NEUTRON SCATTERING

Artist's conception of neutrons flying through matter. Low-energy neutrons (red) scatter to new trajectories by interacting with the atomic nuclei (blue). Since the nuclei are small, the neutrons see mostly empty space and scattering are rare. Cynthia L. Boone created the computer art using Alias software. The 14,000 nuclei were generated by 50 translations of an original lattice plane.

Part 1. Introduction to Neutron Scattering

by Roger Pynn

Neutrons are a little-known but remarkably powerful probe of condensed matter on length scales ranging from interatomic distances to a few thousand angstroms. This article explores the physics of various neutron-scattering processes, introduces the experimental techniques and instruments that make neutron scattering so versatile, and discusses the single equation that unifies the interpretation of neutron-scattering data.

Mathematical Foundations of Neutron Scattering

The director of the Los Alamos neutron-scattering facility discusses the history of the field, its successes around the world, its present problems in the United States. and the plans for opening it to a wide spectrum of users from academia and industry.

by Dianne K. Hyer and Roger Pynn

The accelerator neutron source at the Los Alamos neutron-scattering center is proving at least as effective as traditional reactor sources, This article traces neutrons from their "birth" in the spallation target through beam-tailoring devices and scattering samples to their "death" in neutron detectors. Electronic measurements of times of flight and scattering angles are also discussed, as well as the mechanics of the user program at LANSCE.

Part II. Neutron Scattering Research at LANSCE:

by Jill Trewhella

Small-angle neutron-scattering experiments provide evidence that calmodulin, a protein that mediates calcium regulation of biological processes, is flexible in solution. Such experiments also clearly show that calmodulin closes around one of its target enzymes as it stimulates that enzyme's catalytic function.

Ferrofluids-A New Alignment Technique

What is neutron scattering? Why is it becoming more and more useful in both basic and applied science? And why are materials scientists, in particular, very worried about its future in the United States? At first glance the technique appears forbidding. You shine a large number of neutrons on a sample, but—because neutrons have no electric charge-only a very small fraction are scattered. Therefore generating measurable signals requires a great deal of ingenuity. Also, the neutrons are expensive to produce and are available at only a small number of specialized facilities. It seems like the last technique to choose-except that its results are so valuable. For many questions in condensed-matter physics, in biology, in chemistry, and in the study of modem synthetic materials, neutron scattering is the only technique that provides answers.

Unlike x-rays, neutrons "see" light and heavy elements equally well, interact with the magnetic moments of electrons, and detect atomic vibrations as well as average atomic positions. These properties make neutrons a unique and extremely versatile probe of material structure and dynamics. A look at the table of contents will convince you—measurements of the structure of high-temperature superconductors, of the Bose condensate in liquid helium, of the change in shape of the protein calmodulin as it mediates involuntary muscle contraction, of the catalytic effect of metal atoms on the dissociation of molecular hydrogen, of the fractal structures of precursors to today's materials, of the layered structures of polymer films, of residual strains in metals and ceramics—and these diverse topics are a sampling of the work at only one facility. High-intensity neutron beams, high-resolution spectrometers, and computerized data acquisition and analysis systems are opening a wide range of phenomena to more precise study, especially in materials technology and other fields tied to economic growth.

Although the promise of neutron scattering is growing, its future in the United States is uncertain. At present, all but two neutron-scattering centers are located at aging nuclear reactors whose budgets are escalating every year to pay, not for research and upgrades, but for enhanced safety. Moreover, access to those facilities has not been easy in the past. Even now, few university scientists use the technique— many fewer than in Europe. But a change is in the works. In the early eighties a user program began at the Intense Pulsed Neutron Source at Argonne National Laboratory, and in 1987 the Los Alamos Neutron Scattering Center (LANSCE) officially opened its doors to users from academia and industry. (Both facilities house accelerator sources, which produce pulsed neutron beams by a nuclear reaction called spallation.) A cold-neutron research center is being built at the National Institute of Standards and Technology, and a spectrometer upgrade at the Brookhaven High-Flux Reactor is planned. The response to these initiatives has been gratifying. For example, more than 140 outside users mounted experiments at LANSCE last year—and this year's participation is even greater. A technique that was practiced by a small community of professionals is now accessible to "amateurs" from industry and academia.

