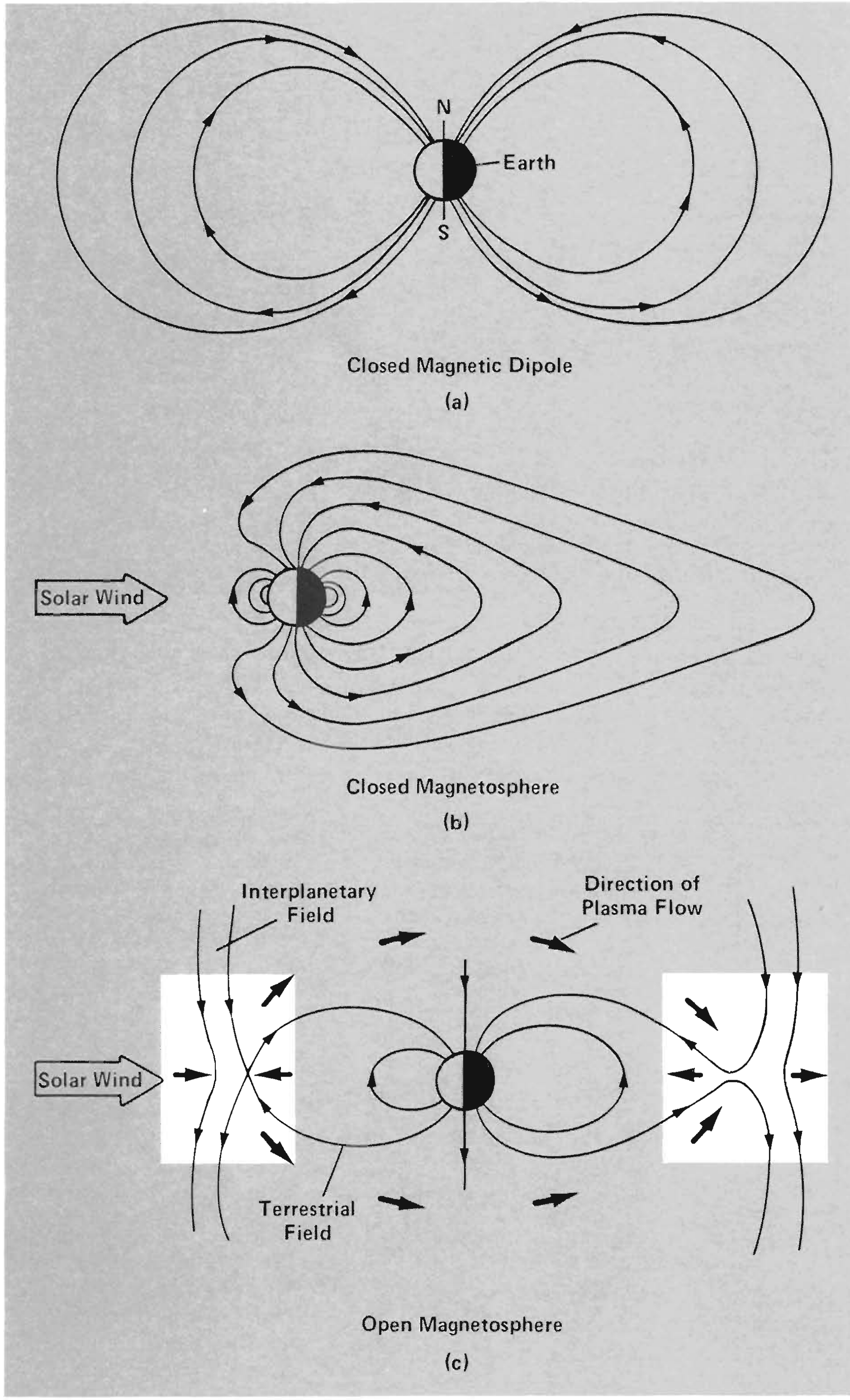


Fig. 2. Three stages in the evolution of our knowledge of the earth's magnetosphere: (a) The original model was a simple, closed magnetic dipole extending into a vacuum. (b) A model in which the solar wind flowed past the earth without the interconnection of interplanetary and terrestrial field lines depicted a distorted magnetosphere, but one that was still closed; that is, all field lines started and ended on the earth, and the solar wind merely flowed around the magnetosphere. (c) An open magnetosphere resulted when the interconnection of the terrestrial field with a weak interplanetary field carried by the solar wind was added to the model. The interconnection of the two sets of field lines allows plasma and energy from the solar wind to move into and through the magnetosphere. (For simplicity the tilt of the earth's axis is not shown.)

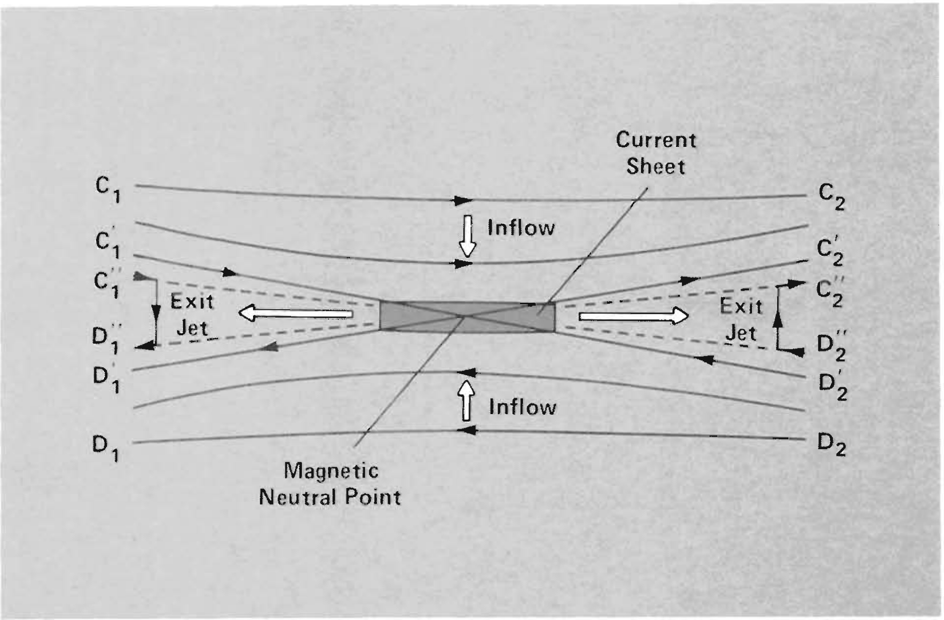


Fig. 3. Magnetic reconnection induced by flow of plasma (arrows). Carrying magnetic field lines of opposite orientation (unprimed) with it, plasma moves slowly from top and bottom toward the neutral point. Here the field lines (primed) break and reconnect. Tension within the reconnected field lines produces fast plasma jets moving to the left and right that carry the reconnected field lines (double primed) with them. (Adapted from a figure on page 75 in Space Science Physics: The Study of Solar-System Plasmas *(National Academy of Sciences, Washington, D.C., 1978).)*

An Incomplete Model of the Magnetosphere. Plasma and magnetic field measurements made by many satellites since 1961 suggest that the magnetosphere is indeed open, substantially in agreement with Dungey's model. However, these data were not decisive about many important details. For example, does solar wind plasma enter only at the front of the magnetosphere or all along its surface? How far down the tail is the distant neutral point? Are auroral disturbances relatively localized events or is the whole magnetosphere affected? What happens far from earth during an energy release? We will first describe what are believed to be the main features of the magnetosphere as pieced together from near-earth observations. Later in the article we will show how measurements far down the tail of our magnetosphere have begun answering these questions by filling out the model with many exciting details.

The very long magnetic tail of the magnetosphere extends more than six million kilometers downstream in the solar wind (Fig. 4), much like the tail of a comet. The transverse cross section of this magnetotail is roughly circular and 250,000 to 350,000 kilometers in diameter. Extending entirely across the magnetic midplane of the tail is a flat *plasma sheet* (seen edge on in Fig. 4). Field lines within the plasma sheet are rooted in opposite hemispheres at the earth; that is, plasma sheet field lines are closed. This region, the plasma sheet, is where much of the solar wind plasma accumulates. The energy extracted from the solar wind and stored in the first 600,000 kilometers of the tail is equivalent to a 1- to 10-megaton bomb, and the mass is several tons.

The electric field generated by the flow of the solar wind past the magnetosphere causes an electric current to move crosswise through the plasma sheet (flowing out and normal to the plane of Fig. 4). The current closes around the circumference of the tail, forming two gigantic solenoids that create the regions of strong magnetic field above and below the plasma sheet. These regions, called *tail lobes*, also contain plasma but generally of a much lower density than that of the plasma sheet. Lobe field lines are open, one end being tied to the earth and the other to the solar wind. This fact allows solar wind plasma to enter the magnetic tail.

Thus, inward flow of plasma appears to be a result of magnetic reconnection. At the front of the magnetosphere, reconnection occurs continuously but at a variable rate. This rate depends on the solar wind flow speed and the degree to which the two fields are antiparallel. The latter condition, in turn, depends on the momentary orientation of the solar wind field, which varies through a full range of orientations (over a period of months the north-south component averages to zero). A southward orientation with interplanetary field lines opposing the direction of the dayside terrestrial field lines (the configuration shown in part (c) of Fig. 2) is thought to result in the greatest rate of plasma accumulation within the magnetosphere.

The surface of the magnetosphere is called the *magnetopause*. Close to but just beneath the magnetopause on lobe field lines is another region of higher plasma density called the *plasma mantle*. This is solar wind plasma that has entered the magnetosphere and is flowing tailward. The cross-tail electric field causes this moving plasma to convect gradually toward the plasma sheet (along the dashed lines in Fig. 4).

At the far right of Fig. 4 is a distant magnetic neutral point where the plasma sheet of closed field lines ends. If reconnection occurs at the front of the magnetosphere, it must also occur here in the tail so that neither the earth nor the solar wind experiences a net gain of magnetic flux. Thus, in the open model we also find a continuous process of reconnection at this distant neutral point. New, closed plasma sheet field lines are being constantly formed here from open lobe field lines, and plasma is ejected into the plasma sheet. In the opposite direction the reconnection process generates unconnected interplanetary field lines that become part of the solar wind as it flows away from the earth.

The picture of the magnetosphere depicted so far uses magnetic reconnection to move solar wind plasma into the magnetosphere where it then drifts tailward toward the plasma sheet. But this process results in a net gain of mass and energy in the magnetosphere that must eventually be released. Will the release be a continuous, steady-state bleeding of plasma from the magnetosphere, or will it be a sporadic release?

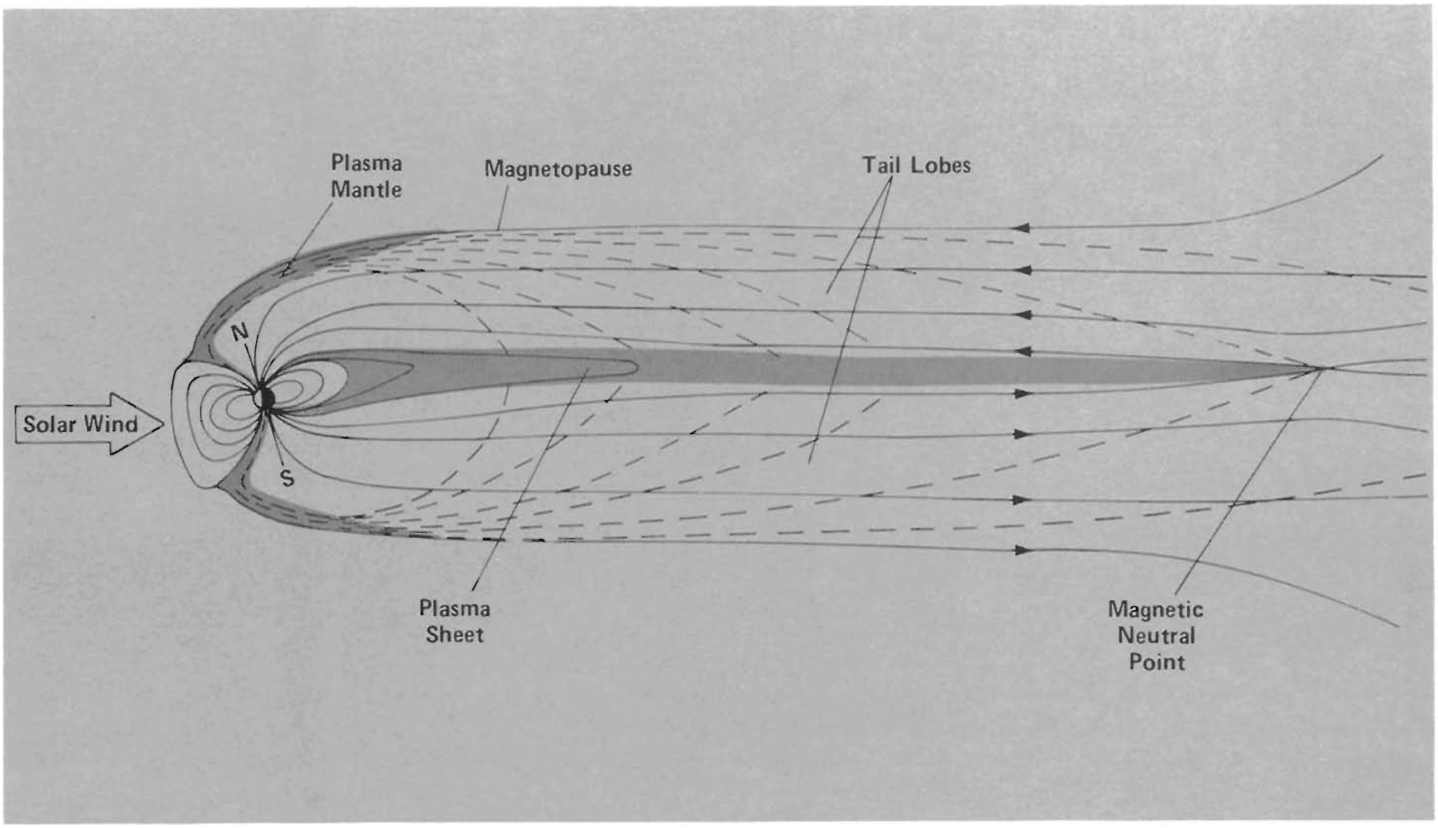


Fig. 4. A modern view of the magnetosphere. Solar wind plasma crosses the magnetopause (the surface of the magnetosphere) on reconnected field lines on the dayside to form the plasma mantle and drifts along the dashed lines through the plasma-poor tail lobes to populate the flattened, horizontal plasma sheet.

Energy Release. Evidence of one kind of sporadic energy release can be found by observing comets. A comet's tail is the result of the comet's interaction with the solar wind, probably similar in some respects to the interaction of the earth's magnetosphere with the solar wind. In fact, the existence of the solar wind was first inferred in the 1950s from an analysis of the ionic tails of comets. (In a sense, studies of the geomagnetic tail and comet tails are complementary. We can directly observe a variety of physical processes in the geomagnetic tail *in situ* with satellites but must infer its overall structure from single point measurements. On the other hand, we can directly observe optically the overall structure of a comet tail but as yet have not had an opportunity to make *in situ* measurements. In September 1985, ISEE-3, renamed ICE, will pass through the tail of a comet to provide the first such direct measurements.)

Dramatic evidence for the manner in which comets can lose energy accumulated from the solar wind is provided by Fig. 5, which shows the tail of a comet breaking off, severed, it is thought, by magnetic reconnection. Here the reconnection occurs sporadically. Does a similar loss mechanism take place in the earth's magnetosphere? Does magnetic reconnection of previously closed field lines occur deep in the plasma sheet? Is such a process responsible for auroral disturbances?

The term *auroral substorm* was coined by Syun Akasofu, who found that for a ground-based observer the auroras in the local midnight sector at high geomagnetic latitudes often brighten dramatically and spread rapidly poleward over a period of from thirty minutes to an hour. The process is repeated at irregular intervals, averaging once every few hours. Figure 6, an ultraviolet image of the earth and its upper atmosphere taken from a satellite about 25,000 kilometers above the north pole, shows an example of such a disturbance on a global scale.

During the past decade instruments on satellites orbiting to distances of about 200,000 kilometers into the magnetotail have returned plasma and magnetic field data

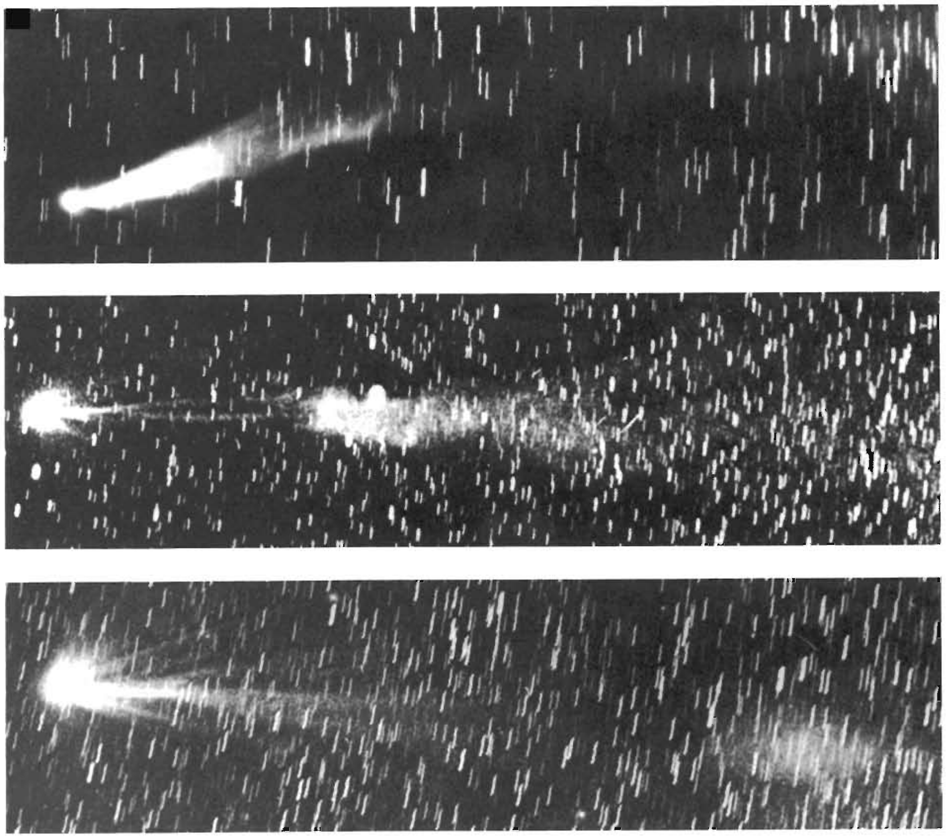
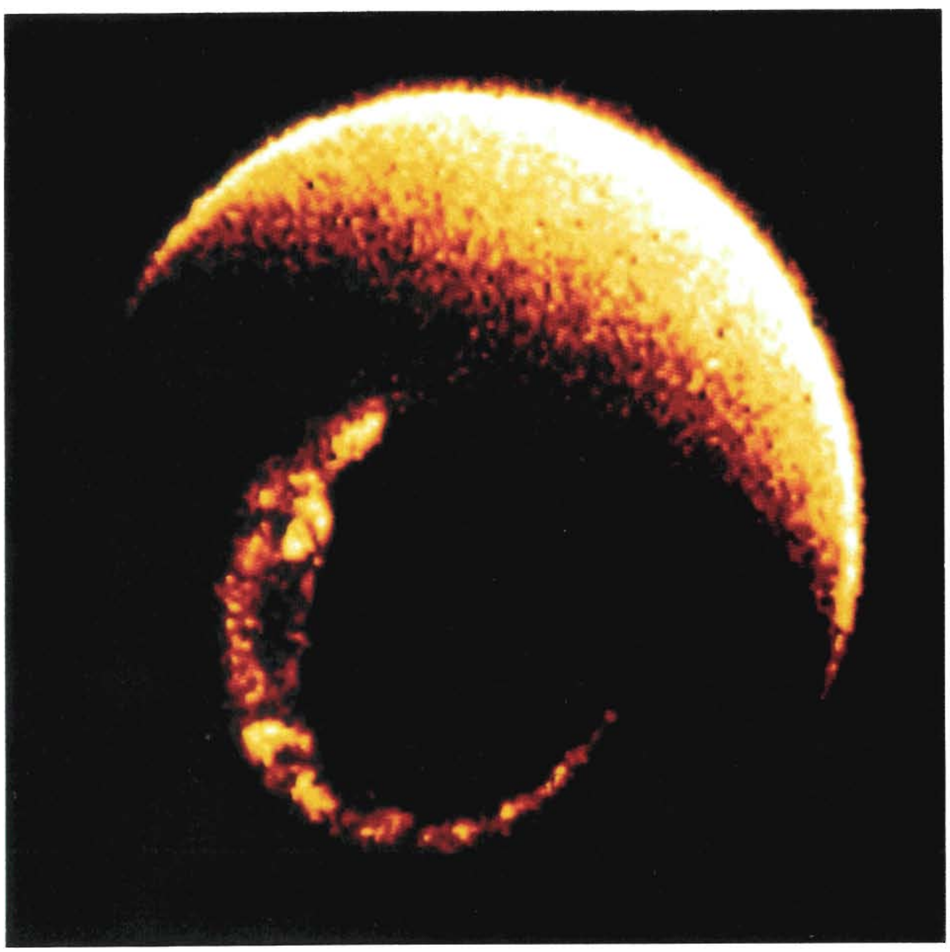


Fig. 5. Photograph (courtesy of Yerkes Observatory) of the comet Morehouse on (top to bottom) September 30, October 1, and October 2, 1908, showing the disconnection and drifting away of a portion of the comet's plasma tail.



relevant to understanding the cause of these auroral substorms. These data, recorded at positions relatively close to the earth (compared to the greater than six-million-kilometer length of the magnetotail), suggested that at about the same time an auroral substorm begins, a longitudinal sector of the tail's plasma sheet is severed by magnetic reconnection (Fig. 7). The plasma sheet field lines earthward of the reconnection point contract rapidly, driving plasma down field lines into the polar atmosphere. This process causes the auroral disturbance seen from the surface of the earth. Thus, in this model, the northern and southern lights are a by-product of the magnetosphere's sudden loss of plasma and energy.

The model also predicts that in the opposite direction a severed plasmoid of closed field lines forms and flows tailward to join the solar wind far downstream. To confirm this idea and to determine just how open the magnetosphere is—in short, to flesh out the model—we needed data from instruments deep in the tail.

Fig. 6. The aurora. This color-coded ultraviolet image of the earth was taken with University of Iowa instrumentation aboard NASA's Dynamic Explorer satellite from a point 3.27 earth radii above the north polar cap. The broad crescent is a portion of the sunlit hemisphere; the narrow crescent is an aurora on the nightside of the earth. Principal contributions are from emission lines of atomic oxygen with wavelengths of about 130.4 and 135.6 nanometers. Reconnection in the plasma sheet of the geomagnetic tail causes plasma to jet both earthward and tailward. The earthward jetting plasma causes an auroral disturbance as it impacts the earth's upper atmosphere.

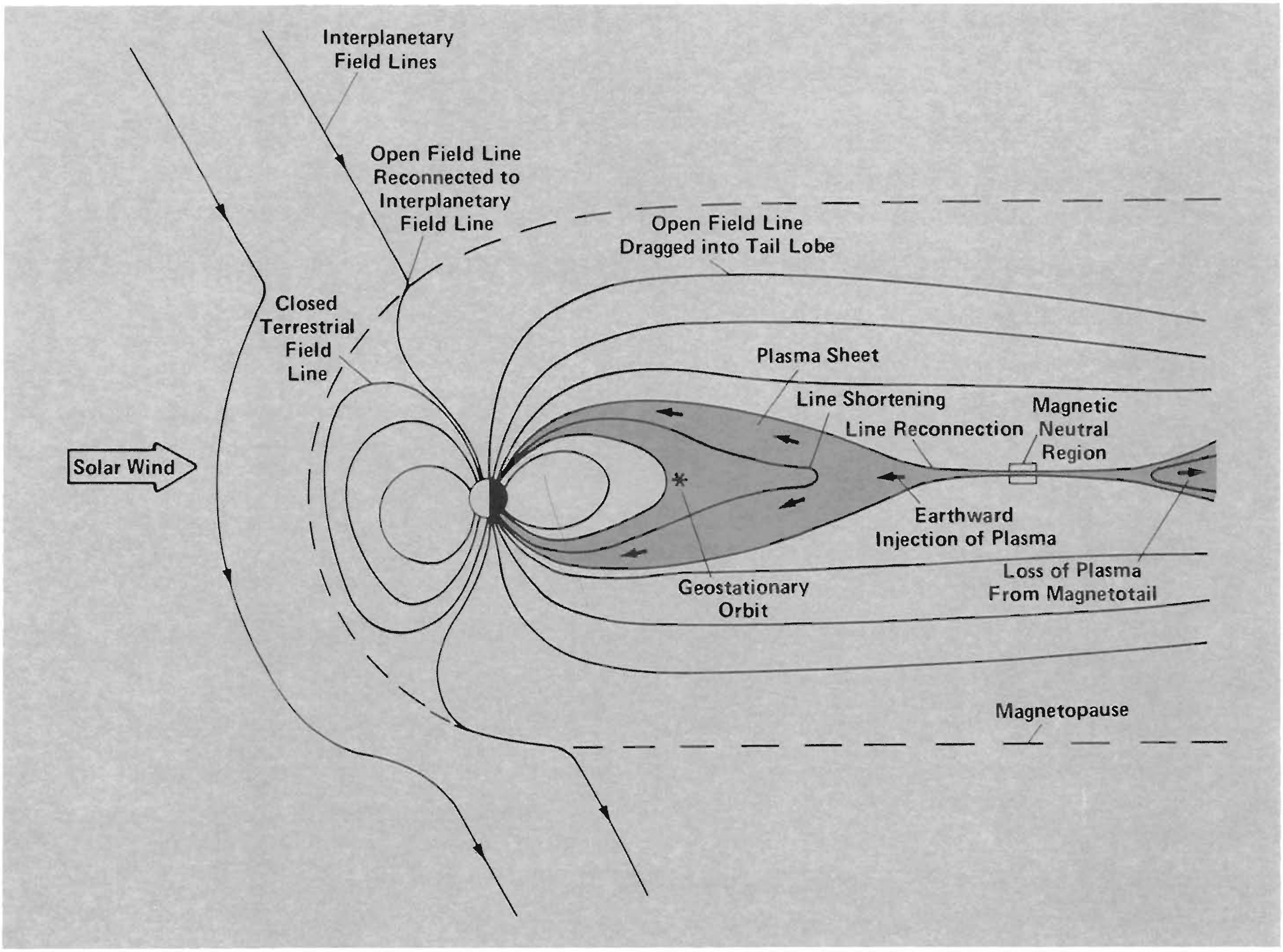


Fig. 7. Features of magnetic substorms inferred from many years of near-earth observations. Interplanetary field lines interconnect with terrestrial field lines near the front of the magnetosphere and are dragged back to form the tail lobes, carrying plasma and energy with them. Far down the tail (off the figure to the right), lobe field lines reconnect to form the plasma sheet. When the midsection of the plasma sheet becomes thin enough (due to increased magnetic pressure from the tail lobes), field lines reconnect there also, forming the magnetic neutral region shown. The release of energy that results from the reconnection heats and accelerates magnetotail plasma, driving part of the plasma back along field lines into the earth's atmosphere at the poles. Another blob of plasma is driven to the right, flowing at great velocity down the tail. Before ISEE-3's journeys into the magnetotail, we could only speculate about what occurred beyond the right-hand side of the figure.