Experience at the Los Alamos facility is proving that data obtained at pulsed neutron sources can compete with data obtained at reactors. Five years ago when several new spallation sources were just coming on line, conventional wisdom said

that small-angle neutron scattering, which requires intense beams to measure the sizes and shapes of biological macromolecules and colloidal particles, could be done well only at reactors. Today small-angle experiments work as well at LANSCE as at the best reactor sources in the world, except for the very smallest angles. Powder-diffraction experiments here are also highly effective; for example, polycrystalline samples weighing less than 50 milligrams have yielded excellent data. And Los Alamos researchers are leading the way in using spallation sources for reflectometry, one of the newest applications of neutron scattering. These successes reflect the high neutron flux available here as well as the efficient utilization of that flux through innovative spectrometer design. LANSCE and other pulsed sources are demonstrating that they can more than fulfill their original promise.

With all its successes the Los Alamos source is not without problems. Beam availability during 1990 has so far been less than last year because of budget cuts, investment in environmental and safety precautions, and some accelerator problems. Also, though LANSCE generates a higher peak flux than any other spallation source, it has yet to achieve the design value. Nevertheless this year's users report they are getting better data in less time than they could elsewhere.

What of the future? Since neutron scattering is a signal-limited technique, progress depends on building higher-intensity neutron sources. The U.S. neutron-scattering community plans to build a state-of-the-art research reactor, the Advanced Neutron Source, which will deliver five times greater neutron fluxes than today's most powerful research reactors. The new reactor, still in the design stage, should begin operation around the turn of the century. Construction will cost about one billion dollars, and increasingly stringent safety regulations will make it very costly to run. The decision to build the advanced reactor was made several years ago, before spallation sources had proven their worth. We asked Roger Pynn, the director of LANSCE, whether in light of recent progress he thought the next-generation facility ought to be a spallation source. His answer: "If societal concerns should make it impractical to build and operate a research reactor, then an advanced spallation source could meet the needs of the community without raising as many safety and environmental issues, But, for a balanced program, the country needs sources of both types." Whatever happens in the long run, it is clear from the presentations in this issue that neutron scattering is a fertile and expanding field and that LANSCE is providing inspiration for the future in the areas of technical innovation, opportunities for young people, outreach to the larger scientific community, and tangible scientific results.

The editors thank all the authors and the numerous other scientists who contributed their thought and energy to the contents of this volume. It was a pleasure for us to learn about the technically difficult but thoroughly absorbing world of neutron scattering.

"Neutron Man" personifies the neutron's dual nature, exhibiting wave and particle properties. Here he enters a crystal lattice as a plane wave (blue), interacts with the crystal lattice (green), and becomes, through interference effects, an outgoing plane wave (red) with a direction dictated by Bragg's law. His particle properties allow him to be absorbed by a helium atom in a neutron detector, and his time of flight is measured.
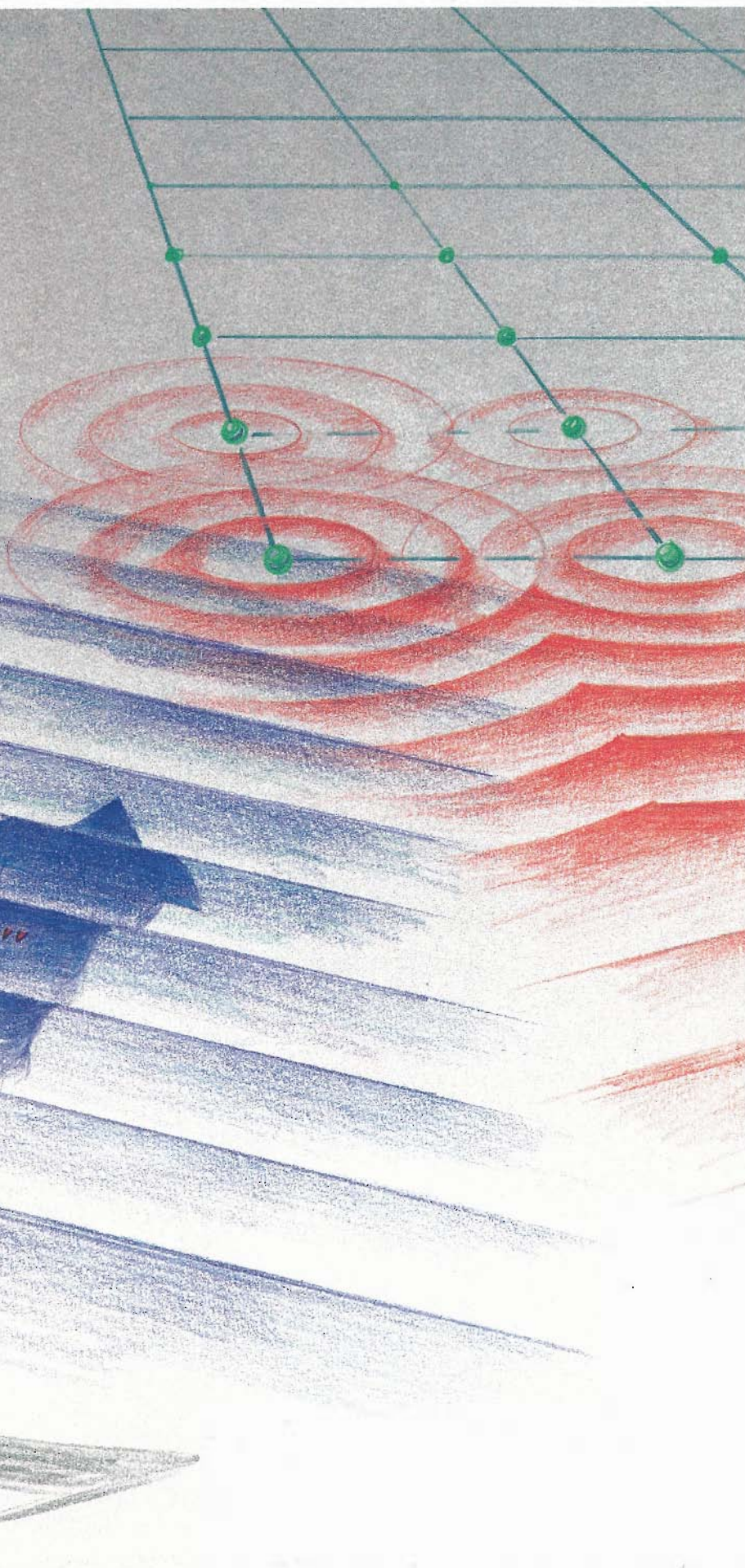
neutron scattering
A PRIMER

by Roger Pynn

How can we determine the relative positions and motions of atoms in a bulk sample of solid or liquid? Somehow we need to see inside the material with a suitable magnifying glass. But seeing with light in an everyday sense will not suffice. First, we can only see inside the few materials that are transparent, and second, there is no microscope that will allow us to see individual atoms. These are not merely technical hurdles, like those of sending a man to the moon, but intrinisic limitations. We cannot make an opaque body transparent nor can we see detail on a scale finer than the wavelength of the radiation we are using to observe it. For observations with visible light this limits us to objects separated by about a micrometer (10^{-6} meter), which is more than a thousand times longer than the typical interatomic distance in a solid (about 10^{-10} meter or so).

X rays have wavelengths much shorter than those of visible light, so we might try using them to find atomic positions. For many crystalline materials this technique works quite well. The x rays are diffracted by the material, and one can work out the relative atomic positions from the pattern of spots the diffracted rays make on a photographic plate. However, not all atoms are equally "visible" to x rays:
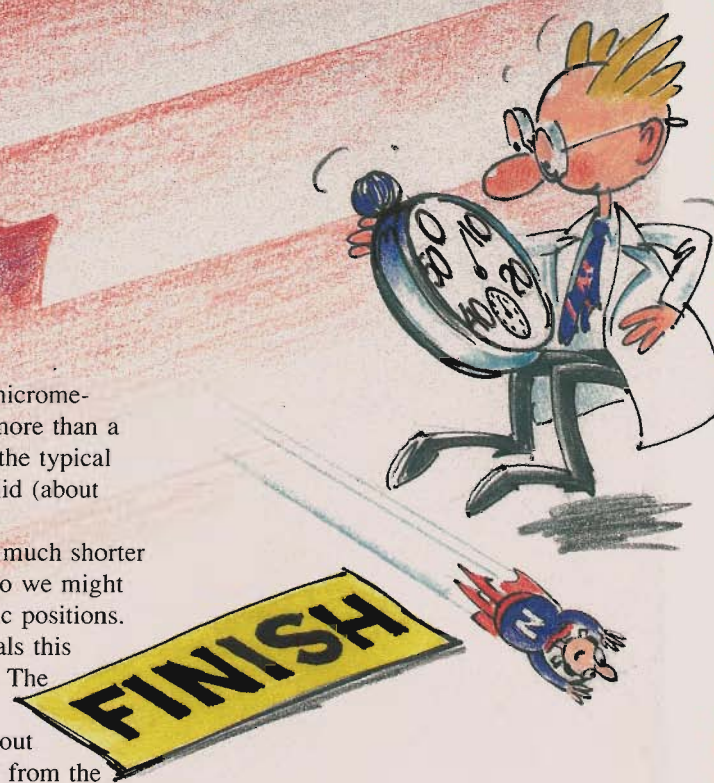
the light atoms in the soft tissue of our jowls do not stop x rays as well as the heavy mercury atoms in the dental amalgam used to fill teeth. Although this phenomenon is useful to the dental profession, it is often an embarrassment for scientists measuring atomic positions.