ISEE-3's Journeys

In August 1978 NASA launched ISEE-3, the last of a triad of spacecraft designed to study the solar wind and its interaction with the earth's magnetosphere. ISEE-3 was in orbit for four years about the sunward Lagrangian point approximately 1,400,000 kilometers (220 earth radii) upstream from the earth where it provided many useful measurements of the solar wind. However, Robert Farquhar at NASA discovered unique three-body orbits involving the moon, the earth, and the spacecraft that were ideally suited for exploring the deep geomagnetic tail. So in the summer of 1982, ISEE-3 was moved out of its original orbit into a lunar-controlled orbit for deep tail exploration.

The spacecraft first crossed the tail in October 1982, still moderately close to the earth (400,000 to 600,000 kilometers). Then in December 1982 it started down the tail, reaching its apogee (or turnaround point) almost 1,500,000 kilometers into the tail on February 8, 1983 (Fig. 8). A shorter journey in April 1983 took it back down the tail to about the 500,000-kilometer point. These and several other journeys since then have allowed us to sample the tail's structure in the previously unexplored region from 400,000 to almost 1,500,000 kilometers (60 to 230 earth radii).

Plasma Data from the Distant Tail. Figure 9 is a photograph of the apparatus aboard ISEE-3 designed (under the leadership of Samuel J. Bame) to measure the velocity distributions of electrons in space. During the 3-second spin of the spacecraft, the instrument counts the number of electrons in fifteen contiguous energy bins (from below 10 to above 1 thousand electron volts (eV)) at each of 16 azimuthal angles spaced at intervals of about 23 degrees. This allows a relatively quick "snapshot" of the electron

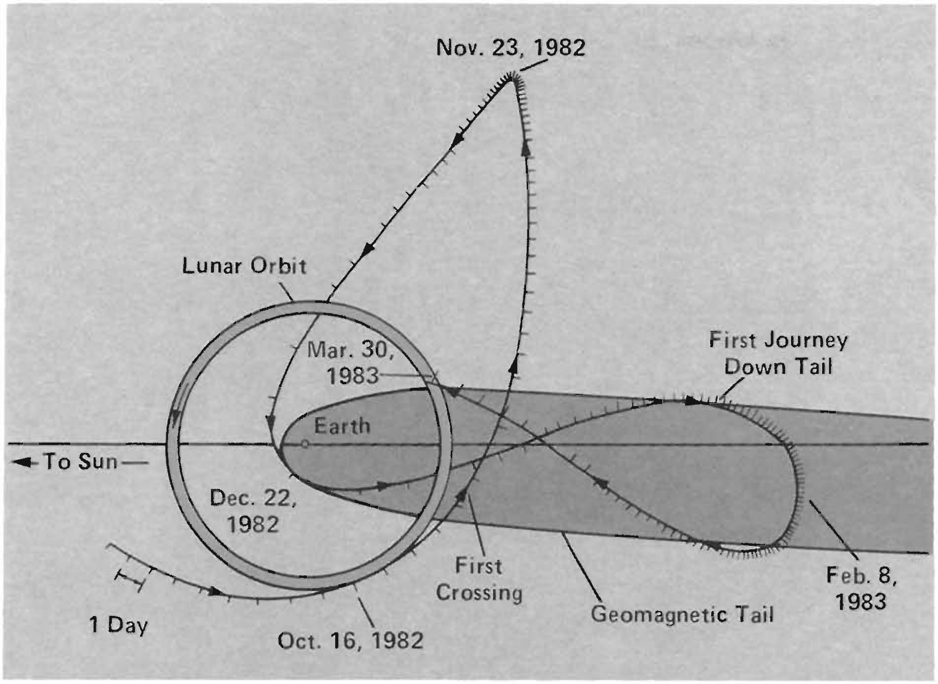


Fig. 8. The initial journeys of ISEE-3 across and down the magnetotail. The orbits, as seen from earth, are projected onto the ecliptic plane; each tic mark corresponds to one day. Lunar swing-bys were used judiciously to set up several orbits that allowed ISEE-3 to remain for long periods in the distant tail at distances up to 1,500,000 kilometers, or 230 earth radii.

Fig. 9. Equipment carried by ISEE-3 for the Los Alamos plasma experiment. The instrument has a 135-degree spherical-section electrostatic analyzer (behind the upper opening) followed by a secondary emitter system (the concave disk in the upper opening) and a 20-stage focused mesh electron multiplier (held by Samuel Bame). The number of electrons in each of fifteen contiguous energy bins from 8.5 eV to 1.14 keV is counted by varying the voltage stepwise on the analyzer plates. From these data the density, flow speed, temperature, and heat flux of plasma electrons everywhere along the spacecraft's orbit can be determined.

thermal distribution surrounding the spacecraft. The measurement is usually repeated every 84 seconds, although repetition rates of 12 seconds are within the capabilities of both the instrument and the spacecraft telemetry and are used on occasion.

Figure 10 shows how the raw data are displayed as a color spectrogram for initial analysis. The four panels give electron energy spectra averaged over the quadrants centered, respectively, on the spacecraft's noon, dusk, midnight, and dawn viewing directions. In each panel electron energy increases logarithmically from bottom to top and time progresses from left to right over a 12-hour span. The color represents a logarithmic scale of counts, ranging from dark blue for very few (10^4) to red for very many (10^4).

The spectra of Fig. 10 were taken on January 24, 1983, when the spacecraft was close to apogee. The variability evident in the figure is typical of data from deep in the tail and includes various crossings of boundaries between the major regions characterizing this part of the magnetosphere. For example, the first two major spectral changes are associated with a crossing that starts outside the tail, passes through the magnetopause into the southern tail lobe (at 1:10 universal time (UT)), and then passes from the lobe into the plasma sheet (at 4:20 UT). Some additional crossings of the magnetopause are also labeled in the figure.

The large number of such transitions between various regions of the distant tail is, at first, surprising because at apogee the spacecraft is nearly stationary relative to the sun-earth line (Fig. 8). Some of the crossings are caused by the daily wobble of the geomagnetic dipole relative to a fixed direction in space and by flapping of the tail in response to changes in the direction of the solar wind. A directional shift of 1 degree at the earth will be seen close to the spacecraft's apogee as a displacement of the entire tail by 3.5 earth radii. Solar wind shifts of this magnitude and even larger are common. However, as we will see, some of the cross-

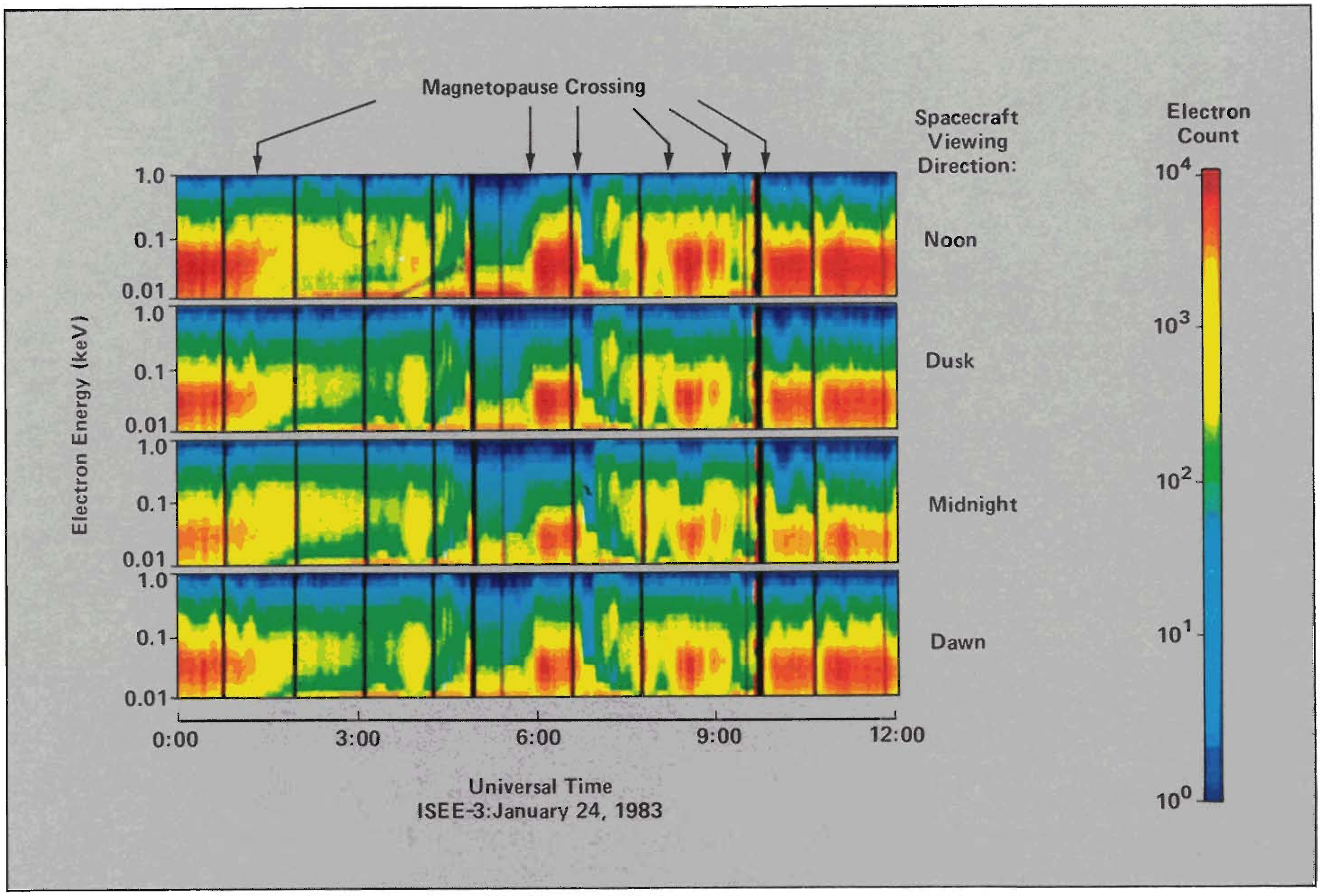


Fig. 10. Electron spectrograms from the instrument aboard ISEE-3 illustrate the variability of the plasma observed in the distant geomagnetic tail. Each panel shows an average of the energy spectra over a quadrant in the viewing direction of the spacecraft, the spectra being color coded according to the logarithm of the count rate measured by the ISEE-3 analyzer. The vertical scale for each panel is electron energy that increases logarithmically; the horizontal scale is universal time (which is based on the earth's rotation) relative to Greenwich. Color spectrograms provide an overview of the data and are used for timing different events, for selecting intervals of data for special study, and for illuminating certain aspects of the electron velocity distributions. For example, here many of the sharp changes in the spectrograms reveal frequent crossings of the magnetopause.

ings are definitely associated with auroral substorms.

Plasma Entry Into the Tail. Before we can understand the dynamic behavior of the distant magnetotail and before we can understand the effect of substorms in this region, we must try to understand the "quiet time" structure and configuration of the tail. We do this by looking at the general patterns of ISEE-3 data throughout the distant tail. Such studies help us discern where and how the energy to fuel substorms moves from the solar wind into the magnetotail. In particular, do ISEE-3 measurements confirm the idea of an open magnetosphere with plasma entry all along the magnetopause?

The data can be analyzed to yield density, temperature, and bulk flow speed of the plasma. These parameters can then be correlated with the magnitude and direction of the magnetic field as measured by the Jet Propulsion Laboratory's magnetometer aboard ISEE-3. Figure 11 shows two such sets of simultaneous plasma and magnetic field data.

In part (a) of Fig. 11 there are three crossings of the magnetopause (dashed vertical lines) characterized by dramatic changes in all three plasma parameters and marked changes in the magnetic field data. An outward crossing of the magnetopause takes the spacecraft from the plasma-poor tail lobe to an outside region with more normal flow of the solar wind. (The solar wind adjacent to and outside the magnetopause is usually called the *magnetosheath* and is separated from the undisturbed solar wind by a collisionless bow shock.) Such crossings are thus identified (as at 19:21 and 20:36 UT) by increases in the density and flow speed of the plasma as well as a drop in its temperature. The magnitude of the magnetic field drops as the spacecraft goes from the stronger terrestrial to the weaker interplanetary field lines. Further, the direction of the field rotates from the radial sun-earth direction (an azimuthal angle of about either 180 or 0 degrees, depending on whether the spacecraft is in the southern or northern lobe, and a polar angle of 0 degrees) to a direction characteristic of the interplanetary field lines (in this case, an azimuthal angle of about 100 degrees and a highly variable polar angle). Inward crossings of the magnetopause (as at 19:51 UT) have oppositely directed changes in the plasma and magnetic field data.

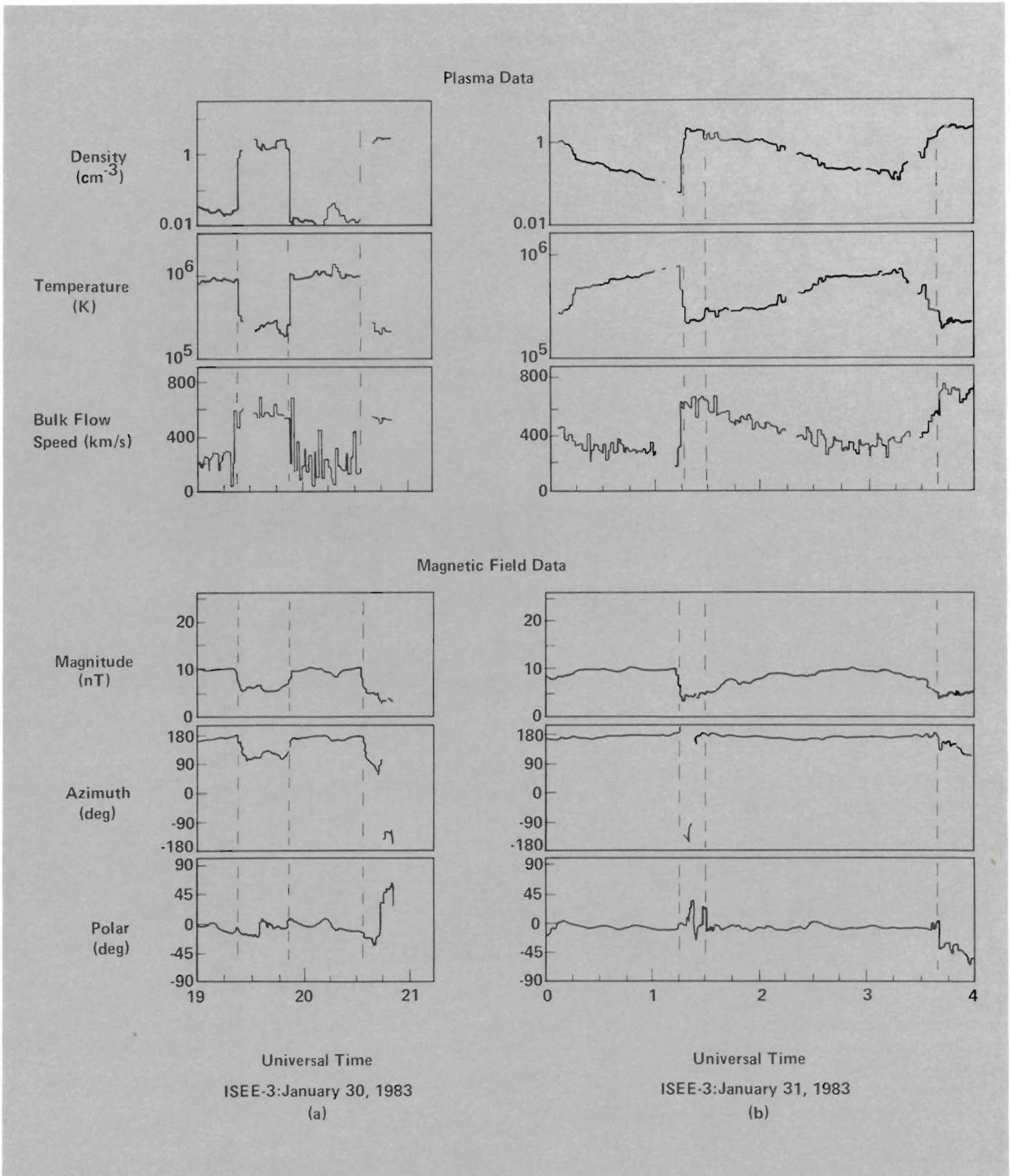


Fig. 11. Plasma and magnetic field data from ISEE-3 illustrating crossings of the magnetopause (vertical dashed lines). (a) The dramatic changes here are typical of a locally closed magnetopause as the spacecraft crosses from the plasma-depleted tail lobe to the denser, flowing, solar wind plasma outside the magnetosphere, or vice versa. (b) The more gradual changes in plasma parameters for the magnetopause crossing at 1:31 UT are typical of a locally open magnetopause in which plasma is drifting across that region of the magnetopause into the tail lobe. (The units of magnetic field magnitude are nanotesla (nT).)

The changes depicted in part (a) of Fig. 11 are what would be expected for a closed magnetosphere without local plasma entry. Frequently, however, the data are similar to those in part (b). Here the outward crossing at 1:14 UT is similar to those of part (a), but the inward crossing at 1:31 UT is quite different in that the changes in the plasma parameters are only gradual rather than dramatic. We know this is a true inward crossing because the magnetic field rotates back to the radial direction and its magnitude increases. These data show a region of plasma just inside the magnetopause that differs only slightly from the plasma outside. In fact, as the spacecraft penetrates with time deeper into the magnetosphere, the plasma gradually changes to one characteristic of the inner tail lobes, that is, with low density and flow velocity and high temperature. Thus, in this latter crossing we have detected the direct entry of solar wind plasma into the magnetosphere.

The third crossing of the magnetopause at 3:41 UT also shows a continuous change in the plasma parameters as the spacecraft crosses the surface in an outward direction, although the changes are more rapid in this case. There are two possible explanations: either the magnetotail has swept by the spacecraft more rapidly, or the region of plasma detected on the inward crossing has thinned, resulting in the more discontinuous transition from tail lobe to magnetosheath. In fact, careful examination of the crossing at 1:14 UT shows the hint of either a thin region of tailward-moving plasma or a very quick crossing of a thicker region. It is difficult to distinguish such effects using data from a single spacecraft.

We feel these data are, indeed, evidence that at times the solar wind plasma outside the magnetosphere but far down the tail flows relatively unimpeded across the magnetopause to populate the distant tail lobes. However, the plasma does not always have free entry to the lobes. Crossings that show a sudden, very large change in density (by a

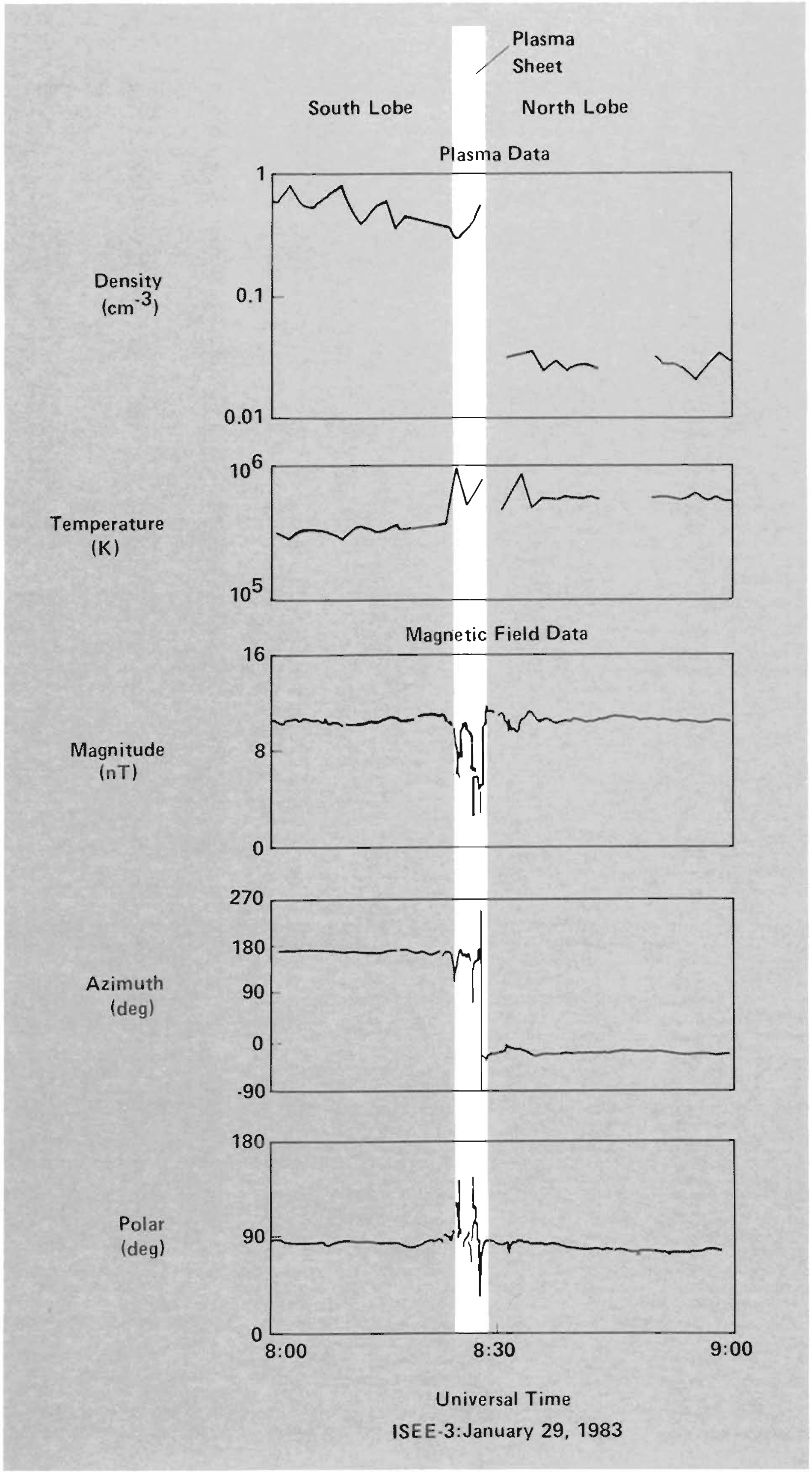


Fig. 12. These data for a traversal of the plasma sheet by ISEE-3 show a large difference between the electron densities of the north and south lobes. Such density differences are controlled by the polarity of the interplanetary magnetic field (Fig. 13) and indicate, at that point in the tail, a closed magnetopause in one lobe and an open magnetopause in the other (Fig. 14).