X rays are scattered by the electrons surrounding the nucleus of an atom. As a result, heavy atoms with many electrons (such as mercury) scatter x rays more efficiently than light atoms (such as oxygen or, worse, hydrogen). Thus, x rays pass right through light materials without being greatly attenuated or deflected. It is for this reason that the structure of the much-heralded high-temperature superconductors was not determined by x-ray diffraction— in spite of the fact that most university physics departments worldwide have an x-ray machine. One of the first high-temperature superconductors discovered contained yttrium and copper, both of which are heavy and scatter a relatively large percentage of the x rays incident on a sample. Unfortunately, the superconductors also contained oxygen, whose feeble scattering of x rays is swamped by that of its heavy neighbors. It was impossible to determine the positions of the oxygen atoms using x-ray diffraction because the x rays passed through the superconductor almost without noticing the oxygen.

We might try to find atomic positions by "seeing" with electron beams. After all, quantum mechanics tells us that particles have wave properties, and the wavelength of electrons can easily be matched to interatomic distances by changing the electron energy. However, as anyone who has ever rubbed a balloon on the family cat knows, the interaction between electrical charges is strong. Not surprisingly then, a charged particle, such as an electron or a positron, does not travel far through solids or liquids before it is attracted or

NEUTRON, ELECTRON, AND X-RAY PENETRATION DEPTHS

Fig. 1. The plot shows how deeply a beam of electrons, x rays, or thermal neutrons penetrates a particular element in its solid or liquid form before the beam's intensity has be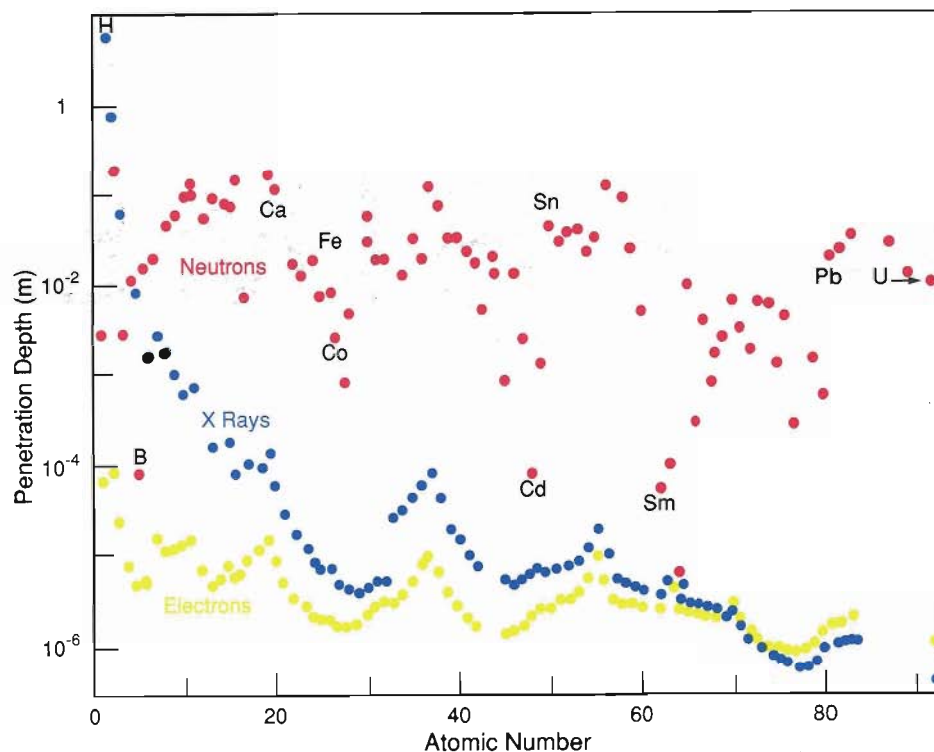en reduced by a factor of $\frac{1}{e}$, that is, to about 37 percent of its original intensity. The neutron data are for neutrons having a wavelength of 1.4 angstroms (1.4×10^{-10} meter).

repelled by the electrons already in the matter. This makes electrons unsuitable for looking inside bulk materials: they suffer from the same opacity problem as light, and specially prepared, thin samples are required for electron microscopy.

Neutron Scattering

What about neutrons? They have no charge, and their electric dipole moment is either zero or too small to be measured by the most sensitive of modern techniques. For these reasons, neutrons can penetrate matter far better than charged particles. Furthermore, neutrons interact with atoms via *nuclear* rather

than electrical forces, and nuclear forces are very short range—of the order of a few fermis (1 fermi = 10^{-15} meter). Thus, as far as the neutron is concerned, solid matter is not very dense because the size of a scattering center (nucleus) is typically 100,000 times smaller than the distance *between* such centers. As a consequence, neutrons can travel large distances through most materials without being scattered or absorbed (see the opening illustration to "Putting Neutrons in Perspective"). The attenuation, or decrease in intensity, of a beam of low-energy neutrons by aluminum, for example, is about 1 percent per millimeter compared with 99 percent or more per millimeter for x rays. Figure 1 illus-
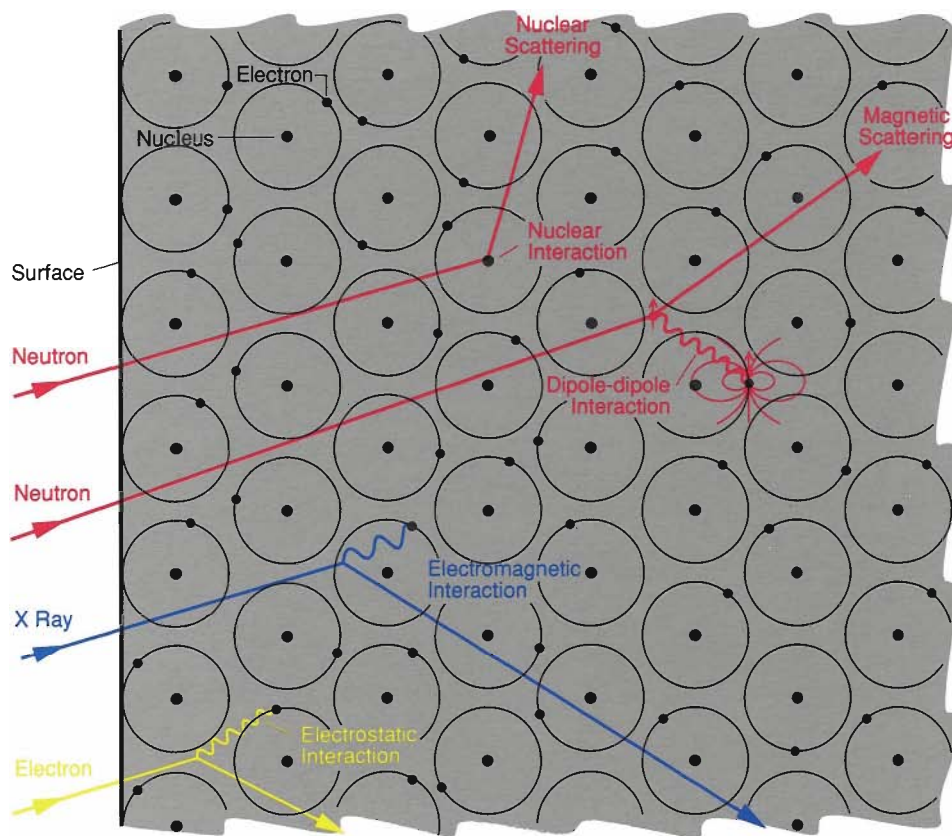
SCATTERING INTERACTIONS

Fig. 2. Beams of neutrons, x rays, and electrons interact with material by different mechanisms. X rays (blue) and electron beams (yellow) both interact with electrons in the material; with x rays the interaction is electromagnetic, whereas with an electron beam it is electrostatic. Both of these interactions are strong, and neither type of beam penetrates matter very deeply. Neutrons (red) interact with atomic nuclei via the very short-range strong nuclear force and thus penetrate matter much more deeply than x rays or electrons. If there are unpaired electrons in the material, neutrons may also interact by a second mechanism: a dipole-dipole interaction between the magnetic moment of the neutron and the magnetic moment of the unpaired electron.

tensities. X-ray instruments at synchrotron-radiation facilities can provide fluxes of 10^{18} photons per second per square millimeter compared with 10^4 neutrons per second per square millimeter in the same energy bandwidth for powerful neutron-scattering instruments.