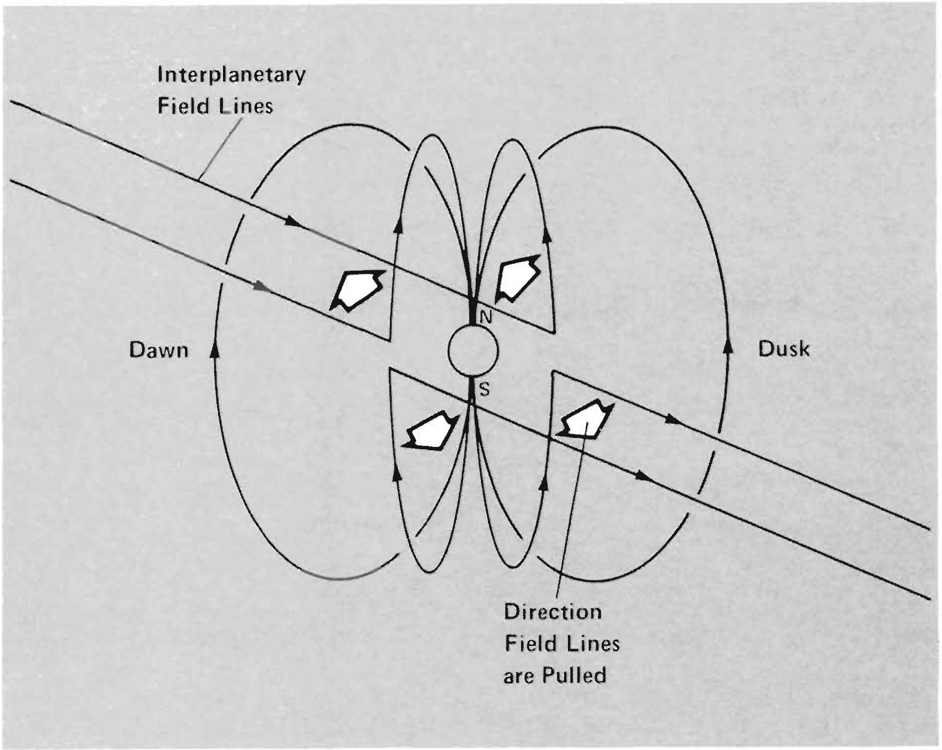


Fig. 13. Magnetic reconnection, as seen from the sun, at the dayside of the magnetosphere. The interplanetary magnetic field lines shown here have their polarity directed away from the sun, have a southward component, and are reconnecting with terrestrial field lines. The reconnected field lines are dragged back into the tail lobes in the asymmetric manner indicated by the arrows. In other words, recently reconnected field lines are pulled by field tension toward opposite sides of the magnetosphere over the northern and southern polar caps. A reversal in the polarity of the interplanetary field reverses the asymmetry of this process.

factor of 50 to 100) clearly are examples where the magnetopause is locally closed to direct penetration by solar wind plasma. We believe that plasma entry to the tail lobes is largely controlled by external factors. In fact, we have observed an asymmetry between the plasmas in the north and south tail lobes that appears to be controlled by the polarity of the interplanetary magnetic field (which changes at the earth an average of two to four times a month).

Asymmetry of Tail-Lobe Densities. Figure 12 is an example of one of the more rapid crossings of the central plasma sheet that were observed by ISEE-3. The plasma and magnetic field data extend over a 1-hour period and can be identified as a crossing from one tail lobe to the other by the reversal of magnetic field direction (the azimuthal angle switches from 180 to 0 degrees). The plasma sheet is identified by the decrease in field magnitude and the fluctuating field orientation. We can further identify this as the plasma sheet rather than plasma outside the magnetosphere by the fact that its temperature is almost ten times greater than that typical of plasma outside.

Of particular interest in Fig. 12 is the fact that the south-lobe plasma (left side) has a density comparable to that of the plasma sheet and more than ten times higher than plasma in the north lobe (right side). Comparable density differences between the two lobes have been observed whenever rapid transits of the plasma sheet occur. However, the plasma of higher density is sometimes in the south lobe, sometimes in the north lobe.

The polarity of the interplanetary magnetic field determines which lobe contains the densest plasma. We learned this from the fact that, for all rapid plasma sheet traversals examined to date, the south-lobe density was higher when the interplanetary field was directed toward the sun and lower when the field was directed away from the sun (all these events occurred when ISEE-3 was on the dawn side of the tail). Thus, the asymmetry in plasma density revealed by ISEE-3 depends on a property of the interplanetary field and thus appears to be linked with magnetic reconnection.

A natural explanation of the asymmetry lies in the corresponding asymmetry of the reconnection of field lines at the front, or dayside, of the magnetosphere. Figure 13 depicts this situation in a view of the earth as seen from the sun. Here the interplanetary field lines are shown with a southward component, the most efficient configuration for magnetic reconnection, and a westward component caused by the spiraling of field lines due to the sun's rotation (see the figure in "The Solar Wind"). For the case depicted, the field lines are directed away from the sun (which can be seen more easily in the side view of Fig. 7, illustrating the same configuration).

The tension resulting from this asymmetric reconnection of field lines (that is, the reconnection of field lines that are not truly antiparallel) is such as to pull the lines in the northern hemisphere toward the dawn side as they are dragged back into the magnetotail by the flow of the solar wind and, likewise, to pull the lines in the southern hemisphere toward the dusk side. When the interplanetary field is directed toward the sun, the manner in which the field lines get pulled back into the tail reverses. These effects lead to asymmetries along the boundaries of the geomagnetic tail and within the distant tail lobes. In particular, for the situation of Fig. 13 in which the field lines are dragged back along the northern dawn flank and the southern dusk flank, we expect, in these regions, to observe a magnetopause boundary of open field lines. Elsewhere we would expect a closed magnetopause boundary.

Asymmetric entry of plasma is the result of this asymmetric reconnection. Figure 14 represents a cross section of the distant tail as viewed from the earth when the interplanetary field is directed away from the sun. It shows entry of plasma through the open regions of the magnetopause and consequent drift toward the plasma sheet. Plasma entry occurs along the length of the tail wherever the reconnected field lines cross the magnetopause and is prevented elsewhere.

Although the northern dusk and southern dawn regions of the magnetopause are locally closed at this particular distance downtail, these regions must be open at some further distance if all lobe field lines are the result of magnetic reconnection at the nose of the magnetosphere. However, because the solar wind is continuously carrying toward the earth interplanetary field lines of varying orientation, the geometry of reconnection is always changing at the front of the magnetosphere. What we observe for a cross section relatively close to the earth are the open regions due to *recently* connected field lines that are being dragged down the tail. Those field lines that were reconnected at an earlier time produce an open magnetopause much further downstream.

This, then, is the general understanding that we have gained from ISEE-3 as to how solar wind plasma and energy enter the magnetotail. In particular, the ISEE-3 measurements have provided substantial evidence that the distant tail is open because of reconnection at the front of the magnetosphere. Now we can return to the question of how this added energy is dissipated during a substorm, and we can examine how the far magnetotail reacts as stored energy is explosively released close to the earth.

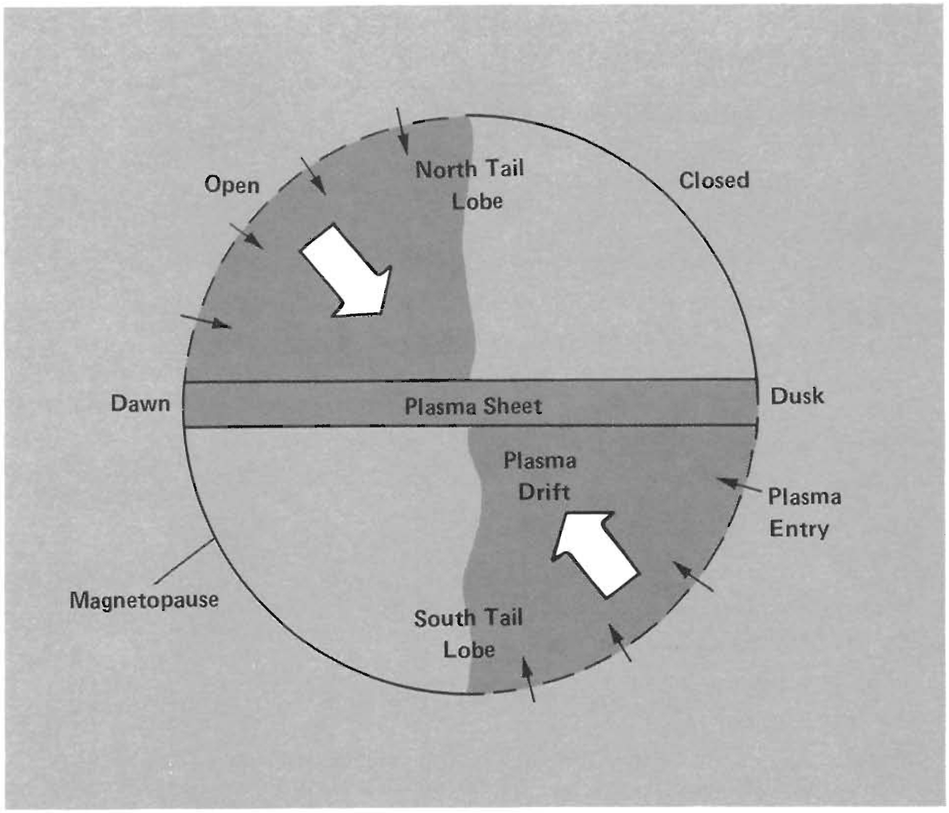


Fig. 14. Cross section of the distant geomagnetic tail as seen from the earth, illustrating the asymmetric manner in which solar wind plasma fills the tail. The open portions of the magnetopause consist of field lines that were reconnected at the front of the magnetosphere and dragged by the solar wind back into the lobes to this point in the tail. The closed portions have field lines that were reconnected earlier and whose open regions are thus much further down the tail.

Mapping a Magnetospheric Substorm

As noted earlier, substorms are believed to be initiated by magnetic reconnection within the near-earth plasma sheet. The near magnetotail can be almost continuously monitored by satellites in geosynchronous orbit 42,000 kilometers (6.6 earth radii) from the center of the earth. Scientifically, this distance is very interesting because it is at the outer edge of the Van Allen belt with its trapped radiation and is also at the inner edge of the magnetotail plasma sheet (Fig. 7). Paul Higbie and Richard Belian at Los Alamos have provided a system of energetic-particle sensors, called the Charged Particle Analyzer (CPA), for an entire series of geosynchronous satellites. The system monitors (for both the U.S. Departments of Energy and Defense) aspects of the near-earth environment, including a variety of magnetospheric conditions. In particular, some of the data can be used to map a magnetospheric substorm by comparing CPA data taken close to earth with simultaneous measurements by ISEE-3 in the deep magnetotail.

The Growth Phase. Data from geosynchronous satellites show that magnetic field and plasma parameters undergo a very regular and predictable sequence of variation in association with substorms. Whenever the flow of the solar wind and the orientation of the interplanetary field cause an enhanced coupling between the solar wind and the magnetosphere, magnetic field lines at geosynchronous orbit begin stretching, going from a dipole-like configuration to a stressed, tail-like configuration. This tail-like field is indicative of enhanced cross-tail currents and, thus, of increased storage of magnetic energy in the tail lobes and plasma sheet. (In effect, the plasma sheet is flattened by in-

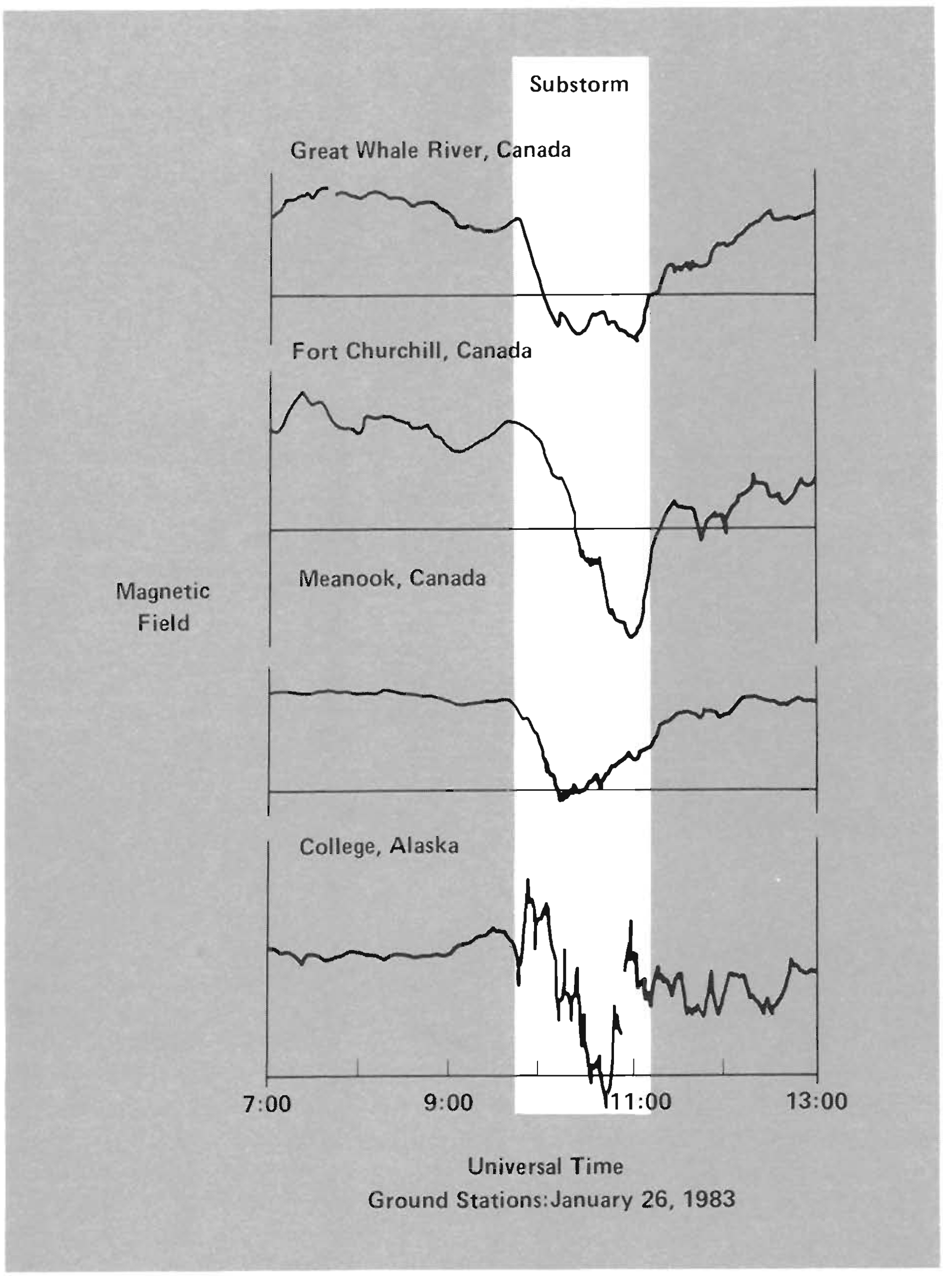


Fig. 15. Magnetic signature of an auroral substorm recorded at different ground stations on January 26, 1983. These widely separated magnetometer stations began recording, almost simultaneously at 9:45 UT, the classic "negative-bay" signature indicative of strong ionospheric current flow. The disturbance persisted for more than an hour, and the magnetic field then gradually returned to its pre-substorm values. Figures 16 and 17, respectively, show data collected during this same substorm by a satellite in geosynchronous orbit and by ISEE-3 in the distant tail.

creased magnetic pressure from the tail lobes.) This stored energy is eventually dissipated during substorms, and the interval of energy storage is known as the *growth phase* of the substorm.

Near midnight at geosynchronous orbit the growth phase has a clear signature—a change in the spatial distribution of energetic plasma particles. In particular, electrons with energies of 10 to 100 keV progressively change from being concentrated perpendicular to the local magnetic field lines to being concentrated along field lines. This change is due to plasma drift in the distorted, tail-like magnetic field and occurs about a half to two hours before the substorm release of energy.

Substorm Onset. Within a minute or so of the start of the substorm, the stressed magnetic field configuration snaps rapidly back to a more dipole-like configuration. At this time hot plasma and energetic particles are injected into the region of geosynchronous orbit. The injected plasma appears to be related directly to the rapid conversion of stored magnetic energy into plasma flow energy at reconnection sites in a limited segment of the plasma sheet. Further, the higher energy particles appear to be accelerated very impulsively, probably because of intense induced electric fields in the region where the magnetic field lines are merging.

One result of this activity in the tail is auroral activity of the kind discussed previously that produces strong magnetic perturbations near the northern and southern poles of the earth. These field disturbances are the electromagnetic signatures of strong currents flowing in the polar ionosphere, and they always accompany the visual auroral displays. Figure 15 shows the classic "negative-bay" signature of a substorm beginning more or less simultaneously at a series of magnetometer stations at about 9:50 UT on January 26, 1983. The fact that these ground sites are separated by thousands of kilometers gives us an indication that the

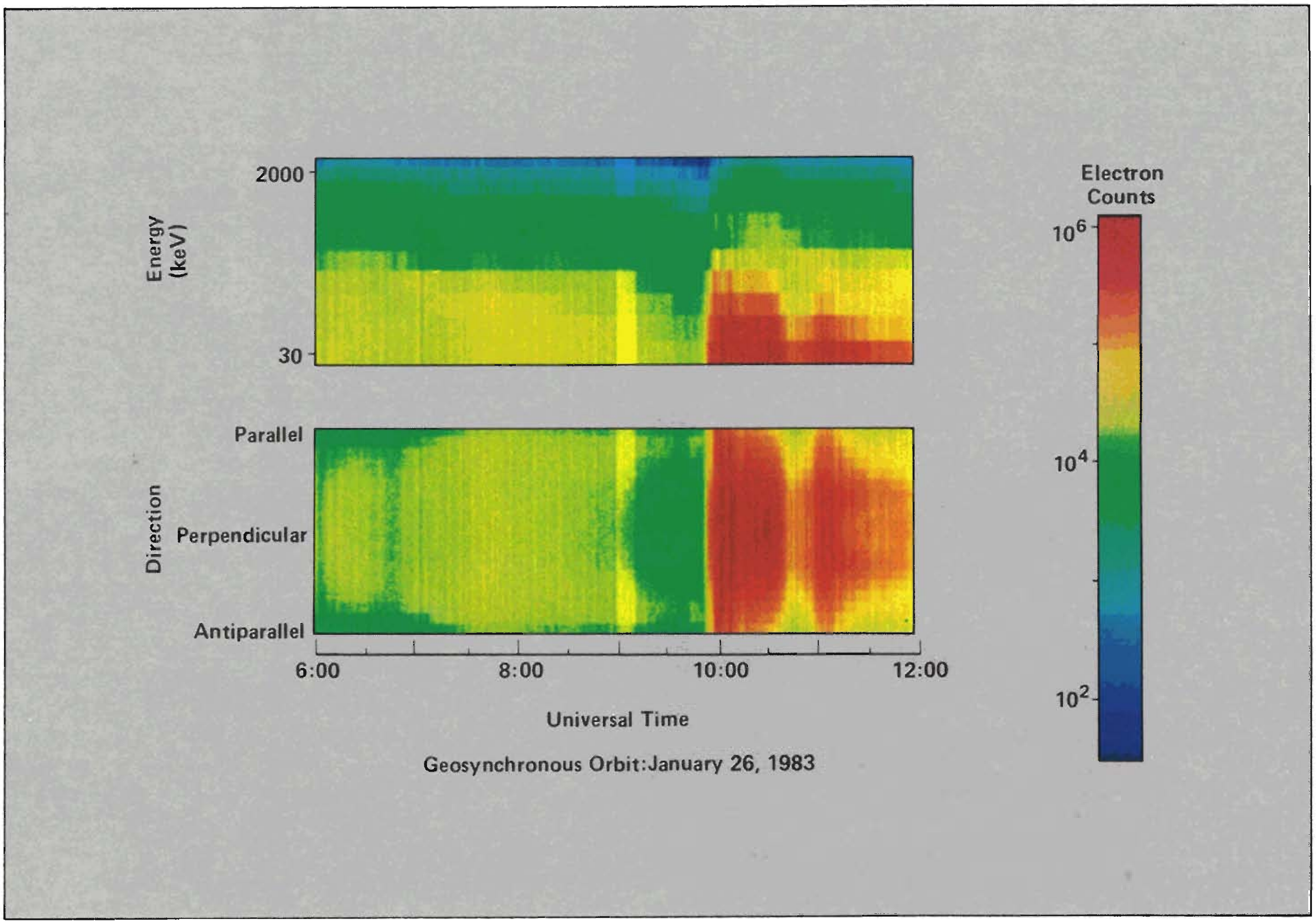


Fig. 16. Electron spectrograms, from a satellite in geosynchronous orbit, for the substorm of January 26, 1983. These panels give the energies and direction of highly energetic electrons encountered by a satellite located near the midnight meridian and 6.6 earth radii from the center of the earth. Note the distinctive "fork" in the lower spectrogram beginning around 8:55 UT. This change in electron motion from perpendicular to parallel to the field is indicative of the growth phase of the substorm. The sharp change at 9:50 UT signals the onset of the substorm itself.

substorm effects may be truly global. The magnetic records are relatively calm until substorm onset, then show large negative deflections that persist for about an hour, and finally return to pre-substorm values.

Near-Earth and Deep-Tail Comparisons. The particular substorm of Fig. 15 was observed both by satellites in geosynchronous orbit and by ISEE-3. First we will examine energy-time and angle-time spectrograms for the geosynchronous data. The two panels of Fig. 16 show the energetic (30 to 2000 keV) electron data from a satellite in geosynchronous orbit near the midnight meridian. The upper panel shows only those particles that are moving perpendicular to the local field. The lower panel gives the direction of motion of the energetic electrons, ranging from antiparallel (bottom) through perpendicular (center) to parallel (top) with respect to the local magnetic field lines.

We see from these panels that at around 8:55 UT the energetic electrons at geosynchronous orbit became largely field aligned. This is especially evident in the second panel where the number of electrons moving

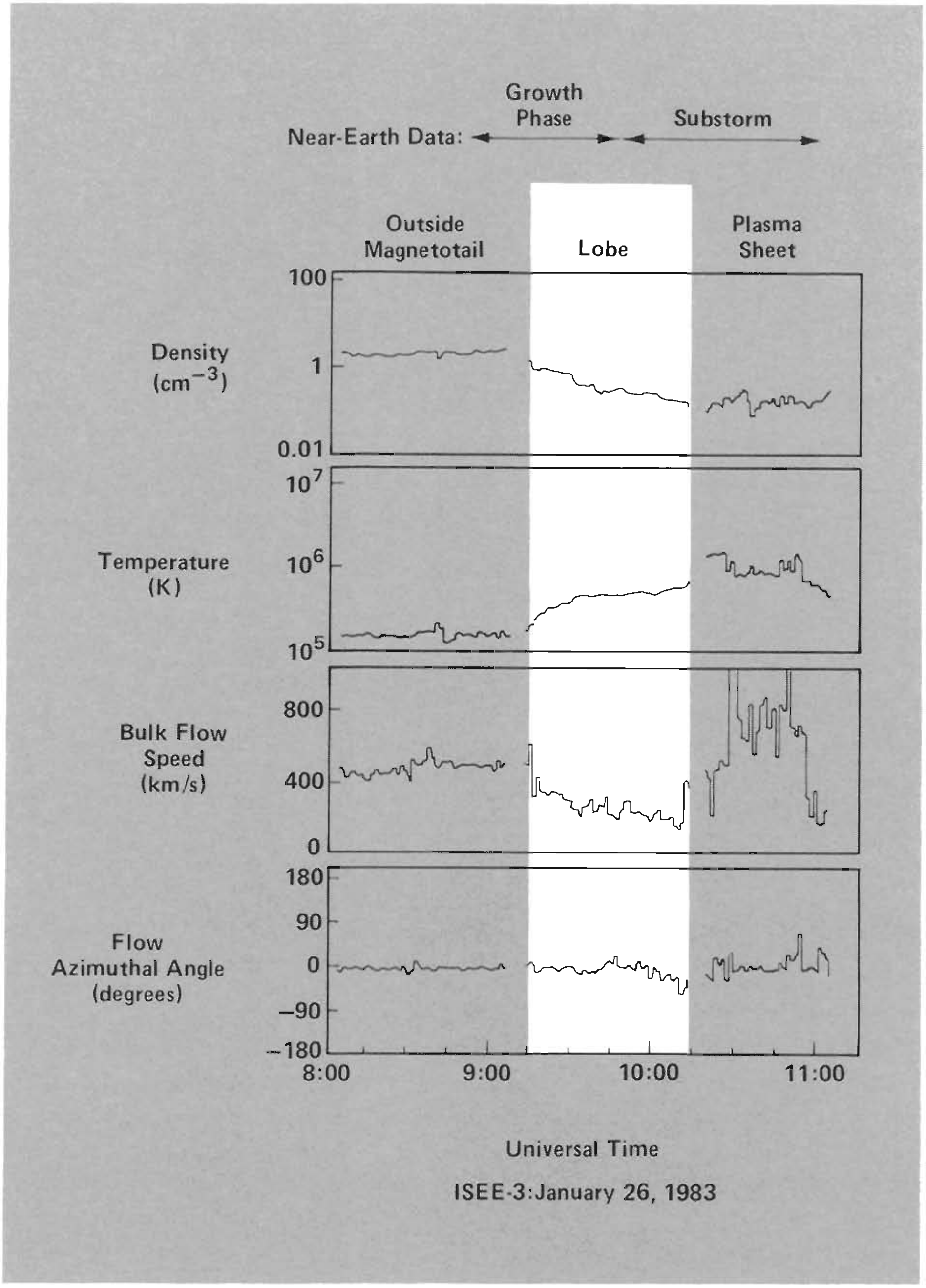


Fig. 17. ISEE-3 data from the distant tail for the substorm of January 26, 1983. Comparison of the near-earth data (arrows at top) with the ISEE-3 data shows that, at about 20 minutes after the start of the growth phase close to the earth, ISEE-3 crossed from outside the magnetotail into one of the tail lobes. Then, about 20 minutes after the onset of the substorm, the spacecraft entered the plasma sheet and encountered very high tailward plasma flows.

perpendicular to the field (center of the panel) drops, whereas the number of those moving along the field lines (top and bottom of the panel) increases. Such behavior is indicative of the growth phase of the substorm.

At 9:50 UT, when the ground stations were picking up the signature of the onset of a magnetic disturbance, the motion of the energetic electrons switches from field aligned to field perpendicular and their number increases dramatically. At this point energetic electrons are being injected from the "accelerator" in the magnetotail into the atmosphere.

What did ISEE-3 detect far down the tail at this time? Figure 17 shows that the spacecraft was outside the magnetotail (in the magnetosheath) for over an hour, but at 9:15 UT it crossed the magnetopause into the south tail lobe. This transition occurred about twenty minutes after the start of the growth phase of the substorm close to earth. At 10:11 UT, ISEE-3 encountered the plasma sheet with its high temperatures and very strong flow of plasma tailward. This latter crossing occurred about twenty minutes after the start of the actual magnetic disturbance near the earth. Thus, there is a close systematic relationship between the development of the substorm close to earth and ISEE-3 plasma observations twenty minutes later. But what changes took place in the magnetotail to cause these boundary crossings?

Growth-Phase Model. Between January 24 and February 8, 1983, we identified approximately 75 substorms at geosynchronous orbit. In about half of these cases, the geosynchronous satellites were in the proper position (near midnight) to see the growth phase. For most (about 30) of these latter, well-characterized substorms, ISEE-3 detected a corresponding transition from outside the magnetotail into one of the lobes. This pattern is so regular we feel that not only does the magnetotail expand during the growth phase but that the entire tail from geosynchronous orbit outward participates in that expansion (Fig. 18).

These results have led us to model the growth phase as follows. A southward turning of the interplanetary field lines enhances the rate at which magnetic reconnection occurs on the dayside of the magnetosphere. As these newly connected lines are dragged into the tail, magnetic and plasma energy are added at an enhanced rate, and the tail grows in diameter. During our late January and early February observations, ISEE-3 was positioned in an almost stationary fashion outside the distant magnetotail surface so that growth of the tail would eventually cause it to envelop the spacecraft. Simple calculations show that during a substorm growth phase we should expect the magnetotail to increase in diameter by several earth radii. Our observations with ISEE-3 are a strong confirmation of this heretofore speculative idea.

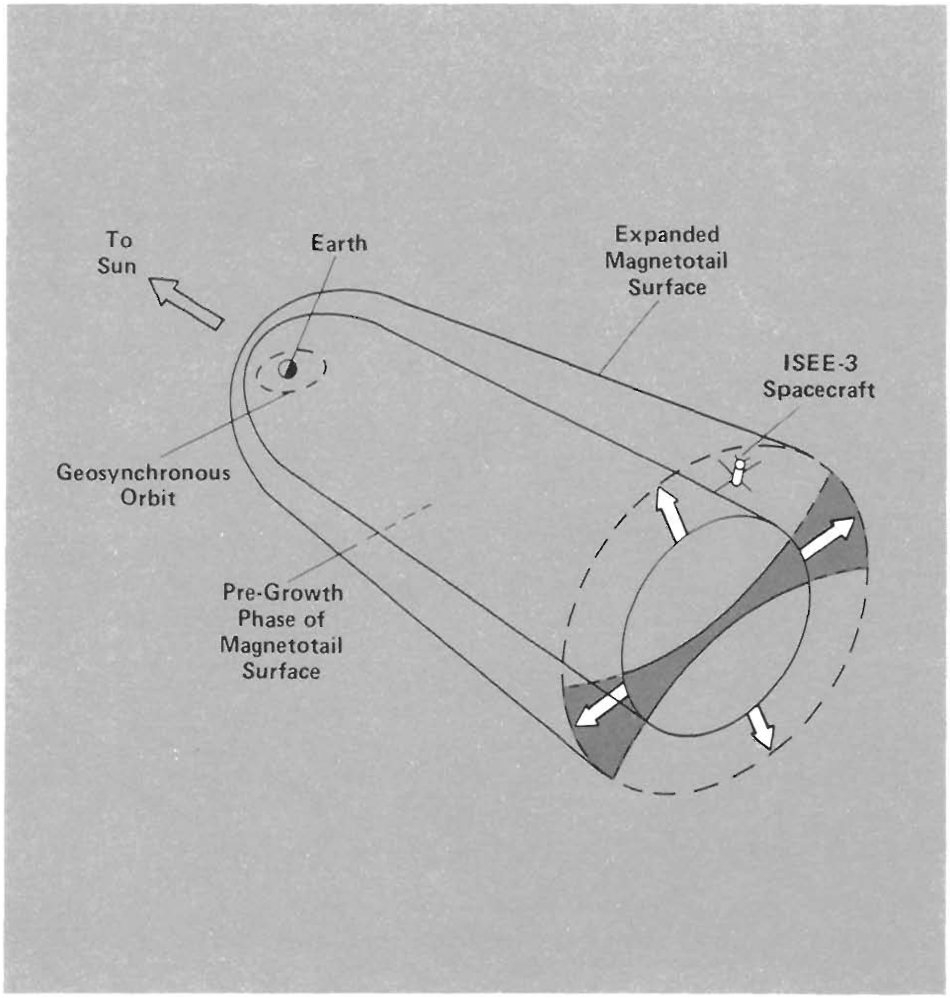


Fig. 18. Growth-phase model. Before the growth phase of a substorm begins, the magnetotail is relatively small in cross section, and, quite often, ISEE-3 resides outside. As plasma and energy are added to the tail, the cross section apparently grows until ISEE-3 is enveloped. The data show this as an inward transition from outside the magnetosphere to one of the tail lobes. Such lateral growth of the tail is seen all the way from geosynchronous orbit (6.6 earth radii) to more than 200 earth radii down the tail.

Substorm Model. The results of Fig. 17 also demonstrate another feature frequently found in our data, namely the envelopment of ISEE-3 by the plasma sheet. This occurs about twenty or thirty minutes after the start of a magnetic disturbance at the earth. The event of January 26, 1983 is typical in that the envelopment includes strong tailward jetting of plasma (between 10:20 and 11:00 UT the bulk plasma flow velocity was regularly greater than 700 kilometers per second). Further, for numerous examples the data recorded by ISEE-3 show the spacecraft passing from one lobe into the plasma sheet and then back into the same lobe again as if a bulge in the plasma sheet had passed by.

We believe such behavior is the signature of plasma release in the manner depicted in Fig. 19. The plasma sheet is severed close to earth, flows tailward as a plasmoid, and reaches ISEE-3 about twenty or thirty minutes later. The delay in arrival is consistent with the bulk flow speeds measured for the plasma of about 500 to 1000 kilometers per second. The plasmoid's dimensions increase substantially as it departs because it is moving through regions of decreasing magnetic pressure. The large size of the plasmoid results in a bulge in the magnetopause that travels along with it.

ISEE-3 provided two other pieces of information that strongly support the plasmoid model as the correct interpretation of the substorm-related encounters with the plasma sheet. First, during these encounters the magnetic field in the newly expanded plasma

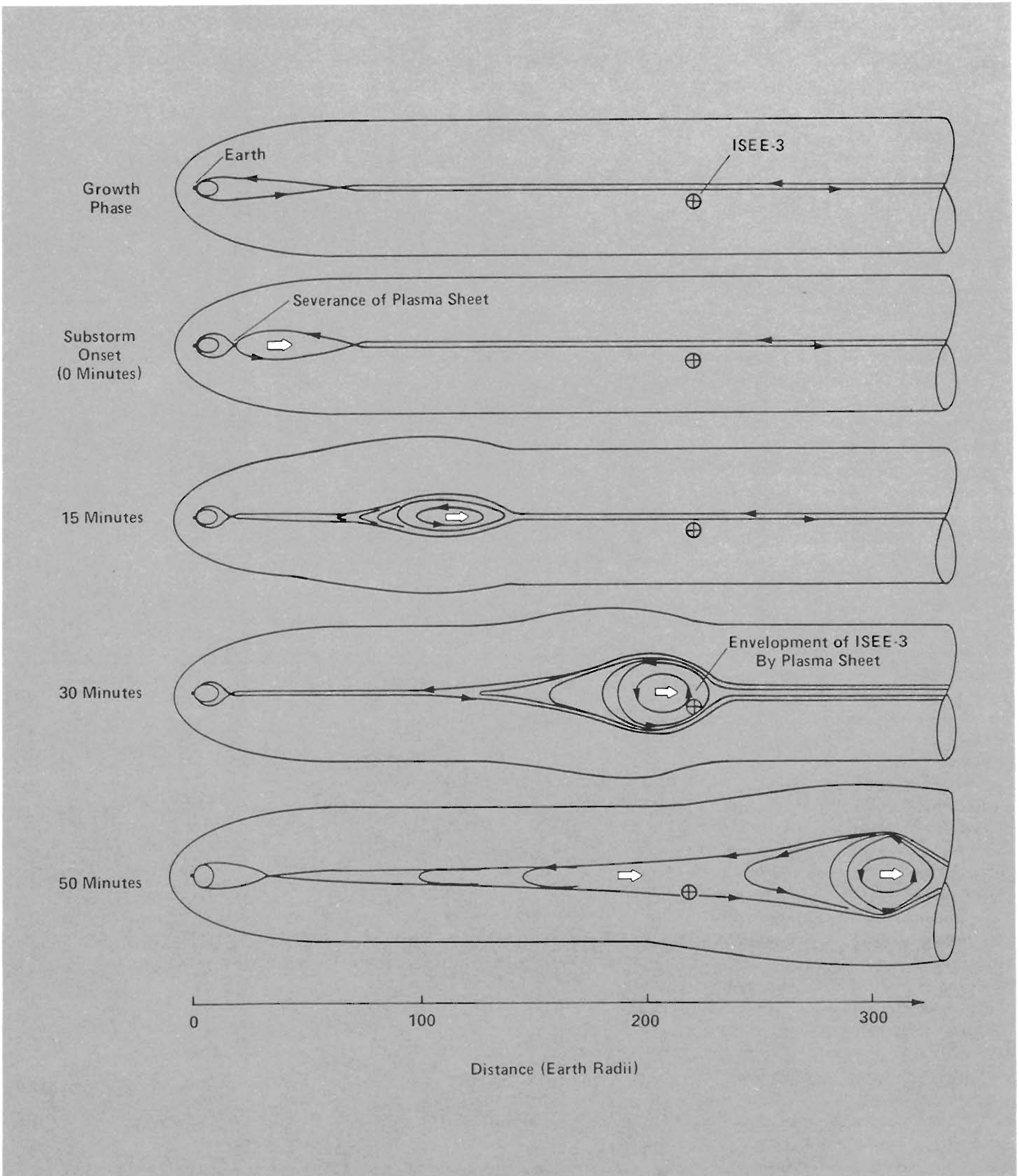


Fig. 19. Substorm model. After plasma and energy have built up in the magnetotail during the growth phase, a portion of the plasma sheet is severed by reconnection close to earth, causing substorm onset. The plasmoid thus formed hurtles down the tail, eventually enveloping ISEE-3 about thirty minutes later at its vantage point nearly 1,400,000 kilometers from earth. The envelopment is seen by instruments in the spacecraft as a transition into the plasma sheet. (White arrows indicate plasma flow; black arrows indicate the direction of the magnetic field.) This model is supported further by data such as those shown in Figs. 20 and 21.

sheet usually points steeply northward for a few minutes, then points southward (often steeply) for a longer period of time. For example, Fig. 20 shows the envelopment of ISEE-3 by the plasma sheet about 30 minutes after the onset of a substorm at the earth on January 25, 1983. This envelopment is indicated by the decreased magnetic field strength and high plasma bulk flow speed between about 5:30 and 6:10 UT, but we also see that the polar angle of the magnetic field turns steeply northward (positive 45 degrees) for several minutes followed by an even steeper and longer southward turning for most of the remaining period of the plasma sheet encounter. Examination of Fig. 19 shows that this is the expected magnetic signature of a passing plasmoid: at the front of the plasmoid the field lines point northward, just past its center they reverse and point southward, and toward the back they gradually resume a more radial direction.

Data from another particle experiment on ISEE-3 on the flow direction for energetic electrons (about 100 keV) give further support of the model. For example, in Fig. 21 we see for an event on February 16, 1983, that energetic electrons appeared at 3:05 UT, a few minutes before the plasma sheet enveloped the spacecraft, and these electrons were almost all streaming tailward. About 12 minutes later (20 minutes after the onset of a substorm near the earth) ISEE-3 encountered the plasma sheet, and, at the same time, the electron distribution became isotropic.

This change in the flow direction of energetic electrons is typical of many observed events and can be explained with the help of Fig. 19. As the plasmoid approaches ISEE-3 the spacecraft will first encounter "interplanetary field lines," that is, lobe field lines that reconnected near the earth after the plasmoid departed. These field lines pass around the outer part of the plasmoid and are contracting behind it, helping to pull it out along the tail. Electrons, accelerated at the near-earth neutral line, flow freely along these field lines and into the solar wind

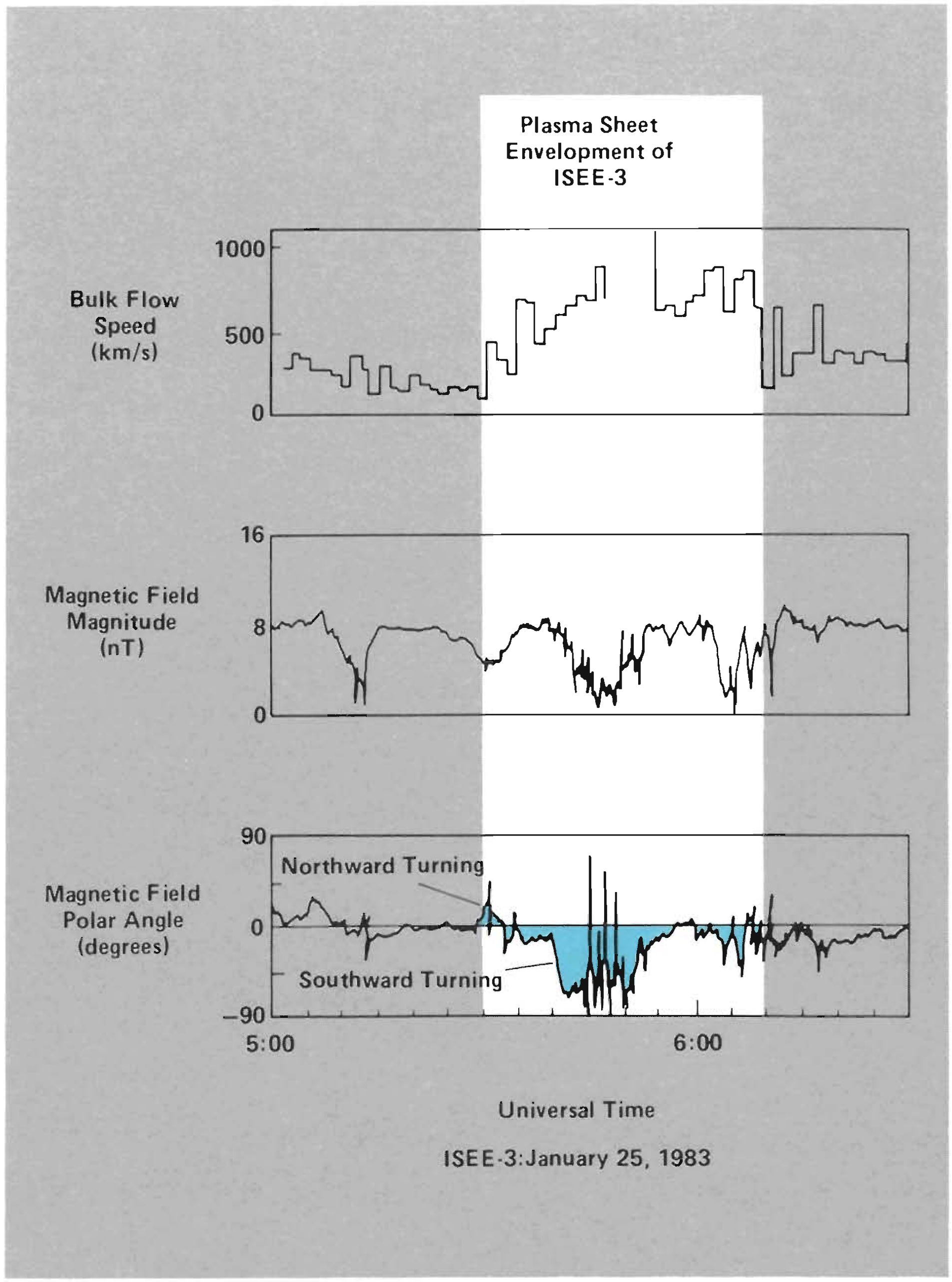


Fig. 20. Data from ISEE-3 illustrating the turning of the magnetic field for a substorm-related encounter with the plasma sheet. The plasma bulk flow speed and the magnetic field magnitude show that ISEE-3 was enveloped by the plasma sheet from 5:30 to 6:10 UT on January 25, 1983. The data for the polar angle of the field show a steep northward turning for several minutes followed by a longer southward turning as predicted in Fig. 19 for the envelopment of the spacecraft by a moving plasmoid.

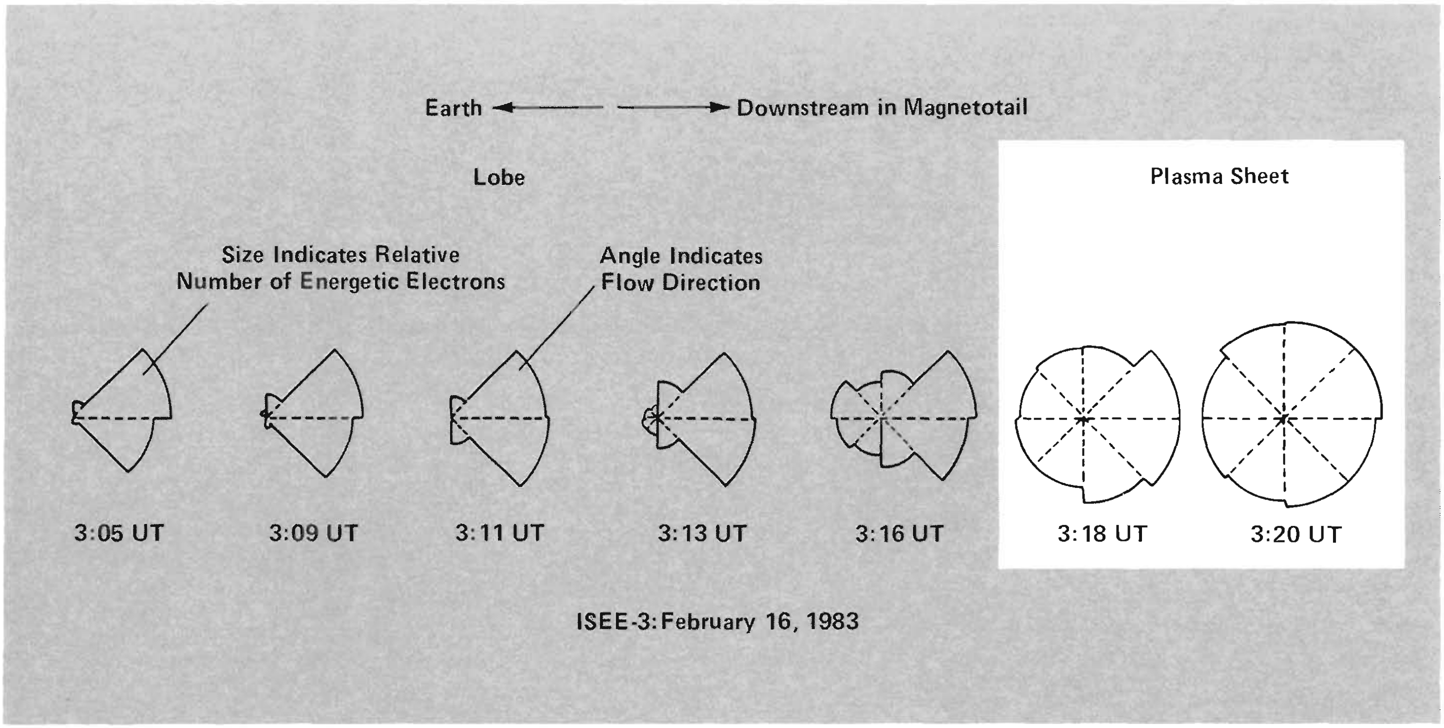


Fig. 21. Electron angular-distribution data from ISEE-3 for a substorm-related encounter with the plasma sheet. The azimuthal orientation of the motion of energetic electrons (about 100 keV) is shown. The area of each sector is proportional to the relative electron count and the orientation of that sector indicates the flow direction (for example, sectors pointing to the right are proportional to the number of electrons flowing down the tail away from the earth). For several minutes before ISEE-3 was enveloped by the plasma sheet on February 16, 1983, electron flow was strongly tailward. When the spacecraft entered the plasma sheet, motion of the electrons became isotropic. These data agree with the model that explains the plasma sheet encounter as the envelopment of ISEE-3 by a moving plasmoid. (Adapted from a paper presented by M. Scholer of the Max Planck Institut für Extraterrestrische Physik in Garching bei München.)

without reflection, thus their unidirectionality. Inside the plasmoid, ISEE-3 encounters closed magnetic loops containing electrons trapped there when the plasma sheet was severed. These electrons have isotropic motion because they have bounced back and forth within the plasmoid.

What happens when ISEE-3 is within the tail but too far above or below the plasma sheet to be enveloped by the plasmoid as it sweeps by? Because the plasmoid creates a moving bulge in the magnetotail, there should be a moderate north-then-south deflection in field direction and a temporary compression of the lobe field. Indeed, such magnetic signatures, called traveling compression regions, are seen frequently.

Thus, ISEE-3 has confirmed, in a remarkably graphic way, the ideas about substorms and their significance for magnetospheric energy gain and loss that were correctly inferred but only dimly seen in earlier near-earth observations. Continued study of the data acquired by the spacecraft during several passes through the distant tail will provide important new insights into the complicated physics of magnetic reconnection.

In the meantime, the busy spacecraft has been flung once more by the moon into an orbit that should allow it to intercept the tail of a comet (see "Comet Exploration and Beyond"). Although there are many differences between them, the geomagnetic tail and a comet's tail both result from interactions with the passing solar wind. The chance to use the same ISEE-3 instruments to measure directly the plasmas and fields within both entities should help us understand many of the similarities and differences in the mechanisms of tail formation in these remarkable solar system objects. ■

Comet Exploration and Beyond

On December 22, 1983, the ISEE-3 spacecraft whipped by the moon and headed off on a new orbit designed to intercept the Giacobini-Zinner comet in late 1985 (see figure). After having completed this maneuver by passing just seventy miles above the moon's surface, the spacecraft was renamed ICE (for International Cometary Explorer) and its mission was switched from an exploration of the earth's magnetotail to an exploration of the local environment of a comet.

Comets are formed under conditions totally different from those that lead to the planets of our solar system. They appear to consist principally of ice mixed with dust, though cold chemistry within them has also produced more complex molecules such as methane, ammonia, and cyanogen.

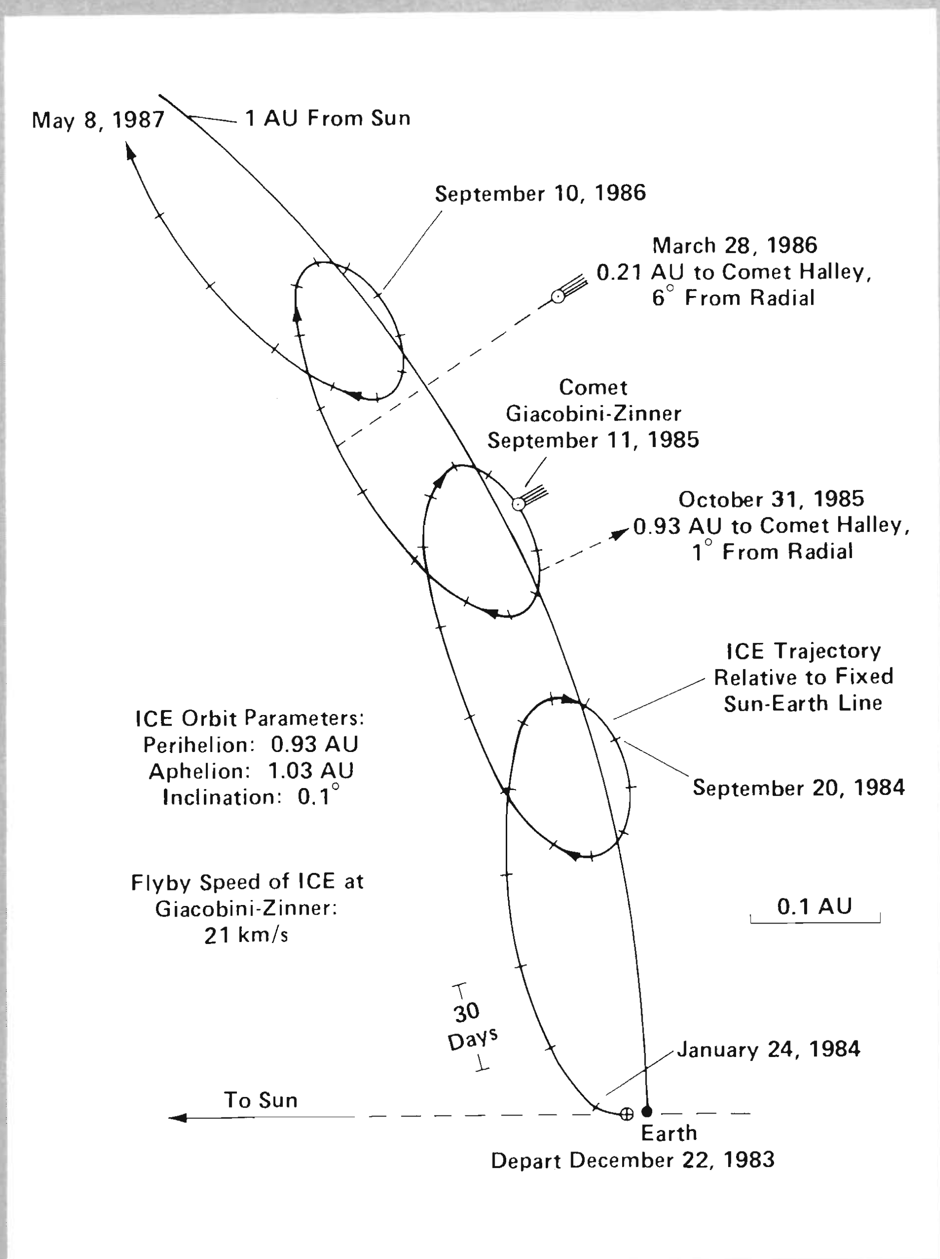
Giacobini-Zinner is an old, relatively small comet that passes by the earth approximately every thirteen years. The comet's glowing head of gases and dust blowing off its core reaches a visible size of 50,000 kilometers (intermediate in size between the earth and Jupiter) and then tapers into a tail that stretches out nearly 800,000 kilometers. Like other comets, Giacobini-Zinner is believed to have been formed in the vast reaches of space beyond Pluto and then pulled into our solar system by gravity. It has been observed since 1900, trapped in a circuit between the sun and the orbit of Jupiter.

The comet will be met by ICE within 70.8 million kilometers of earth. Because its orbit is known to vary as much as half a million kilometers from one solar passage to the next, astronomers will be working to pin down its precise orbit. At the time of intersection (September 11, 1985), the spacecraft hopefully will pass through the comet's tail within about 3000 kilometers of its head.

The likelihood of encountering dust from the comet's core in this close approach may result in damage to the spacecraft instruments. However, the danger is expected to be minimal since the earth has moved across Giacobini-Zinner's debris path several times in the past and encountered spectacular meteor showers, but with relatively short path lengths indicating small particle size for the dust.

As ICE moves through the tail, its instruments will provide the first *in situ* measurements of a comet and its environment. We hope to measure the nature and rate of ionized-gas release and learn how the comet's head and tail are affected by the solar wind. Two questions of particular interest are whether a bow shock forms upstream in the solar wind and whether the solar wind interaction accounts for the mysteriously high rate at which ionized material spews from a comet. The loss rate is known to be ten times higher than can be explained by evaporation due to solar radiation. It will also be interesting to search for magnetic reconnection in the tail.

The passage of Giacobini-Zinner in September 1985 will be followed only a few months later by Halley's comet. Observations of both comets will offer an un-



Trajectory of the ICE spacecraft after its lunar boost (bottom) on December 22, 1983. Here the line from earth to sun has been fixed in space, and we see that the relative motion of the spacecraft takes it around the sun with a radius of about one astronomical unit (AU), but in a manner that causes it to spiral ever further from earth. ICE will cross through the tail of the Giacobini-Zinner comet on September 11, 1985 and will make observations of Halley's comet from upstream positions in the solar wind around October 31, 1985 and March 28, 1986.

paralleled opportunity to compare two quite different visitors from interstellar space. Halley's comet is younger, larger, and moving faster than Giacobini-Zinner. As it passes the earth every seventy-six years, the size of its head becomes nearly as large as Jupiter, and its tail stretches 100 million kilometers. Significant differences between the two comets are expected in composition, structure, and their interaction with the solar wind.

To study Halley's comet, Russia, the European Space Agency, and Japan will launch a fleet of probes ahead of its path. Five spacecraft will make visible and ultraviolet images of the comet's sunlit core, measure the dust it throws off, and detect any bow shock the comet may create in ramming its way through the solar wind.

ICE will also make measurements near Halley's comet by moving to two positions upstream of the comet after leaving Giacobini-Zinner (see figure). The first position (October 31, 1985) will place the spacecraft 138.4 million kilometers from the comet and 76 million kilometers from the earth; the second (March 28, 1986) will place it 35 million kilometers from the comet and 96.5 million kilometers from the earth. From both upstream positions ICE will collect data on the solar wind a day or so before the plasma reaches the comet. The effect of this measured solar wind on the comet will then be observed both by the international probes close to the comet and by telescopes on the earth.

This spacecraft's exploration of the magnetotail and two comets will conclude in 1987, almost a year after its final observation of Halley's comet. Though it will continue measuring the interplanetary solar wind, the spacecraft by this time will be 121 million kilometers from earth. At this point radio signals from its low-gain antennas, originally designed for use only to a distance of 1.6 million kilometers, will be growing too weak to be detected. ICE will return to near-earth orbit in 2015 as gravity pulls it around the sun-earth system. ■

AUTHORS

John T. Gosling received his Ph.D. in physics from the University of California, Berkeley in 1965, having completed a thesis on the emission of x rays from the auroras. On a postdoctoral appointment at the Laboratory and then, starting in 1967, at the High Altitude Observatory in Boulder, Colorado, he did some of the first research on the solar wind as well as helping to photograph the solar corona outside an eclipse with an experiment flown on NASA's Skylab. He returned to Los Alamos in 1975 to work on a variety of topics in space plasma physics including solar mass ejections, the origin and evolution of disturbances in the solar wind, and magnetic reconnection at the earth's magnetopause. Recently, using data from plasma experiments flown on a variety of space probes and collaborating with numerous colleagues from around the world, he has dealt with ion dissipation and acceleration at collisionless shocks and with plasma entry into the geomagnetic tail. Gosling has just finished a term as associate editor of the space physics volume of the *Journal of Geophysical Research.*

Daniel N. Baker, head of the Space Plasma Physics Group of the Earth and Space Sciences Division, has been at Los Alamos since 1977. Prior to this he was a graduate student at the University of Iowa and a Research Fellow at the California Institute of Technology and was involved with the interpretation, analysis, and modeling of space plasma data. His research includes a variety of experimental and theoretical work on energetic plasma phenomena ranging from spacecraft instrument design to an analysis of Voyager data in the upstream region of Jupiter to the theoretical modeling of the development of magnetotail instabilities. Since coming to Los Alamos, Baker has devoted much of his effort to understanding substorms in the magnetosphere, and he has shown how these disturbances contribute to anomalies in the operation of near-earth spacecraft. At present he is interested in using modern computer techniques to enhance the acquisition, dissemination, and display of spacecraft data. He is now serving on several NASA advisory committees as well as the National Academy of Sciences Space Science Board Committee on Solar and Space Physics.

Edward W. Hones, Jr., received his Ph.D. in physics from Duke University in 1952. After working seven years in nuclear reactor physics with E. I. DuPont de Nemours and Co., he became interested in space research, which he pursued at the Convair Corporation in San Diego, the Institute for Defense Analyses in Washington, D.C., the University of Iowa, and, beginning in 1965, Los Alamos. He pioneered the observation and interpretation of plasma flow in the magnetosphere and, using a long sequence of satellite observations, developed new, compelling evidence in 1976 and 1977 that substorms involved magnetic reconnection and the formation of plasmoids. Thus he was particularly happy when he found, in the ISEE-3 observations described in this article, the rapidly departing plasmoids that he had predicted six years earlier.

AUTHORS

Acknowledgments

Successful participation in a space endeavor such as the ISEE-3 project almost always draws on the skills and expertise of a large number of people. Responsibility and credit for the results described in this paper are thus shared with many individuals. Within the Earth and Space Sciences Division, J. R. Asbridge, S. J. Bame, W. C. Feldman, D. J. McComas, and R. D. Zwickl all have had active roles either in the design, testing, and integration of the experiment, or in the analysis and interpretation of the data, or both, and it is a pleasure to acknowledge their participation and contributions. Magnetometer data from the Jet Propulsion Laboratory's experiment on ISEE-3 have been essential for the interpretation of our own data and have been freely shared with us by E. J. Smith and his colleagues. Finally we wish to thank the many individuals within the ISEE project office at Goddard Space Flight Center and at NASA headquarters who have made possible ISEE-3's excursions into the deep geomagnetic tail.

Further Reading

S. J. Bame, R. C. Anderson, J. R. Asbridge, D. N. Baker, W. C. Feldman, J. T. Gosling, E. W. Hones, Jr., D. J. McComas, and R. D. Zwickl. "Plasma Regimes in the Deep Geomagnetic Tail: ISEE-3." *Geophysical Research Letters* 10(1983):912-915.

J. C. Brandt and D. A. Mendis. "The Interaction of the Solar Wind with Comets." In *Solar System Plasma Physics, Vol. II*, edited by L. J. Lanzerotti, C. F. Kennel, and E. N. Parker, 255-292. Amsterdam:North-Holland Publishing Co., 1979.

S. W. H. Cowley. "Plasma Populations in a Simple Open Model Magnetosphere." *Space Science Reviews* 26(1980):217-275.

E. W. Hones, Jr. "The Magnetotail: Its Generation and Dissipation." In *Physics of Solar Planetary Environments*, D. J. Williams, editor, 558-571. Washington, D.C.:American Geophysical Union, 1976.

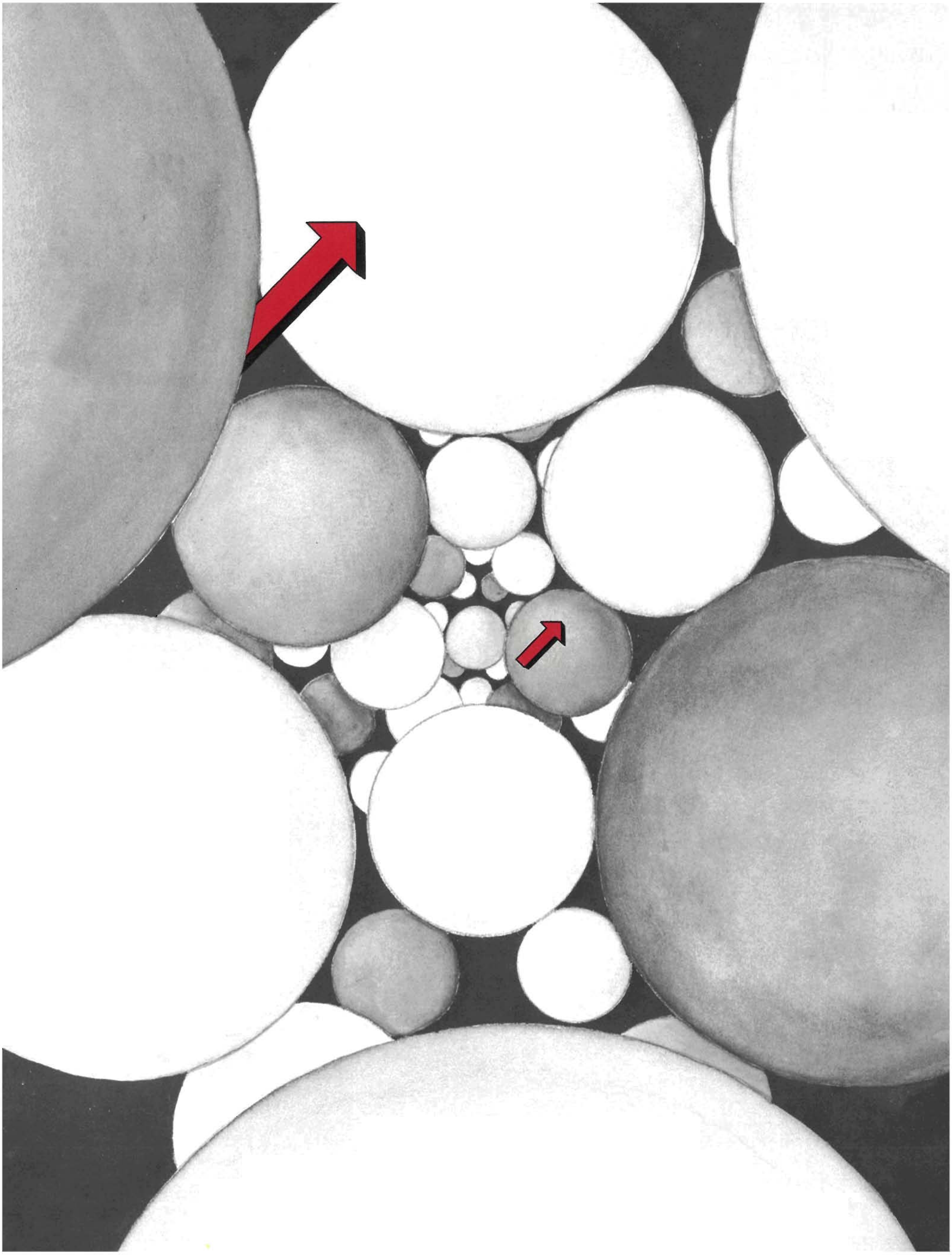
E. W. Hones, Jr. "Transient Phenomena in the Magnetotail and Their Relation to Substorms." *Space Science Reviews* 23(1979):393-410.

E. W. Hones, Jr., D. N. Baker, S. J. Bame, W. C. Feldman, J. T. Gosling, D. J. McComas, R. D. Zwickl, J. Slavin, E. J. Smith, and B. T. Tsurutani. "Structure of the Magnetotail at $220R_E$ and Its Response to Geomagnetic Activity." *Geophysical Research Letters* 11(1984):5-7.

A. J. Hundhausen. *Coronal Expansion and Solar Wind.* New York:Springer-Verlag, 1972.

B. U. Ö. Sonnerup. "Magnetic Field Reconnection." In *Solar System Plasma Physics, Vol. III*, edited by L. J. Lanzerotti, C. F. Kennel, and E. N. Parker, 47-108. Amsterdam:North-Holland Publishing Co., 1979.

V. M. Vasyliunas. "Theoretical Models of Magnetic Field Line Merging." *Reviews of Geophysics and Space Physics* 13(1975):303.



p-state superconductivity?

A routine test leads to an extraordinary discovery.

by Gregory R. Stewart, Zachary Fisk, Jeffrey O. Willis, and James L. Smith

An attractive interaction between pairs of electrons with parallel spins may be responsible for the observed but unexpected superconductivity of UPt_3. (Adapted from a drawing by author James L. Smith.)

Some experimental findings are so unexpected, so outside the limits of previous experience, that their interpretation lags well behind the facts, awaiting new insight. Our finding of September 16, 1983, was certainly unexpected, but a possible interpretation was immediate—and exciting. We were measuring the low-temperature electrical resistance of a tiny whisker of the intermetallic compound UPt_3 to see how defect-free its crystal lattice was, and, as the whisker slowly cooled, its resistance suddenly fell to zero, a clear indication of superconductivity.

What was unexpected was not the superconductivity per se but its occurrence in a material we were investigating as a likely candidate for the greatly enhanced spin fluctuations characteristic of almost ferromagnetic materials. This phenomenon reflects a tendency toward magnetism (and hence is often called near magnetism), and a large body of experimental evidence supports the view that, like ferromagnetism and superconductivity, substantial spin fluctuations and superconductivity are mutually incompatible. Not one of the thousands of known superconductors had exhibited convincing evidence of enhanced spin fluctuations, nor had any of the few known "spin fluctuators"* exhibited superconductivity.

How, then, did we interpret what we had seen? The idea immediately came to mind that perhaps UPt_3 is a "*p*-state" superconductor (see "Superconductivity and Spin Fluctuations"). This type of superfluidity, which would *not* be incompatible with spin fluctuations, had been considered twenty-odd years ago as a generalization of the BCS theory and had been observed a decade ago in liquid helium-3 at millikelvin temperatures. Many other materials had been examined as possible *p*-state superconductors because of their relatively large magnetic susceptibilities, but all had failed a crucial test involving extreme sensitivity of the superconducting transition to lattice defects. Could UPt_3 be the first?

Before recounting the tale of our work on UPt_3, we point out that it is but one of many esoteric materials we investigate not only for their inherent scientific interest but also for their possible technological value. (Spin fluctuators, for example, are related to catalysts and hydrogen-storage media.) The materials are drawn from the alloys and intermetallic compounds of the elements known as the transition elements, the lanthanides, and the actinides. All of these elements are characterized by the presence of electrons in inner *d* or *f* shells, and the variable behavior of such electrons is responsible, on an atomic level, for the diversity found in the crystalline solids containing the elements. In some cases the electrons are localized on the ions in the lattice; in others the electrons are itinerant, that is, are free to move about the lattice as do conduction electrons in metals. These extremes of behavior can result in magnetism and superconductivity, respectively. Of particular interest are those materials in which the electrons are "indecisive," easily pushed toward one or the other extreme. Among these materials had been found two of the three known spin fluctuators, not to mention the two known "heavy-fermion" superconductors (of which more later).

Why UPt_3?

Our interest in UPt_3 as a possible spin fluctuator was aroused in the fall of 1982, when J. J. M. Franse, Universiteit Amsterdam, sent us a collection of papers by his

We use the term "spin fluctuator" as shorthand for a material exhibiting enhanced spin fluctuations.

Superconductivity and Spin Fluctuations

by David Pines

Electrons in normal metals are a normal Fermi liquid in that, no matter how strong their interaction, the particles retain the essential properties of a noninteracting fermion system: the ground-state distribution of particles can be characterized by a Fermi surface in momentum space (which is spherical if one neglects anisotropy introduced by the lattice); excited states can be placed in one-to-one correspondence with those of a free-electron gas; the specific heat varies linearly with the temperature; and so on. In superconductors, on the other hand, as first shown by Bardeen, Cooper, and Schrieffer in their microscopic theory developed in 1957, both the ground state and the excited states of the system are altered in a fundamental way. A net attractive interaction between pairs of particles near the Fermi surface gives rise to an instability of the normal state, and the superconducting ground state becomes a single quantum state, the *condensate*, which is a *coherent* superposition of bound particle pairs and which can flow without resistance. To produce single-particle excitations from the superconducting ground state requires a finite amount of energy, the energy gap, so that the specific heat of a superconductor is drastically altered from that of a normal metal.

In BCS theory the net attractive interaction between conduction electrons near the Fermi surface arises from the exchange of phonons, the quanta of crystal lattice vibrations. The coherent pairs that make up the condensate are in ^0S states (that is, states with zero total spin and angular momentum), corresponding to pairs of particles with opposite spins and momenta. Other pairings, such as *p*-state pairing (in which the condensate would be a coherent superposition of pairs of particles with parallel spins) or *d*-state pairing, are in principle possible; however, both experiment and microscopic calculations to date suggest that where electron-

group on various magnetic and nearly magnetic systems. Among the papers was one by P. H. Frings and coworkers entitled "Magnetic Properties of U_xPt_y Compounds," which had been presented during the summer at a magnetism conference in Kyoto that none of us had been able to attend. In this paper were data on the specific heat and magnetic susceptibility of UPt_3 at temperatures above 1 kelvin. These data clearly hinted at enhanced spin fluctuations.

One sign of such behavior is a magnetic susceptibility whose order of magnitude lies approximately midway between that of a nonmagnetic metal ($\sim 10^{-4}$ electromagnetic unit per mole) and that of a ferromagnetic metal ($\sim 10^{-1}$ emu/mole). Frings et al. reported a susceptibility of 0.8×10^{-2} emu/mole for UPt_3, a value of the right order of magnitude and similar to those of the two other known metallic spin fluctuators $TiBe_2$ (discovered at Los Alamos) and UAl_2. (Liquid helium-3, the other spin fluctuator known at the time, is nonmetallic.) Another sign of near magnetism is an increase in the susceptibility at some high magnetic field, indicating the transition from near magnetism to magnetism. Such an increase occurs for UPt_3, according to Frings et al., between 150 and 200 kilogauss. A final, sure sign of enhanced spin fluctuations is an increase, rather than a steady decrease, in the specific heat with decreasing temperature.

phonon interactions are sufficiently strong as to bring about superconductivity, an s-state condensate will be energetically favorable.

Under some circumstances a normal Fermi liquid may become almost ferromagnetic in that particle interactions give rise to internal magnetic fields that act to enhance substantially the usual Pauli paramagnetic susceptibility. In such a system low-frequency spin fluctuation excitations are likewise greatly enhanced, and the strong coupling of particles near the Fermi surface to these spin fluctuations (sometimes called paramagnons) leads to an effective mass that is frequency (and temperature) dependent. A signature of this dependence is a term in the specific heat that varies with temperature as $T^3 \ln T$ (compared to the T^3 variation characteristic of normal Fermi liquids). Three such almost ferromagnetic metallic Fermi liquids have thus far been discovered: $TiBe_2$, UAl_2, and, most recently, UPt_3.

Liquid helium-3 is an example of a fermion system that is both nearly ferromagnetic and, at temperatures less than 2 millikelvins, *superfluid* (the analogue for neutral systems of superconductivity). Its specific heat contains a $T^3 \ln T$ term, and neutron-scattering experiments provide direct evidence for the strongly enhanced low-frequency spin fluctuation excitations responsible for that behavior. It is moreover a p-state superfluid; that is, the condensate is formed from coherent combinations of pairs of particles of parallel spins in $^3P_1$ states. This p-state superfluidity is not an accident: the short-range repulsion between helium-3 atoms is so strong that s-state pairing is strongly suppressed, and the interplay between that strong repulsion and the Pauli principle is responsible for the almost ferromagnetic behavior. Put another way, the particle correlations responsible for the enhanced spin fluctuations tend to oppose s-state superfluidity and to favor formation of a p-state condensate, in part as a result of the particle-spin fluctuation coupling.

It is natural therefore to hope that in metals exhibiting strongly enhanced spin fluctuations, one might possibly have p-state superconductivity of purely electronic origin. UPt_3 appears to be a particularly promising candidate for such an electronic analogue of liquid helium-3. Not only might it be the first metal for which electron interactions alone give rise to superconductivity, but its identification as an anisotropic superfluid could open the way to a quite new family of superconducting phenomena, in much the same way as the study of superfluid helium-3 has vastly expanded our understanding of neutral superfluid phenomena. ■

Laboratory consultant **David Pines** is Professor of Physics and Electrical Engineering and a member of the Center for Advanced Study at the University of Illinois. He has carried out pioneering studies on classical and quantum plasmas, electrons in metals, collective excitations in solids, superconductivity, superfluidity, nuclear structure, and, most recently, on superfluidity and the internal structure of neutron stars, the theory of compact x-ray sources, and elementary excitations in liquid helium-3 and helium-4. He is the author of three books and serves as editor of *Frontiers in Physics* and *Reviews of Modern Physics*. The latest of his many honors was receipt in 1983 of the newly established Friemann Prize in Condensed Matter Physics.

This increase follows from the presence of a term proportional to $-T^3 \ln T$ in the electronic specific heat. Frings et al. reported an upturn in the low-temperature specific heat of UPt_3 but gave no detailed analysis of its temperature dependence. In light of these suggestive data, we planned a more thorough investigation of UPt_3.

We were also interested in UPt_3 because of our research on the new class of materials described in the sidebar "Heavy-Fermion Superconductors." The intermetallic compound $CeAl_3$ was regarded as a likely member of this class and yet showed no superconductivity. A study of UPt_3 might help explain why, since UPt_3 and $CeAl_3$ have the same crystal structure.

The Serendipitous Experiment

Before proceeding with our plans for UPt_3, we wanted single crystals of very high quality. By June of '83 we had grown some crystals in the form of tiny whiskers (see "Single Crystals from Metal Solutions"). The best measure of the quality of a metallic crystal is its electrical resistance near absolute zero. At such low temperatures the resistance is due primarily to scattering of electrons from lattice defects since scattering from lattice vibrations is suppressed. The resistance of the whiskers was still dropping at 1.3 kelvins, our lowest easily obtainable temperature, and so we planned on further measurements in our dilution refrigerator, which can attain temperatures as low as 0.01 kelvin (see "Getting Close to Absolute Zero"). But first we worked on improving sample quality and size.

In August we cooled the UPt_3 in the refrigerator but obtained no data because of a problem with the electrical leads. Since experiments in the refrigerator are extremely time-consuming, we delayed another attempt on UPt_3 until samples of several other materials were ready and could be cooled at the same time.

On Friday, September 16 all the samples were in the refrigerator and well chilled. Since UPt_3 held (we thought) the least promise of interesting results, it was the last sample to be measured. So it was at 4:30 p.m. when we saw the resistance of the whisker start to plummet at 0.54 kelvin (Fig. 1). Then began a month of intensive effort to confirm what would be a remarkable discovery—the coexistence of superconductivity and enhanced spin fluctuations.

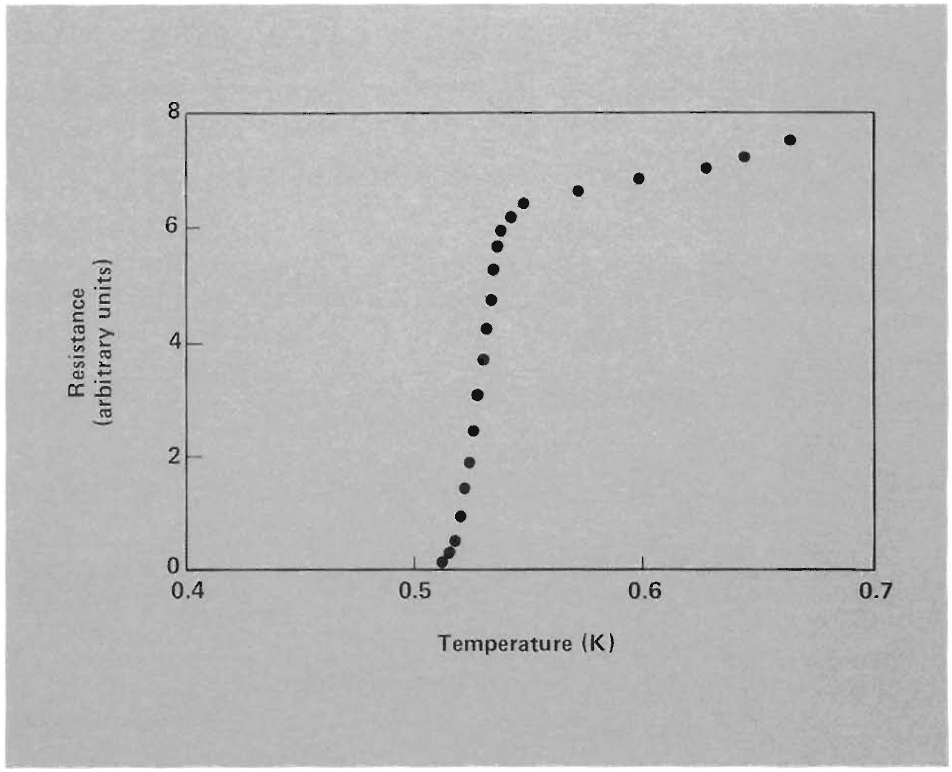


Fig. 1. Data obtained during the first measurement of the low-temperature electrical resistane of a single crystal of UPt_3. The abrupt disappearance of resistance, a sign of superconductivity, was quite surprising since we regarded UPt_3 as a likely spin fluctuator.

Were We Right?

Our first concern was whether the observed zero resistance was due to UPt_3 itself or to some other, undetected superconducting phase. Even if present as only a minor constituent (say 1 percent), such a phase can produce misleading indications of superconductivity in measurements of both resistance and magnetic susceptibility. The simplest test for bulk superconductivity (that is, of the major phase) is to measure the susceptibility of the sample as a ground powder. Grinding breaks up any field-excluding layers formed by a superconducting minor phase, and the measured susceptibility more truly represents the behavior of the major phase.

We immediately carried out this test on a ground powder of UPt_3 , using an apparatus cooled by simple evaporation of liquid helium-3 and thus much less time-consuming than the dilution refrigerator. By Sunday,

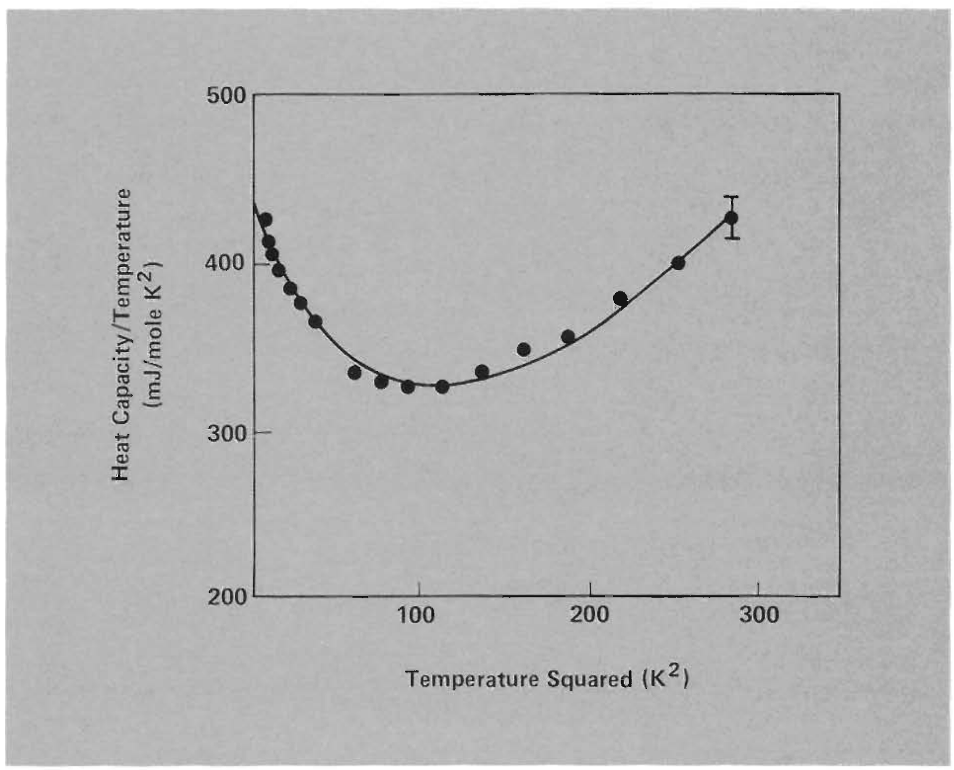


Fig. 2. Specific heat data for unannealed UPt_3 whiskers at temperatures greater than 1.5 kelvins. The curve is a least-squares fit to the temperature dependence predicted for a spin fluctuator. The extremely good fit constitutes strong evidence of enhanced spin fluctuations in UPt_3.

Heavy-Fermion Superconductors

This descriptive name has recently been attached to a class of superconductors in which the effective mass of the fermions (in this context, electrons) responsible for the superconductivity is several hundred times greater than the mass of a free electron. Such a large effective mass implies that the electrons behave less like a gas of independent particles (the usual picture of conduction electrons) and more like a liquid of interacting particles. The unmistakable sign of a heavy-fermion superconductor is an extremely large value of γ, the proportionality constant relating electronic specific heat to temperature. (The effective mass is deduced from this parameter, which is determined experimentally by extrapolating low-temperature specific heat data to absolute zero.)

Two heavy-fermion superconductors are now known: $CeCu_2Si_2$, the first, and UBe_{13}. In each case the superconductivity is surprising (so much so that the fact was initially reported, almost apologetically, in a footnote) since near room temperature the magnetic susceptibility follows a temperature dependence like that of a material with local magnetic moments. Thus magnetism (an ordering of the moments), not superconductivity, is the expected response at lower temperatures.

We became involved in heavy-fermion superconductivity by being the first to grow high-quality single crystals of $CeCu_2Si_2$. These crystals helped to dispel some of the confusion about the properties of this material, which had varied wildly from sample to sample. Then, in collaboration with H. R. Ott and H. Rudigier of Eidgenössische Technische Hochschule–Hönggerberg, we showed that UBe_{13} was another heavy-fermion superconductor. Its properties are very similar to those of $CeCu_2Si_2$ and fortunately vary little among different samples.

The existence of a second example has made heavy-fermion superconductivity more appealing for study but as yet little better understood. More examples must be found before interesting questions about the phenomenon, such as whether *p*-state superconductivity is involved, can be answered. ■

September 18 we had some disappointing news—the ground powder was *not* superconducting down to 0.45 kelvin. More measurements followed. We cooled the powder in the dilution refrigerator but again found no indication of superconductivity, this time down to 0.050 kelvin. We measured the specific heat of UPt_3 at temperatures down to 1.5 kelvins, and the news from this front was good. The data fitted beautifully to the $-T^3 \ln T$ dependence predicted for a spin fluctuator (Fig. 2).

We now knew that UPt_3 was a bona fide spin fluctuator and that the ground powder was not a superconductor. Why, at this point, did we persist with further, perhaps fruitless, tests for superconductivity? We had several reasons. One was the lack of a reasonable suspect for a superconducting second phase. Uranium is a superconductor, but its presence in UPt_3 is not to be expected since two other phases of the uranium-platinum system (UPt and UPt_2, neither of which are likely superconductors) are closer in composition to UPt_3, and a second phase is usually adjacent to the major phase in composition. In addition, crystals in the form of whiskers are generally free of other phases. A second reason was the behavior of a single crystal of UPt_3 prepared by Franse's group in a totally different way than our samples. (Franse had sent this crystal to us earlier as an encouragement to measure its heat capacity in a magnetic field.) We had measured its susceptibility in the dilution refrigerator along with that of the ground powder and found a superconducting transition at 0.35 kelvin. This fact made the negative result from the ground powder more suspect than the positive result from the whiskers. The final reason for persistence was the chance that our initial interpretation was correct. If UPt_3 *was* a *p*-state superconductor, our measurements on a ground powder could easily be misleading since grinding introduces defects into the lattice that would be extremely destructive of *p*-state superconductivity. (*p*-State superconductivity is more strongly inhibited by lattice defects than is *s*-state superconductivity because the effective diameter of the interacting electron pairs is greater and thus encompasses a greater number of defects.)

Fortified by these arguments (hopes?), we proceeded to look for the only sure sign of bulk superconductivity in UPt_3—a large upward step in its specific heat curve. A superconducting second phase present at a concentration of less than about 5 percent (the limit we had established by x-ray diffraction techniques) would produce some increase, depending on its concentration, but the increase would be nowhere near that expected if UPt_3 itself was a superconductor. (The BCS theory predicts an increase of about 150 percent.)

The whiskers we could gather at the time for the specific heat measurement amounted to only 20 milligrams, but, fortunately, we have developed techniques and equipment for measuring specific heats of very small samples. We spent nine days hovering over the refrigerator, and by Friday, September 30 the data definitely showed a sizable discontinuity. However, because of experimental difficulties below 0.3 kelvin, there remained a nagging uncertainty about its precise shape.

Such an important discovery deserved the best possible data, so we decided to repeat the heat capacity measurements, this time using annealed whiskers. (We had learned from susceptibility measurements in the helium-3 apparatus that annealed whiskers had much sharper superconducting transitions, and this increased sharpness would be reflected in the heat capacity curve.) Since we were running out of whiskers, we took the unannealed ones out of the refrigerator, annealed them, and had them cold again by Monday, October 3. That weekend turnaround was the fastest we had ever achieved.

As shown in Fig. 3, the specific heat of our annealed single crystals of UPt_3 increased by only about 50 perecent, and the transition was quite broad (and had been even broader for the unannealed crystals). Nevertheless, an increase of this magnitude unequivocally ruled out the possibility that the superconductivity was due to a minor second phase. We now felt confident that superconductivity and enhanced spin fluctuations coexisted in UPt_3.

During these experiments we had repeatedly attempted to produce better samples and had significantly increased the size of the crystals but not their lattice perfection. In

Single Crystals from Metal Solutions

Given a free choice, any solid-state experimentalist would characterize a material by making measurements on a single crystal rather than a polycrystalline sample. A single crystal more accurately represents the material (since it is free of grain boundaries at which impurities can hide) and is in fact required for measuring the directional dependence of various properties. Yet growing a single crystal can be exceptionally difficult, and a large number of important experiments await the preparation of appropriate single crystals.

Numerous techniques exist for growing crystals, but finding one that works for a particular material can be frustrating and time-consuming. A method we use quite often in our research is growth from slowly cooled solutions of the desired material in a molten metallic solvent. (This method is an easy extension of the observed natural growth of single crystals from aqueous solutions.) We have used as solvents such metals as aluminum, indium, tin, copper, bismuth, and gallium. The solvent provides a clean environment for crystal growth, and the relatively low temperature at which growth occurs often results in low defect concentrations. Offsetting these advantages is the possibility that solvent atoms may appear at lattice sites and in voids of the crystal. In addition, one must find a container that is not attacked by any component of the solution and a chemical to remove the solvent without attacking the crystal. We have built up a collection of workable "recipes" and are constantly including new "ingredients." Still, success demands a certain flair.

When applying this technique to a new material, one unknown is always present: the material may be one that nature simply refuses to provide as nice crystals. Also, the appropriate phase diagram is usually lacking. Then we must rely on educated guesses and hunches, since determining the phase diagram for a system of at least three elements is not a job to undertake merely for exploratory work on crystal growth.

To grow the single crystals of UPt_3, we used bismuth (melting point: 280 degrees Celsius) as the solvent. As usual, the phase diagram for the system was not available. But we knew from published work that UPt_3 has a melting point of 1700 degrees Celsius and is chemically quite stable, that reasonably large amounts of uranium and platinum can be dissolved in bismuth at temperatures on the order of 1000 degrees Celsius, and that compounds of both uranium and platinum with bismuth exist. But the shapes of the uranium-bismuth and platinum-bismuth phase diagrams indicated that these compounds are not exceptionally stable. Our guess—that UPt_3 would crystallize preferentially—was correct, provided that the solution was not cooled below about 1100 degrees Celsius (where a competing crystallization takes place). We obtained good yields by using atomic percentages of uranium, platinum, and bismuth in the ratio of 1:3:4 and an initial temperature of 1450 degrees Celsius. Since that temperature is near the boiling point of bismuth, we sealed the crucible in a tantalum can to prevent its evaporation. We used a crucible of BeO rather than the more usual Al_2O_3 because uranium might attack Al_2O_3 at such a high temperature.

As we improved the technique, we obtained crystals of UPt_3 with a length of up to 1 centimeter and a cross section of 1 millimeter by 1 millimeter. Nature shows her hand here. The material seems always to have a needle-like habit. ■

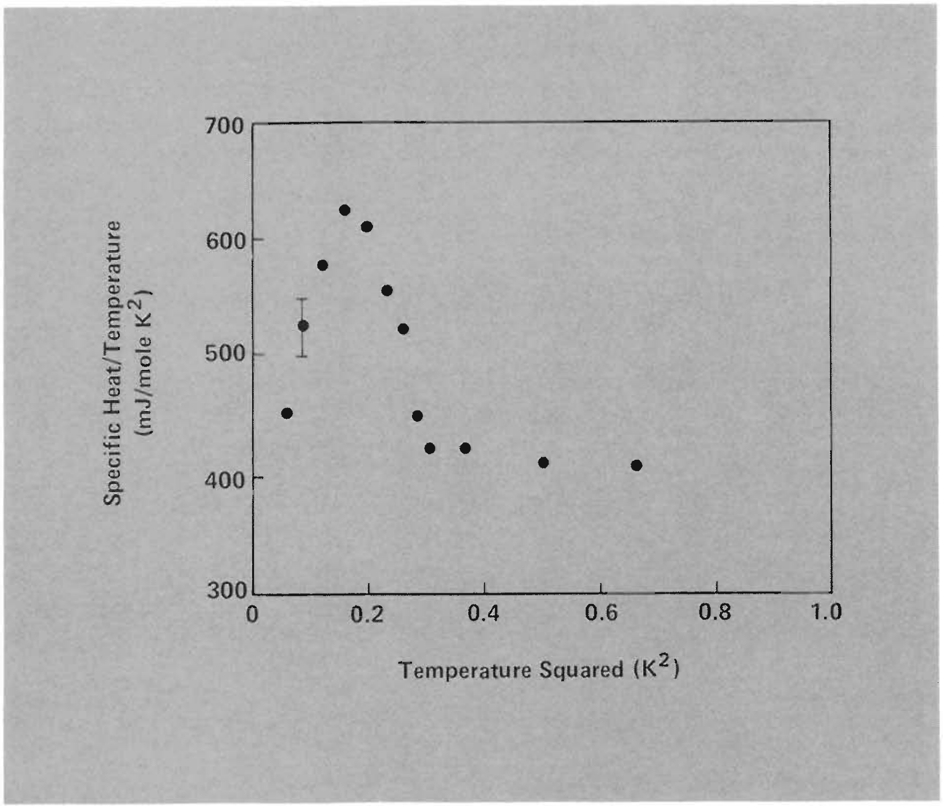


Fig. 3. Specific heat data for annealed UPt_3 whiskers near the temperature at which the resistance of the whiskers fell to zero. The sizable discontinuity rules out the possibility that a minor superconducting phase was responsible for the zero resistance.

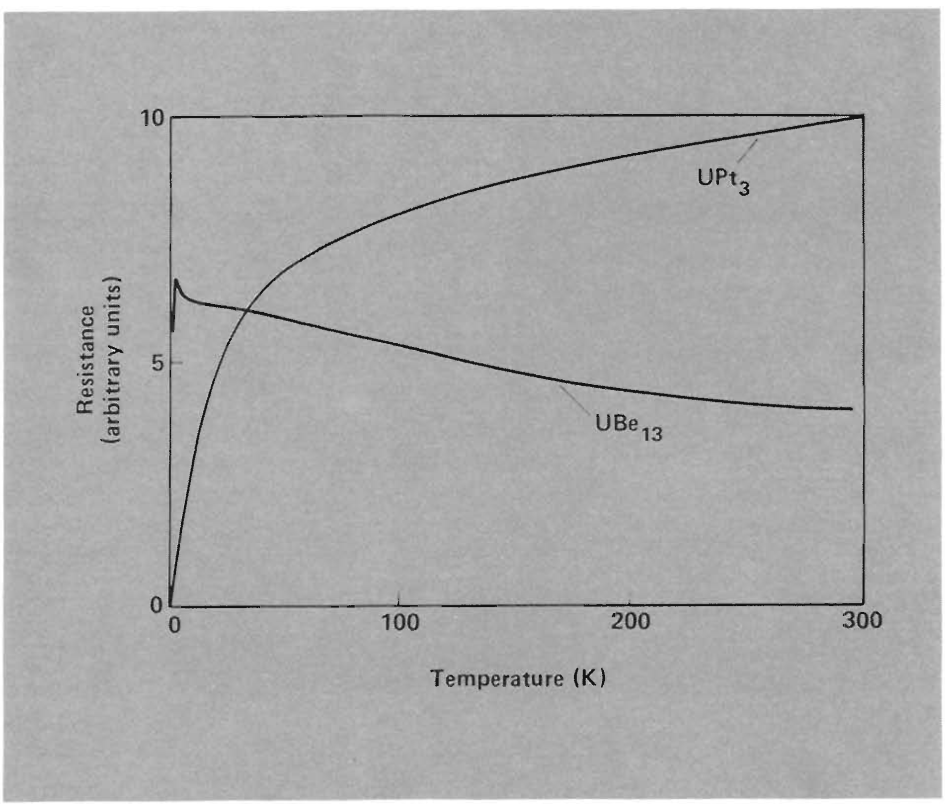


Fig. 4. The remarkably different resistance-versus-temperature curves of UPt_3, a possible p*-state superconductor, and UBe_{13}, a heavy-fermion superconductor. Explaining this and other differences presents an interesting challenge to theory.*

fact, low-temperature resistance measurements indicated that the lattice perfection was near the limit expected for a compound like UPt_3. Therefore we felt that the accuracy of the specific heat data could be increased further only by better calibration of the calorimeter. We ran a piece of high-purity copper and then, as a check on the systematic errors, ran one-half of that piece. We finished the calibration by October 13 and had the manuscript in the mail to *Physical Review Letters* on October 18.

What Next?

We have now two new superconductors, UPt_3 and UBe_{13} (see "Heavy-Fermion Superconductors"), as different from each other (Fig. 4) as they are from all other superconductors (except $CeCu_2Si_2$). Many questions come to mind about these materials; the most intriguing is that of *p*-state superconductivity. Of the tests that have been proposed for this phenomenon, we mention the more obvious.

One test we plan to carry out in collaboration with a group at the University of California, Riverside, is to measure the shift of the nuclear magnetic resonance frequency of platinum-195 in UPt_3. (A similar measurement is already in progress on beryllium-9 in UBe_{13}.) This "Knight shift" is due to shielding of the nucleus from an applied magnetic field by the counter magnetic field of the conduction electrons. The predicted temperature dependence of the Knight shift in the vicinity of the transition temperature is quite different for *s*- and *p*-state superconductors.

A test we have already mentioned is sensitivity to lattice defects. Our measurements on the ground whiskers of UPt_3, although suggestive, need considerable elaboration. In particular, we must demonstrate that the sensitivity to magnetic defects is equal to (rather than greater than, as is the case with *s*-state superconductors) the sensitivity to nonmagnetic defects. The difficulty with this test is finding suitable magnetic impurities to incorporate into the lattices of these materials (nonmagnetic impurities come free). Another test is based on the fact that a "supercurrent" would not flow through a loop containing a junction between an *s*- and a *p*-state superconductor. However, a poor junction would also kill a supercurrent, and good junctions are extremely difficult to prepare.

Clearly, much work remains to be done, but the data now available at least refute the conventional wisdom of a dichotomy between superconductivity and a tendency

Getting Close to Absolute Zero

Liquid helium-4 and helium-3 rank with vacuum as sine qua nons for many scientific experiments. Some phenomena occur only at temperatures achievable with these unusual liquids, and others become much more tractable to theoretical interpretation.

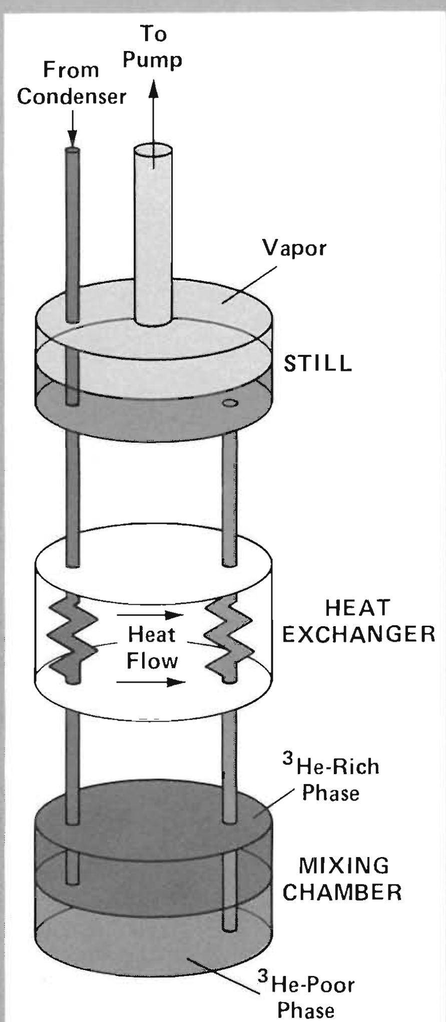
Gaseous helium-4 occurs on the earth as a product of alpha decay and is found in reasonable concentrations in some natural gas fields. It was first liquefied in 1908 by Heike Kamerlingh Onnes (whose discovery of superconductivity soon followed). Temperatures between about 1 kelvin and the boiling point of liquid helium (4.2 kelvins) can be attained simply by pumping on the liquid. The atoms crossing the liquid-vapor phase boundary absorb heat, and the remaining liquid cools. Somewhat lower temperatures (routinely down to between 0.5 and 0.3 kelvin, depending on the system) can be reached by pumping on liquid helium-3. (This stable but naturally extremely rare isotope is a by-product of the manufacture of nuclear weapons.) For both liquids the lower temperature limit is set not by freezing (as it is for normal liquids) but by a rapid decrease in vapor pressure.

Even lower temperatures (down to about 0.005 kelvin) can be reached with a "dilution" refrigerator. This device exploits the natural tendency of liquid helium-3 to "evaporate" into the "mechanical vacuum" of liquid helium-4. (These two liquids, despite both consisting of isotopes of the same element, interact very weakly because one (helium-4) follows Bose-Einstein statistics and the other follows Fermi-Dirac statistics.) The atoms of helium-3 absorb heat (corresponding to the heat of evaporation) as they cross the phase boundary between these two dissimilar liquids. The lower temperature limit is set not by a decrease in the "vapor pressure" as the temperature falls but by a decrease in the heat of "evaporation."

The accompanying diagram illustrates schematically the continuous operation of a dilution refrigerator. Liquid helium-3 dissolves in liquid helium-4 in the mixing chamber, and the dilute solution is pumped to a heated still where helium-3 evaporates preferentially. For economy the helium-3 is condensed and the liquid returned to the system. The photograph shows author Jeffrey O. Willis examining a UPt_3 whisker in the cryostat of the Physical Metallurgy Group's dilution refrigerator. A dewar containing liquid helium encloses the cryostat when the refrigerator is operating. About twenty-four hours are required to cool a sample to the desired temperature.

Temperatures in helium-3 and helium-4 evaporation refrigerators are determined simply by measuring the vapor pressure. Thermometry in a dilution refrigerator involves use of a material whose magnetic susceptibility is known to be quite closely inversely proportional to the temperature. The susceptibility versus temperature curve for this material is calibrated against vapor pressure measurements in a helium-3 evaporation refrigerator, and lower temperatures are obtained by extrapolation. ■

toward magnetism. Granted, the two phenomena had been found to coexist in $ErRh_4B_4$, but in that material they originate on different electrons. Now in UPt_3 spin fluctuations and superconductivity are known to coexist on the same electrons and at the same temperature. These results breathe new life into experimental and theoretical studies of superconductivity. Perhaps David Pines' interpretation is correct, and UPt_3 is a metallic analogue of liquid helium-3. ■



AUTHORS

Gregory R. Stewart has shifted the emphasis of his research twice since joining the Laboratory in 1977: from superconductors, particularly those with high transition temperatures, to nearly magnetic materials and, finally, to the recently discovered heavy-fermion superconductors. The connection found in UPt_3 among these diverse interests joins the finding (with L. R. Newkirk and F. A. Valencia) of the new A-15 structure superconductor Nb_3Nb as his most exciting discoveries at Los Alamos. Author of more than fifty refereed papers since 1977, Greg is a specialist in specific heat measurements but enjoys learning other research techniques, such as using high magnetic fields to alter the electronic properties of materials. He earned his B.S. from Caltech and his Ph.D. from Stanford. A postdoctoral appointment followed at the University of Konstanz in West Germany, where his research included resistivity and Hall-effect measurements on solar-cell materials. Recently he spent a three months' sabbatical at Kernforschungszentrum Karlsruhe.

Zachary Fisk was educated at Harvard University and the University of California, San Diego. He received his Ph.D. in physics at the latter in 1969 under Bernd Matthias. A postdoctoral year at Imperial College, London, was followed by a year as Assistant Professor of Physics at the University of Chicago. He returned to UCSD, becoming Research Physicist and Adjunct Professor of Physics before joining the Laboratory in 1981 as a staff member in the Physical Metallurgy Group of what is now the Materials Science and Technology Division. His research interests include the low-temperature electrical and magnetic properties of metals and the growth of single crystals of these materials. The latter interest originally developed from a consulting agreement with Bell Laboratories.

Jeffrey O. Willis earned his B.S. and Ph.D degress in physics from the University of Illinois, Urbana-Champaign, in 1970 and 1976, respectively. He then spent two years at the Naval Research Laboratory in Washington, D.C., as a National Research Council Research Associate studying the superconductive properties of new materials at ultralow temperatures. In 1978 he joined the Laboratory as a postdoctoral fellow in the Condensed Matter and Thermal Physics Group of the Physics Division. There he investigated the magnetic and superconductive properties of materials primarily using the Mössbauer effect. In 1980 he became a staff member in the Physical Metallurgy Group of what is now the Materials Science and Technology Division. He is currently engaged in the study of new materials at ultralow temperatures, using the dilution regrigerator, and at very high pressures, using diamond anvil cell and other techniques. He is a member of the American Physical Society.

AUTHORS

James L. Smith received his B.S. in physics from Wayne State University in 1965 and his Ph.D. in physics from Brown University in 1974. He has been a staff member in the Physical Metallurgy Group of what is now the Materials Science and Technology Division since 1973. In 1982 he was appointed a Laboratory Fellow for his scientific insight and experimental expertise. His work began at Los Alamos on materials at dilution refrigerator temperatures, his special expertise then. That work evolved into addressing the question of how superconductivity in elements from the left side of the periodic table crosses over to magnetism in elements from the right side. This has led to interesting speculation on such things as catalysts and the stability of stainless steel. He is now a leader in the field of actinide materials and gives several invited talks on the subject each year at various conferences and workshops. Despite his experience, he was completely surprised by the behavior of UPt_3 reported in this article.

Further Reading

P. W. Anderson and P. Morel. "Generalized Bardeen-Cooper-Schrieffer States and the Proposed Low-Temperature Phase of Liquid He^3." *Physical Review* 123(1961):1911-1934.

R. Balian and N. R. Werthamer. "Superconductivity with Pairs in a Relative *p* Wave." *Physical Review* 131(1963):1553-1564.

W. A. Fertig, D. C. Johnston, L. E. DeLong, R. W. McCallum, M. B. Maple, and B. T. Matthias. "Destruction of Superconductivity at the Onset of Long-Range Magnetic Order in the Compound $ErRh_4B_4$." *Physical Review Letters* 38(1977):987-990.

F. Steglich, J. Aarts, C. D. Bredl, W. Lieke, D. Meschede, W. Franz, and H. Schäfer. "Superconductivity in the Presence of Strong Pauli Paramagnetism: $CeCu_2Si_2$." *Physical Review Letters* 43(1979):1892-1896.

Angelo L. Giorgi, Gregory R. Stewart, James L. Smith, and Bernd T. Matthias. "High-Temperature Superconductivity: A Metallurgical Approach." *Los Alamos Science* Vol. 1, No. 1(1980):28-39.

G. R. Stewart. "Measurement of Low-Temperature Specific Heat." *Review of Scientific Instruments* 54(1983)1-11.

J. L. Smith and E. A. Kmetko. "Magnetism or Bonding: A Nearly Periodic Table of Transition Elements." *Journal of the Less-Common Metals* 90(1983):83-88.

P. H. Frings, J. J. M. Franse, F. R. de Boer, and A. Menovsky. "Magnetic Properties of U_xPt_y Compounds." *Journal of Magnetism and Magnetic Materials* 31-34(1983):240-242.

H. R. Ott, H. Rudigier, Z. Fisk, and J. L. Smith. "UBe_{13}: An Unconventional Actinide Superconductor." *Physical Review Letters* 50(1983):1595-1598.

G. R. Stewart, Z. Fisk, and J. O. Willis. "Characterization of Single Crystals of $CeCu_2Si_2$. A Source of New Perspectives." *Physical Review B* 28(1983):172-177.

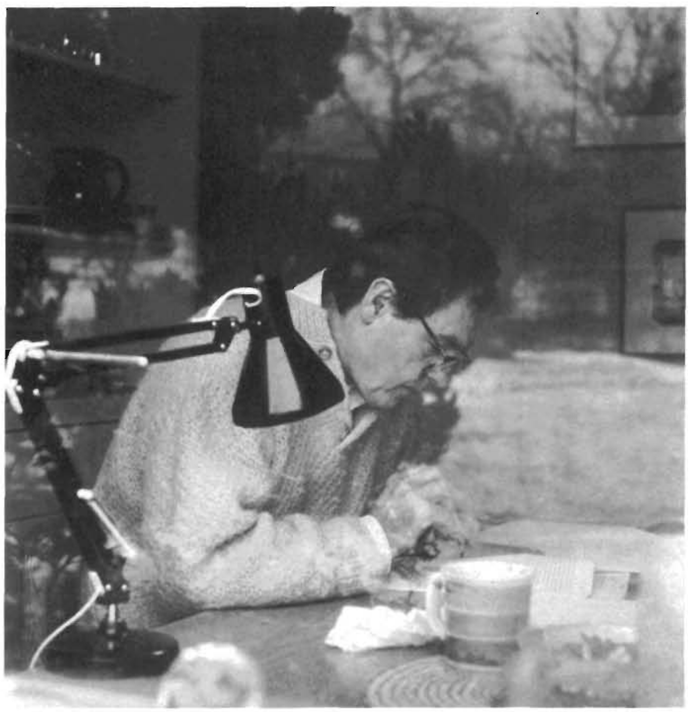
"Second Heavy-Fermion Superconductor." *Physics Today*, December 1983, pp. 20-22.

G. R. Stewart, Z. Fisk, J. O. Willis, and J. L. Smith. "Possibility of Coexistence of Spin Fluctuations and Superconductivity in UPt_3." *Physical Review Letters* 52(1984):679-682.

straight man *for* nonlinearity

An interview with Alwyn Scott

When Los Alamos Science *arranged to interview Alwyn Scott, his secretary warned us that if we weren't careful, we'd wind up being the ones interviewed. He listens well. You have a sense he is absorbing everything you're saying and judging by impeccable standards what's valuable and what is not. As Chairman of the Center for Nonlinear Studies, he is popular for this very reason. In fact, all his time could easily be occupied by the myriad projects of the Center were it not for his fierce need to do his own research (he manages this during morning hours at home). Energetic, efficient, enthusiastic, he nevertheless confesses his antipathy for administrative matters. "If you were to look for me at a staff meeting," he comments drily, "I'd be the one doing calculations under the desk."*



Alwyn Scott holds degrees in physics, engineering, and electrical engineering from the Massachusetts Institute of Technology. After earning his doctorate he began teaching at the University of Wisconsin and remained on its staff for twenty years, although his work has taken him at various times to Bern, Switzerland; Naples, Italy; Sendai, Japan; Woods Hole, Massachusetts; Tucson, Arizona; and Copenhagen, Denmark. Throughout his career he has been concerned with the theory of nonlinear wave propagation and its applications; he has written books and holds patents in that field. "My work is sometimes quite theoretical and sometimes experimental . . . my aim, however, is always to bring theory and experiment closer together in applied science." In 1981 Alwyn Scott became Chairman of the new Center for Nonlinear Studies at Los Alamos, where his special interest in solitons in biology has evolved into a project reported in this issue.

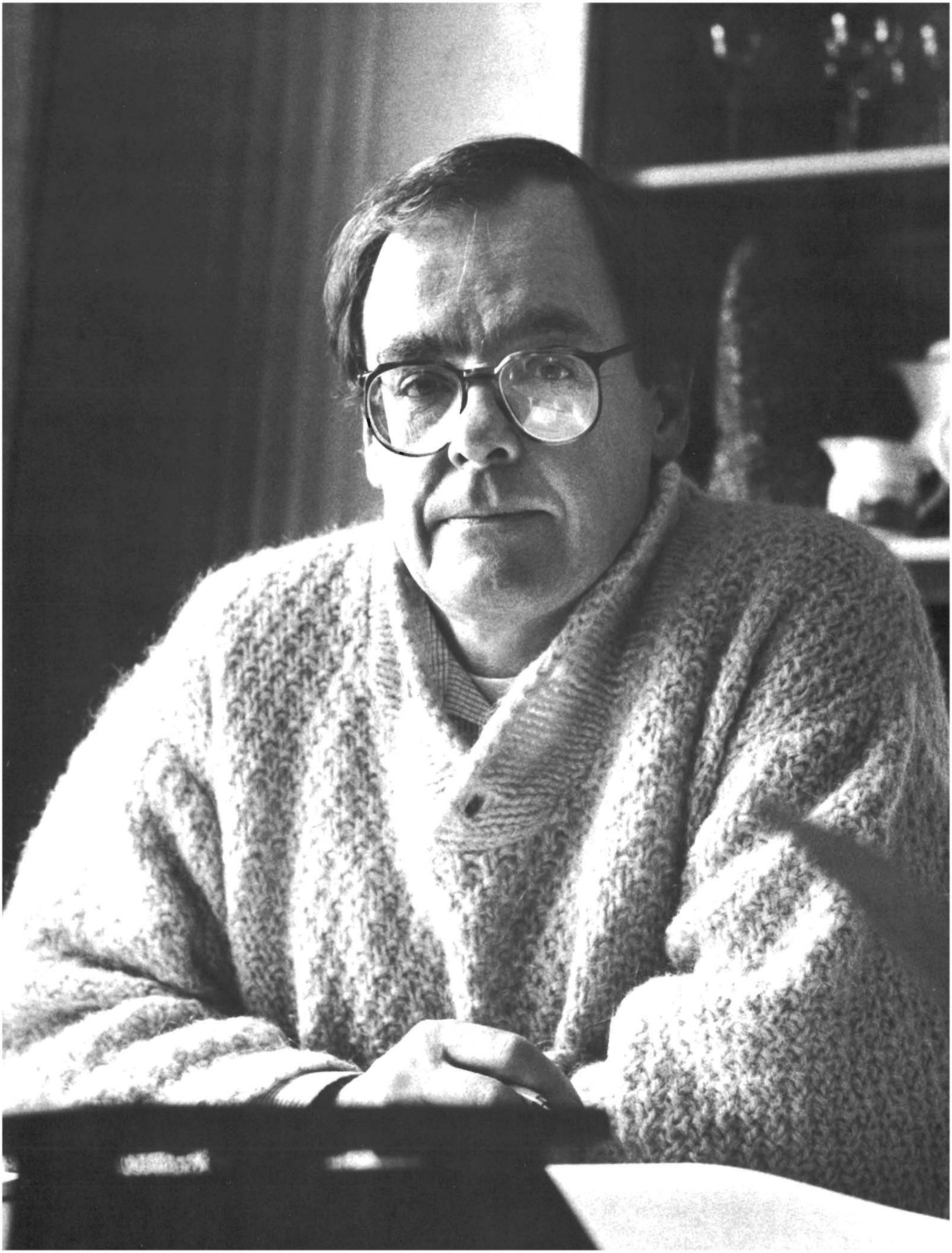
SCIENCE: You've been exploring solitons in biology for quite a while. What's the story behind this project?

SCOTT: In 1978 I attended a soliton conference in Sweden (it was for a week or two out in the country somewhere), and the Soviet physicist Davydov was there. Davydov had originated this work, and he gave a talk on it. He had a model showing how biological energy could be self-trapped as solitons in helical protein and thus provide a mechanism for energy transport and storage.

SCIENCE: Why is that important?

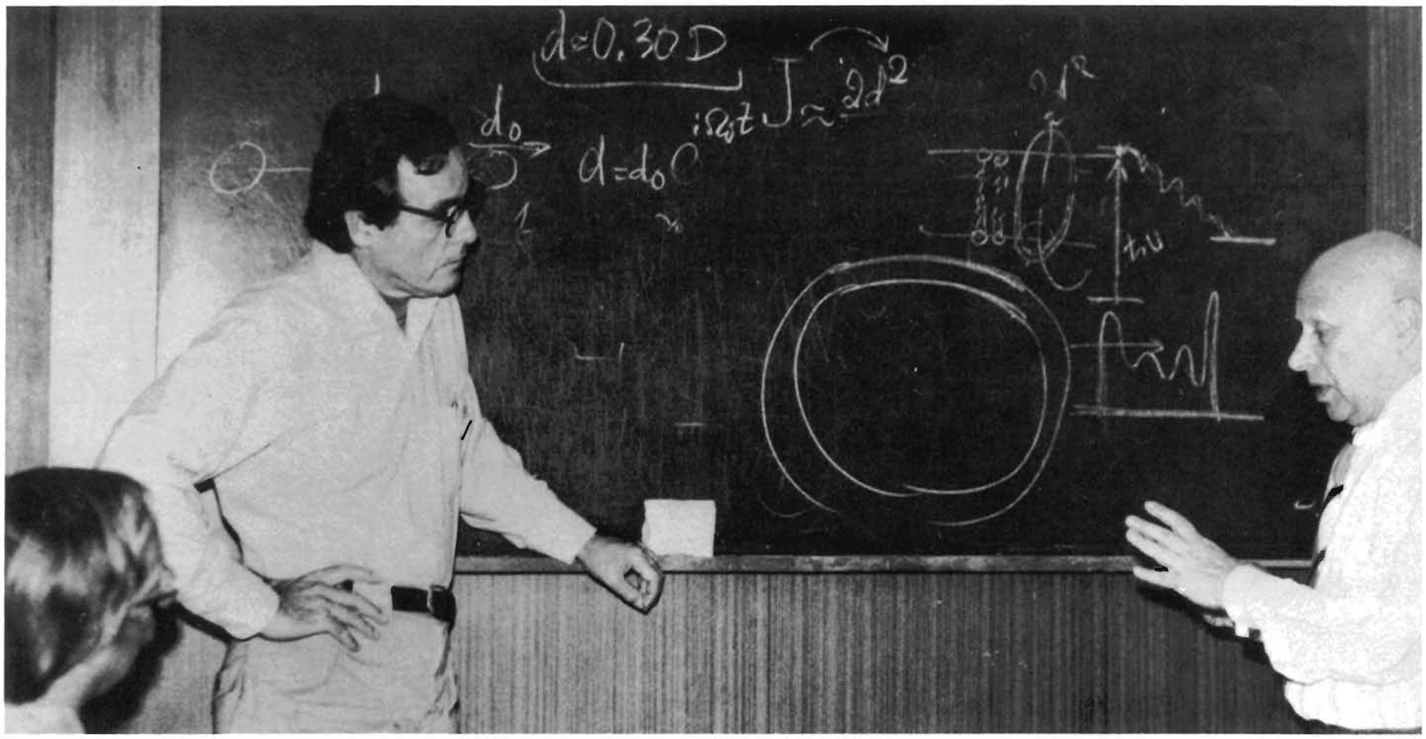
SCOTT: It's a whole change in thinking. A fundamental unanswered question in biology (labeled in 1973 the "crisis in bioenergetics") is how energy is stored, transported and transduced, and used to generate motion or change in form. Davydov showed how nonlinear concepts that are accepted in solid-state physics can be used in biochemistry. What he said seemed reasonable.

At the end of the conference Davydov gave me a stack of reprints and asked me to discuss his ideas with some biochemists here in the States. I tried, but I simply couldn't make clear to them what Davydov was talking about. So it seemed it might be useful to do some numerical calculations and show pictures, or a movie film or



Davydov showed how nonlinear concepts that are accepted in solid-state physics can be used in biochemistry.

INTERVIEW



something, of what was going on. Since I knew Mac Hyman and Dave McLaughlin at Los Alamos, we ran some calculations on the CDC 7600 computer here. These calculations started showing threshold effects that were interesting to investigate further, and that's how I was drawn into the problem.

SCIENCE: So originally you were to be a conduit between Davydov and some biochemistry people. Do you want to say some more about the ideas behind your current project and what you hope from it?

SCOTT: The impression I got from trying to communicate Davydov's ideas to the biochemists was that they are, in many cases, ill-prepared by their background and training to think about dynamical effects in large biochemical molecules like proteins or DNA. They know a lot about the structure of these molecules, but they are trying to go from a knowledge of structure to knowledge of function without thinking about dynamics. One thing I hope to see coming out of our project is a contribution from Los Alamos, with all its high-speed computers, to the understanding of nonequilibrium biochemical dynamics.

Of course, it's a dangerous quest because we physical scientists are very naive about biological reality. Physical scientists often come into some area of biological science and are just—you know—typically arrogant. But we do have good contacts with the Life Sciences Division. Mark Bitensky, their Division Leader, has been connected with this project since the beginning, and Scott Layne, one of our postdocs working on it, is an M.D.

SCIENCE: When I first heard about your project, I was intrigued that physical scientists were going into this biological area. Initially I wondered whether they just had some cute mathematical model they were somewhat arbitrarily trying to plant in these proteins.

SCOTT: This has been typical of what is called biophysics or mathematical biology. You find a physical scientist who has some idea or differential equation and then goes around in the biological world looking for a place to apply it, making all sorts of assumptions and approximations that aren't justifiable from the biologist's point of view. Biological scientists are sensitive to that; they're concerned about somebody who just waltzes into their field from the physical sciences. And rightly so, I think.

SCIENCE: In moving into this area, did you have any specific intuition in terms of the biological context? Or have you gained some?

SCOTT: I've always been interested in systems in the real world that are nonlinear, because they can do more interesting things. Of course, that makes them harder to solve because the unexpected can happen. When you look around in the biological world, essentially everything is, somehow or other, nonlinear.

Proteins, for instance, are marvelous molecular machines. This alpha helix that we are studying is a major constituent of hair. It's a long, helically shaped, fibrous protein, and you can readily see why its structure is appropriate to hair. In this case, going from structure to function is not a difficult leap to make. But if you think more generally of what is happening with proteins, of the many kinds of dynamic roles that they're playing all the time in every cell, then the leap from structure to function cannot proceed without considering the dynamics—the how. And immediately the thought arises that nonlinear effects are playing a vital role. It seems that it must be so—of course, that's different from proving it.

SCIENCE: What do you expect from this particular tack that's being taken, that is, solitons in proteins somehow transferring energy?

SCOTT: Well, in the first place, I would describe our work as more general than trying to study solitons in proteins. As I see it, Los Alamos Laboratory is in a unique position to analyze biological materials from the experimental, theoretical, and computational points of view. With the experimental physicists here and all their equipment on one side and the computing facilities on the other side, we can take material for which the structure has been determined, look at it in the laboratory, and then analyze it on the computers. I see that as broader than "looking at solitons in the alpha helix," although that was Davydov's original, seminal idea.

SCIENCE: So you'd like to look at solitons in materials in general?

SCOTT: Rather at nonlinear effects that are related to solitons. For example, the notion of a soliton makes sense if you think about a protein like this alpha helix as a major constituent of muscle. The protein runs along in a strand, and you can think about a region of excitation that travels along it, rather the way a pulse travels along a nerve fiber. However, if you have a globular protein in which the protein folds up on itself, the notion of a neat solitary wave may no longer apply. The dynamics may be related to what Davydov was talking about along a linear protein, but geometrically it will be a lot more complicated.

SCIENCE: Peter Lomdahl has some interesting numerical results for the globular protein lysozyme, but he is reluctant to discuss it in this issue.

SCOTT: There are four or five refinements that he'd like to have in the model, and those could easily change details of his results. One thing Peter's calculations do show is that nonlinear calculations for such a complicated object are not at all out of the question: it's something that can be done here at Los Alamos with our computing facilities.

SCIENCE: Are there practical engineering applications for biological solitons?

SCOTT: People are talking about "biochips" as a way to use biological molecules for computing at much higher density and smaller scale. And Davydov's soliton is one of the mechanisms that people are thinking about at the Naval Research Laboratory and in industry for doing the computing.

SCIENCE: How do you feel about that?

SCOTT: Optimistic—and, you know, throughout the history of scientific prediction, optimists have almost always erred on the conservative side. I see our work at Los Alamos helping provide the scientific foundation that will make these developments possible.

SCIENCE: I want to switch the topic now. You've traveled widely and worked at many places around the world. Why did you choose to come to Los Alamos?

SCOTT: Well, the Center for Nonlinear Studies started up with aims that I share. I've always been very keen on attempts to understand how nonlinear dynamic effects can play a role in real scientific problems, but for a long time this has been an area of science that was pushed into the background.

SCIENCE: Why was it pushed into the background? Because people couldn't make any progress?

SCOTT: It really wasn't fashionable until 1970 or so. There are a lot of fashions in science, as in other things. But now here is the Center with the aim of encouraging young people to get started in the field. I feel kind of psychically in tune with the charter of the Center.

SCIENCE: Can you give us more details about the aims of the Center?

SCOTT: As I see them? Sure. One is to increase, to improve contacts among nonlinear scientists at the Laboratory and in other parts of the country, particularly those at the universities, and in other parts of the world. The Center is running quite a number of workshops, a large conference every year, and sponsoring the visits of many fine scientists every year. A second aim is to stimulate nonlinear science inside the Laboratory, to start programs here that haven't been going on before, and to help branches of the Laboratory that haven't been interacting with one another in the past but could do so fruitfully. I think this biological project is a good example because it's bringing together people in the Chemistry Division and in the Life Sciences Division and people who are engaged in computational physics—they're all working together on the same project. A third thing is the one I mentioned: the Center is trying to get young people who are starting out in science into the area of nonlinear science, nonlinear dynamics. One of the important things going on in the Center is a vigorous postdoctoral program.

SCIENCE: What are the areas in the Laboratory where nonlinear dynamics is important and needs to be investigated?

SCOTT: Almost everywhere. The traditional areas of Laboratory activity—combustion and explosions—are fundamentally nonlinear. Energy released by a moving front in turn reacts to drive the front. The candle flame is a useful paradigm. Its heat releases energy (vaporized wax) from the candle, and this energy acts to keep the flame hot. And this process of nonlinear diffusion is just what is going on as a nerve impulse propagates along a fiber. Electrostatic energy

... now here is the Center with the aim of encouraging young people to get started in the field. I feel kind of psychically in tune with the charter of the Center.

stored in the cell membrane's capacitance is released and drives the pulse. Michael Faraday said once that the best way to begin understanding natural philosophy is to study the candle.

SCIENCE: What can the Center for Nonlinear Studies do to help in the area of understanding explosions? The Laboratory has had a large research program for forty years.

SCOTT: It does sound presumptuous when you put it that way. One question that I feel we can help with there—and only help with—is to determine the level of complexity at which, at the present time, it is possible to do real science. For example, for what mixture of suitably simplified explosive gases is it possible to write down all the equations and then to solve them using a combination of theoretical and numerical techniques, so that one can predict from first principles what's going on. And to determine where it is necessary to use rules of thumb and heuristic computational procedures, which may work well and be refined over a period of time but which do not proceed from first principles. I think it's important to understand where that line is, and people at the Center have been trying to help define it.

SCIENCE: That's interesting. Is there any group working in fusion, in plasma physics?

SCOTT: Yes, because this is another area where the physics is extremely nonlinear. People who have been attempting to design large energy-producing machines have been concerned for a long time with solving their nonlinear systems, with understanding what stationary states they might have and what instabilities might arise. Some of the work that has been going on in the Center is of interest to those people. The traditional approach has been to determine conditions for instability of particular stationary states. Recently Darryl Holm and collaborators in the University of California system have developed general techniques for determining stability of classes of stationary states. This is exciting and potentially very important.

SCIENCE: Do you have much interaction with the University of California?

SCOTT: Thanks for asking. Yes, we have a special program to encourage scientific collaboration between UC and the Laboratory. They have made funds available for faculty to spend time here, and we are helping Los Alamos staff members make arrangements for research visits to UC campuses.

SCIENCE: It's been suggested that people outside the Laboratory know the Center much better than people on the inside. How do people become associated with the Center?

SCOTT: There are various forms of involvement, from having a sympathy with the aims of the Center and some collaboration with one of the programs to being fully supported, somehow or other. There are postdocs; there are guests for various lengths of time; there are people inside the Laboratory being partially supported by money from the Center in order to work on certain projects. There isn't one typical form of association, and I think that's good. One of the nice things going for the Center is that the budget is flexible: thus it's possible to do whatever makes sense to do without concern for precedent.

SCIENCE: So projects just evolve, and people merge into them?

SCOTT: It's diffuse. A typical way for a person to get involved is that he or she has a subject for a workshop. We talk about it, and it is interesting to invite so and so from New York, so and so from Britain, so and so from California, and so and so from Peoria because these are the best people in the world in that field. We can decide to do that, and a month or two later they are here, talking to one another, and in some cases getting acquainted for the first time. People in the Laboratory who feel that they could become interested in the subject get drawn into the workshop and have the advantage of interacting with all these people. And maybe something starts up.

SCIENCE: And then there would be money to bring postdocs in that particular field to continue the work.

SCOTT: Yes, if it makes sense.

SCIENCE: Will you tell us some of the topics that are subjects of research at the Center now?

SCOTT: Well, we don't want to be totally unorganized, so we have several research themes which give directions to the effort. At present there are five themes: reactive flow; instabilities, material interpenetration and mixing; coherence and chaos in dynamical systems; polymers in synthetic metals; and energy transport mechanisms in biological polymers.

SCIENCE: What's happening that is really exciting?

SCOTT: We've talked some about the first two. The work on chaos is also very interesting at the present time.

SCIENCE: Why?

SCOTT: In many areas of science one wants to understand the dynamics of a dissipative system that has many degrees of freedom and that is being supplied with energy. One of the plasma machines, for example, or the surface of the ocean: on very large scales of distance and time the dynamics of the earth's crust driven by internally generated heat and on very small scales the dynamics of a protein molecule driven by metabolic energy. If the rate of energy input is raised above a certain level, the dynamics becomes chaotic. But its behavior can be described rather simply because the motion in the phase space remains on an attractor with a dimension that is often much less than the number of degrees of freedom of the system.

SCIENCE: Why is that important?

SCOTT: It means that the number of dependent variables necessary to describe the chaotic motion is much smaller than one would expect. In fact, numerical techniques are being developed at the Center to determine the dimension of the attractor from the chaotic motion of a single variable. Doyne Farmer and Erica Jen, working in collaboration with experimentalists at the University of Texas, recently found that the attractor for Couette-Taylor flow has a fractal dimension less

than five when the Reynolds number is 30% above the onset of chaos.
SCIENCE: You mentioned polymers as a research theme. Isn't this more appropriate to the Center for Materials Science?
SCOTT: It certainly is appropriate for them and, in fact, this effort is being carried on as a collaboration between the two centers. Roughly speaking, the theoretical work is in the Center for Nonlinear Studies and the experimental activities are in the Center for Materials Science. There has recently been some exciting experimental progress—Mahmoud Aldissi and Rai Liepins have reported the synthesis of polyacetylene in solution. This could be a key step in the commercial production of plastic metals. For example, it could be used to reduce the reflection of airplanes to radar, could be used in large but lightweight batteries for electric automobiles, etc.

SCIENCE: You have said that you like the work of the Center because it is really fun and worthwhile. You've also said that work in the nonlinear field is different than work in many scientific, technical areas today. Will you expand on that?
SCOTT: Well, in many ways World War II was good for science, for physics and mathematics in particular. After the war governments recognized that science was an important source of national power. For that reason they began to support science heavily, especially in the United States. But I don't see it as an unmixed good. What happened was that science changed from something that people did because they liked it—before World War II there wasn't any other incentive—into a reasonable way to make a living. And so although the quantity, the amount of scientific activity, has vastly expanded since World War II, the general level, and that certainly includes the morale, is not the same.

But nonlinear science was unfashionable, as I mentioned before, not the kind of work that *serious* scientists did. And so the people who were getting into this field in the late fifties and in the sixties were doing so not because they thought it was going to be useful for their careers but because they felt, regardless of what other people said, that the world really wasn't flat, that it was kind of curved. To these people it was necessary to do something that was important rather than to go along with everyone else.

So when people in nonlinear science finally (around 1972) began to get together and find out that they had been thinking the same ideas and working on the same problems for ten or even fifteen years, there was a real recognition there, a kind of family feeling. That feeling continues to set the tone of the activities in this area.
SCIENCE: How did you get into nonlinear waves?
SCOTT: I was an undergraduate in physics, and I remember at that time wanting to find some particle-like solutions for the electromagnetic equations. Maxwell's equations, of course, are linear, but it seemed that the electron should somehow be part of the field. Of course my efforts were totally unsuccessful. There wasn't any way of getting a localized dynamic entity as a solution of Maxwell's equations because the basic solutions of these linear equations extend over all space.

Then when I had completed my bachelor's degree in physics (and I made a solemn oath that I was never again going to be inside a university as long as I lived), I worked in New York City for a couple of years in an engineering job. I remember at that time having interest in the nerve problem: what was going on in the brain, what was really happening with the pulses? I can remember going into the library there on Fifth Avenue, past the lions, and getting Rashevsky's book out and reading it and starting to think about how nerves work. But you know, I had never learned or thought anything in a formal sense about nonlinearity.

A few years later I went back to graduate school and was working

What was missing from everyone's intuition . . . was the notion that because of the effects of nonlinearity, energy could organize itself, focus itself, into spatially localized packages. . . .

on an experimental thesis that involved making some very large Esaki diodes, tunnel diodes. (The Esaki diode is a semiconductor device that essentially shows a negative resistance on a surface.) I was making these things and analyzing them from a linear point of view. Then a Japanese visitor, Professor Nishizawa from Tohoku University, remarked that they were also analyzing large Esaki diodes, but from the point of view of nonlinear wave theory. I had never even heard those words before, but they rang a bell. That was in 1959, and I began spending a lot of time in the library, getting going on my own in the nonlinear direction. And also becoming somewhat angry over the sort of formal graduate training I had received, because although it was recognized that there were nonlinear effects in the world, the presumption was that these really weren't a problem. Whenever nonlinearities appeared, you just divided the nonlinearity up into piecewise linear sections, solved the linear problem for each section, and then matched these solutions at the angle points. That was all there was to it, nothing to worry about. That was the way it was taught, and that's completely wrong.

SCIENCE: Along that line, do you want to explain the nature of nonlinear phenomena? Why can't you just add up the pieces?

SCOTT: In a sense what one means by nonlinearity or linearity is a statement about cause and effect. If a dynamical process is linear and you try a certain cause, say cause A, and you get a certain effect, effect A, and then you try a different cause B and you get effect B, then doing cause A and cause B together gives you just the sum, effect A and effect B together. If a process is nonlinear, when you use cause A and cause B together, you get an effect that's not the sum of effect A and effect B.

A simple example of this is a match and a candle. You take a match (that's a cause A) and you light the candle (that's effect A). Then, after you put out the flame, you take another match (that's cause B) and you light the candle (and that's effect B, another flame). But if you put those two things together (if you light it twice), you don't get two flames—you get only one flame. The reason is that the effect is dynamically self-sufficient, and you can see that life itself is rather like that: once it gets started, it can continue on its own.

SCIENCE: Something about nonlinearity seems to go against people's common intuition, their accustomed thought patterns. I'm thinking of the Fermi-Pasta-Ulam computer experiment.

SCOTT: Well, in that case it wasn't going counter to just anybody's intuition—it was going counter to Enrico Fermi's intuition, which didn't very often happen. What was missing from everyone's intuition at that time was the notion that because of the effects of nonlinearity, energy could organize itself, focus itself, into spatially localized packages that would hang together even though they bounced around and hit the walls or hit each other. What was happening in that original, very germinal, computer experiment was that they started a numerical computation by putting all of the initial energy into the lowest mode of vibration on something like a violin string—the single, half-wave sinusoidal vibration. What they observed was that at initial moments of time this energy seemed to be spreading itself out into the various modes, but it eventually refocused and almost completely recaptured the original distribution of energy in the lowest mode.

Kruskal and Zabusky finally explained what was going on. They saw that this original placement of energy could be viewed as organizing itself into a number of soliton components. Since the soliton components would all have individual velocities and bounce back and forth in various ways, after a while the conditions would become just right for them to get reorganized into the positions they were in at the beginning. It would then look as if all the energy were back in the original mode, organized in the original way. But without the concept of nonlinear, self-trapped packages of energy, people just simply didn't have the tools to put that together. Of course, the concept had been around a long time, since 1834, but scientists never took it seriously.

SCIENCE: I'm thinking about the story of the man with the horse.

SCOTT: Yes, that one. I've done a lot of reading about that.

SCIENCE: Did that happen?

SCOTT: Oh yes, that happened, but the story about Russell and the horse gives a wrong impression, the impression that John Scott Russell was out riding one afternoon—maybe watching butterflies or picking wild flowers—and just happened to see this wave on the canal and followed it along for a while and then wrote something cute in his notebook. Actually, Russell, although only twenty-six years old at the time, was well launched into a career as a civil engineer. He was involved in many engineering designs, and among them was the question of redesigning horse-drawn barges for motor power because railroads were in competition with the canals. Russell organized a series of studies to measure the drawing force versus the velocity of a canal boat. Just a year or two earlier a peculiarity of the horse-drawn barge had been discovered: if the horse got the barge up to a certain speed, depending on the canal and so forth, and if you whipped him so that he jumped and got the boat going faster, then he'd actually have to pull less, do less work. Russell knew about this anomaly, and he had an elaborate experimental setup there on a canal near Edinburgh to pull the boat with a constant force, using weights and pulleys, so that he could measure the velocity very carefully. During the experiments one of the ropes broke, and the boat stopped. And that was when he saw this wave going off the front. He immediately jumped on a horse that was nearby and followed the wave, thinking of it as the source of the anomaly and something very important to understand. It turned out that the anomaly was caused by the boat interfering with the solitary wave it was creating. Russell went on to become preeminent in the nineteeth century as a naval architect.

SCIENCE: Where were we? I guess we were discussing the question of scientists taking nonlinearity seriously.

SCOTT: Well, my point is that Russell knew the solitary wave was important—all his life he knew that. Yet it was treated just as a peculiarity, of no scientific importance. If you look at Lamb's great book on hydrodynamics published in 1932, there are only three pages devoted to John Scott Russell's solitary wave. Yet Russell had insisted all his life that it was a revolutionary idea not just in hydrodynamics, but in acoustics and in electromagnetic waves and throughout physics.

SCIENCE: Is it a revolutionary idea? Is it having that impact now?

SCOTT: Well, of course I'm an enthusiast. But objectively I think anyone would agree that it has been revolutionary in solid-state physics since 1970. Just in understanding solids, these nonlinear states, once you start looking for them, are found in many, many different contexts.

Take, for example, a polymer like polyacetylene; in it there's a solitary wave that's essentially a charge transport mechanism. People hadn't recognized it was there before, and yet that's an important property of the material. It makes it possible to produce a lightweight material out of very cheap components, carbon and hydrogen, that has a high electrical conductivity and can store electrical charge. That's extremely important, but it wasn't appreciated, couldn't be appreciated until people had the notion that a nonlinear mechanism could self-trap charge or energy or magnetic flux, whatever is conserved in a particular context.

SCIENCE: Going back, why did you, when you received your undergraduate degree, swear that you would never go back into a university?

SCOTT: Hm . . . it felt like being in a box. I've always liked to work in a real scientific laboratory or to tinker, but I hated laboratory practice in school. You'd go into this situation where you'd just do the same experiments that everybody else had been doing forever. It wasn't a question of learning something or having fun; it was a matter of doing precisely this and getting the *right* answer.

SCIENCE: Did you always do both experimental and theoretical work?

SCOTT: When I first went to work with a bachelor's degree, I did

. . . throughout the history of scientific prediction, optimists have almost always erred on the conservative side.

primarily experimental work. I was working for a company that made microwave tubes, traveling wave tubes, backward wave oscillators, and things like that. I was making these things, making measurements on them, and doing a little theory, but not much. For a number of years I did mainly experimental work.

SCIENCE: And now you mainly do theoretical work?

SCOTT: Four years ago I spent a year doing experiments on electrophysiology at the Zoological Station in Naples. During that year I probably didn't write more than ten equations.

SCIENCE: So for you it hasn't been an evolution from one to the other, from experiments to theory?

SCOTT: I have always enjoyed the opportunity to do some of one and some of the other. It depends, of course, on what needs to be done. One of the reasons I was in Naples was that I had been interested for a long time in some theories of how the nerves work, but I hadn't been able to get electrophysiologists to do experiments that were directed that way.

SCIENCE: Will you tell us something about that work?

SCOTT: The experimental subject was a squid that is common in the ocean, particularly in the Mediterranean. It's a good subject because it has a very large nerve fiber. The aim of the experiment was to study the nerve pulses on this fiber, to treat them seriously as dynamic entities, and in particular, to see how they would interact with one another when they came together at branching points. One of the notions that people have had is that a nerve cell can actually do computing tasks at the branching points where the fibers come into the nerve cell, not just inside the body of the nerve cell. The aim of this research was to see whether the pulses could interfere with and cancel one another at branches.

SCIENCE: If these pulses were solitons, they would just go through?

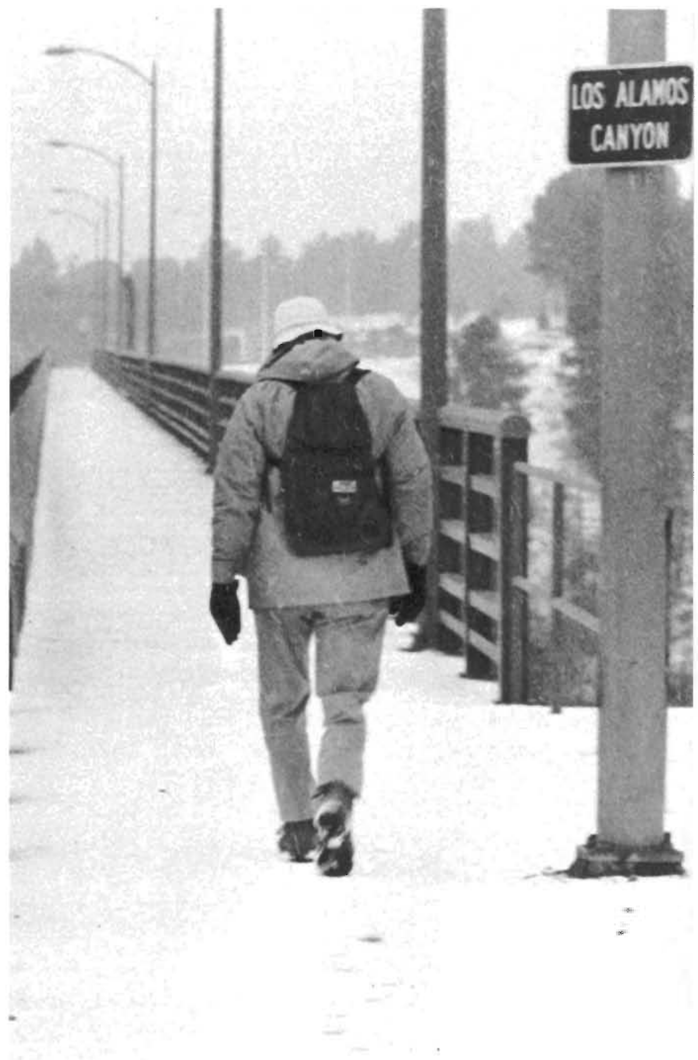
SCOTT: Yes, but they're not solitons because a nerve fiber is very much like the candle we were talking about, always releasing energy and dissipating it. Nerve pulses don't go through each other for the same reason grass fires or candle flames don't go through each other: they've exhausted all the fuel by the time two of them meet.

SCIENCE: What was happening at the branch points?

SCOTT: You can think about a situation where going toward the cell body you have two fibers that are merging into a single branch, like two branches merging into the trunk of a tree. And you can imagine a situation where the geometrical constraint is such that a pulse coming along one of the branches can't ignite the trunk, but two pulses simultaneously coming along both of the branches and getting to the trunk together can ignite the trunk. In computer terms you would speak of that as an *and* connection, because you need pulse A and pulse B in order to get an output pulse.

SCIENCE: Did the intuition for such problems, for working on nerves, come from your work on large Esaki diodes?

SCOTT: Yes, indeed it did. It turns out that the nonlinear wave process that takes place on large Esaki diodes is very similar to what is happening on a nerve fiber. That's when I first became interested in nerve fibers.



SCIENCE: Intuition has to come from some place, doesn't it?

SCOTT: Where motivation comes from is hard to say. I have the feeling that my motivation for the kind of work I'm doing has always been there.

SCIENCE: Then you had an interest in science in your childhood.

SCOTT: A ham set when I was fourteen, an interest in physics and chemistry in high school, lots of tinkering in electronics.

SCIENCE: I wanted to ask you what questions you're now interested in answering, especially in biology. Perhaps we've covered that already.

SCOTT: Biologists really have a whole set of interesting and difficult questions. And some things they are getting answers to. Knowing the structure of DNA and the genetic code is an enormous advance in biological knowledge. But think about the questions they're still faced with: here is an egg and a sperm that come together and start the development of a new organism—how does it somehow manage to take the form that it does? It's not a question of what's going on but of *how* it's going on. It's clearly nonlinear, of course, but the basis for it is not understood at all. That's exciting.

SCIENCE: I've been hearing scientists in many disciplines talking about fractal dimensions, and I associate that term with nonlinear problems and the work of the Center for Nonlinear Studies. I get the feeling that all of science will eventually be permeated by some new understanding about how to look at problems.

SCOTT: That's right, that's right.

SCIENCE: What do you see as the paradigms for modern research?

SCOTT: Well, the whole outlook that an applied mathematician or a physicist has now is totally different from what it was fifteen years ago—and in ways that couldn't have been imagined. In the mid-sixties the general feeling was that a nonlinear partial differential equation was so complicated as a mathematical entity that getting any kind of result beyond a numerical solution on a computer was hopeless. That was in the basket of things that everyone had given up on. Essentially through the work of Kruskal, it became apparent that you could take nonlinear partial differential equations and find very simple ways to analyze them and get exact answers. In a certain sense Kruskal's solution procedure was analogous to solving linear equations. He showed that solitons are a nonlinear generalization of the idea of a Fourier component, or a normal mode of a linear system. In other words, a general solution to a certain class of equations could be rigorously constructed by using solitons as the constituent entities. So, many problems considered hopeless turned out in many cases to be, if not trivial, at least quite do-able and lots of fun. The Fermi-Pasta-Ulam problem was what set Kruskal on the road to working out this concept.

On the other hand, all through the sixties it was considered that ordinary differential equations with two or three independent variables were so simple to solve that there wasn't any reason to be interested in them. And then during the seventies people recognized that a very simple system of three ordinary differential equations with a little nonlinearity added can become chaotic and can no longer be solved in a predicted manner on the most powerful digital computers.

At that point everything was turned topsy-turvy. The problems that were considered so simple to solve that they weren't even worth looking at turned out to be impossible to solve with the best numerical machines we have because they led to chaos. On the other hand, the problems that were considered impossible to solve turned out to be easy because their solutions could be constructed from solitons. The landscape is totally different now than it was fifteen years ago when you consider the kinds of problems we are interested in working on, what sort of tools we use, and what we expect to happen.

SCIENCE: I always thought of the soliton as a curious solution to a particular equation, not something that you would necessarily expect to find in nature. But you are suggesting that it may be quite common.

SCOTT: Yes, the soliton, or more generally the self-trapping of energy or charge into solitary waves, is really a paradigm. It has been around, you might say as a latent paradigm, since 1834, but it began to be recognized only around 1970 as something a typical scientist might expect to find in a typical problem. And the recognition is growing. Every month or so some scientist discovers yet another equation that has soliton solutions. On the other hand, people whose training is entirely with linear equations think of what is going on as a kind of Fourier reconstruction of sine waves without trapping.

SCIENCE: This soliton paradigm seems to evoke an emotional response, with some people *believing* in solitons and some not.

SCOTT: My feeling is that that's the way these changes work. Think of somebody who has spent his entire life developing a certain point of view and has a substantial position in the power structure of science—and then some young people come along with new ideas that he doesn't understand at all, ideas that, if true, would be harmful to his historical position and present power. He's in a position not to believe.

SCIENCE: What do you think of in terms of training people to work on nonlinear problems? What kind of training would be best for them?

SCOTT: Well, for a long time there has been a need in physics graduate programs for a good course in nonlinear science. Many graduate schools are getting something like that now. Probably it's not necessary to have too much more than an introduction to modern concepts. The most important thing is the thesis: many people get pointed in the direction of nonlinear science by the choice of their thesis problem.

SCIENCE: There seem to be some problems in communication. Some people, especially in the biological and biomedical disciplines, seem to be throwing around the word *soliton* without knowing precisely what it means. (We hope to clarify its meaning in this issue.) Then there are the language barriers that exist between disciplines.

SCOTT: Yes, there are language barriers and also concept barriers. But exchange between disciplines has been fruitful in the past: x-ray crystallography is one example, another is nuclear magnetic resonance, and still another is neutron scattering. Solitons may be another area where the interaction between biochemists and physicists could be important. If Davydov's ideas are essentially correct, then they will be important for understanding the way proteins function; if not, then the work will fizzle out in a year or two. But I won't believe that

So, many problems considered hopeless turned out in many cases to be, if not trivial, at least quite do-able and lots of fun.

nonlinear dynamics will not have a role to play in understanding how biochemicals function—that much seems certain.

SCIENCE: It seems the start of an exciting venture. Does the work on solitons dovetail with the programmatic work already going on in Life Sciences?

SCOTT: Yes, indeed. For example, Mark Bitensky has studied for a number of years the succession of dynamic events that take place after a photon of light comes into the eye—all of the activity in the protein rhodopsin before a pulse is perceived as a signal on the optic nerve. At the present time, as I understand it, much is known about *what* is happening with the various components of rhodopsin. Just *how* they're happening and *why* is not very well understood. I think Mark is interested in our work because it may provide a basis for answering some of those questions. In fact, the life sciences are just dripping with such questions.

SCIENCE: How do you find the intellectual climate of Los Alamos? How do you feel about having this as a place to work?

SCOTT: Well, all large organizations have some problems just because they are large and complicated. But it seems to me that there are interesting differences between the social dynamics of a university and those of a national laboratory. It's been particularly impressed upon my mind since I came here from the University of Wisconsin, which is one of our larger educational institutions.

In the university the essential group, the cellular unit, is the department. The administration of the university can't penetrate the department very much. To eliminate a department would take an act of the state legislature, so a department has the possibility of challenging or even ignoring directives of the administration. Thus there is a difficulty getting departments to interact with one another and in organizing projects between departments. Which dean will control the project? Where is the money actually going to go? What is the advantage of one department giving aid to another department?

Here at the Laboratory we have a strong administrative component on the one hand and a large scientific staff on the other, but the groups here don't have the same legal status and coherence that characterize departments in a university. Here it is relatively easy to get people in widely different disciplines together on the same project; it can be organized in a matter of weeks. It's a great advantage to be able to get biochemists and laser physicists and computer scientists and mathematicians all working on the same project.

SCIENCE: How about the financing of the Center? Since you are funded by the Laboratory, how much autonomy do you have? Is the money yours to use, without strings attached?

SCOTT: To use *responsibly*. And it's interesting that to deal with the charge of the Center for Nonlinear Studies to use this money flexibly and imaginatively precludes doing those things that lead to bureaucratic stability. If we were opting for that, we would try to convert the budget into permanent positions, not use it for postdocs, visitors, workshops, and staff support throughout the Laboratory.

SCIENCE: As the Center's many projects become established, it seems that the postdocs who come to work on them might want to stay here.

SCOTT: Well, more than half the postdocs who come to Los Alamos do stay. If the Center draws good postdocs to the Laboratory and they go into other groups and divisions, that will certainly benefit the Laboratory.

SCIENCE: And they will presumably continue their interest in nonlinear phenomena.

SCOTT: Yes, and those who go into the universities will also seed interactions between Los Alamos and the universities.

SCIENCE: Have the divisions offered to donate their own money toward these projects? Say someone in a division wants to work on a project that relates to the Center, will the division support that person rather than your having to?

SCOTT: Sure. For example, in connection with the biology project, there's a significant component going on under the direction of Irving Bigio in the Chemistry Division. It gets some support directly from Chemistry and some support from an Institutional Supporting Research and Development request that is apart from the budget of the Center. And we hope, of course, that in the future some support will come from outside.

SCIENCE: How is communication among the people at the Center? Do you serve as an intellectual guide to the work that is going on?

SCOTT: That's hard to say because there are so many things going on, so many different kinds of activities. I'm in a position to act as a guide in areas that I know something about. But, for example, I don't know very much about chaos business. I try to keep up with what's going on, but there isn't any way I could be described as an intellectual leader in that area.

SCIENCE: But you do fulfill that important function of deciding where the money will go, what the aims of the Center are, and how those aims will be implemented—for example, the emphasis on, belief in, young people.

SCOTT: Well, it's important to have taste. I feel good about this kind of scientific development, and to be in a position to help manage it in a responsible way is something very satisfying to be involved with.

SCIENCE: We've talked about *belief,* that some people believe in solitons and some don't. There must be a certain sense of mission at the Nonlinear Center.

SCOTT: I would prefer the phrase *understand solitons*, but you are right about the sense of mission, and that is extremely important. The Center for Nonlinear Studies is prospering because many people here at the Laboratory and throughout the country recognize it as a good idea and have been willing to work to help it get started. That's absolutely vital: without this commitment from many people the Center wouldn't fly at all. ■

☆ U.S. GOVERNMENT PRINTING OFFICE: 1984 — 776-101/1025

LA-UR-84-430