The combination of a weak interaction and low fluxes makes neutron scattering a *signal-limited technique*, which is practiced only because it provides information on the structure of materials that cannot be obtained by other means. This constraint means that no generic instrument can be designed to examine all aspects of neutron scattering. Instead, a veritable zoo of instruments has arisen with each species specializing in a particular aspect of the scattering phenomenon.

In spite of its unique advantages, neutron scattering is only one of a battery of techniques for probing the structures of materials. All of the techniques, such as x-ray scattering and electron microscopy, are needed if scientists are to understand the full range of structural properties of matter. In most cases, the different methods used to probe material structure give complementary information because the nature of the interaction between the radiation and the sample are different. For example, neutrons interact with nuclei, whereas x rays and electrons "see" only the electrons in matter (Fig. 2). To a certain extent the method of choice depends on the length scale of the structure to be investigated (Fig. 3). When two techniques address the same scale, additional information, such as the size and chemical composition of the sample, is required to choose the optimal technique.

Scattering by a Single Fixed Nucleus

The scattering of neutrons by nuclei is a quantum-mechanical process. Formally, the process has to be described in

trates this point for other atoms and for electrons as well as x rays and neutrons.

Like so many things in life, the neutron's penetrating power is a two-edged sword. On the plus side, the neutron can penetrate deep within a sample even if it first has to pass through a container (necessary, for example, if the sample is a fluid or has to be kept at low temperatures or high pressures). The corollary is that neutrons are only weakly

scattered once they do penetrate. Also, detection of a subatomic particle involves the observation of that particle's interaction with some other particle, so neutron detection requires a certain ingenuity (in practice, detectors make use of one of the few atoms, such as boron, helium-3, or lithium, that absorb neutrons strongly to produce ionizing radiation). To make matters worse, available neutron beams inherently have low in-

Crystallography   Microstructure   Structure

Proteins   Viruses   Bacteria

Micelles

Polymers

Atomic Structures   Grain Structures

Porous Media   Precipitates

Neutron Diffraction

Neutron Small-Angle Scattering

Optical Microscopy

X-Ray Diffraction

X-Ray Small-Angle Scattering

Electron Diffraction

Transmission Electron Microscopy

Scanning Transmission Electron Microscopy

Field-Ion Microscopy

10^{-11} 10^{-9} 10^{-7} 10^{-5} 10^{-3}

Size (meters)

STRUCTURE PROBES

◀ Fig. 3. A variety of techniques can be used to probe structure, but one of the main determining factors in the choice of a technique is the length scale of the structure being examined. Techniques range from neutron diffraction, which can be used to study atomic structure with length scales of 10^{-11} to 10^{-9} meter, to optical microscopy, which can be used to study bacteria and crystalline grain structures at much greater length scales.

terms of the wave functions of the neutron and the nucleus. The wave function of the neutron, as its name suggests, has the form of a wave—that is, a function that oscillates sinusoidally in space and time. The square of the amplitude of this wave at any point gives the probability that the neutron will be found at that point. It does not matter whether we talk about the wave that represents the neutron or the probability that a particle called the neutron is at a given location. Both descriptions will give rise to the same mathematics and are, therefore, equivalent. Sometimes it is convenient to refer to the neutron as a wave because the picture thus conjured is easier to understand. At other times it is more useful to think of the neutron as a particle. We can switch from one description to the other at will, and if we do the mathematics correctly, we will always get the same answer.

The neutrons used for scattering experiments usually have energies similar to those of atoms in a gas such as air. Not surprisingly, the velocities at which they move are also comparable with those of gas molecules—a few kilometers per second. Quantum mechanics tells us that the wavelength of the neutron wave is inversely proportional to the magnitude of the neutron velocity $v = |\mathbf{v}|$ (throughout the text we will use a bold variable to represent a vector quantity and a nonbold ver-

sion of the same variable to represent the corresponding magnitude). For the neutrons used in scattering experiments, the wavelength, λ, turns out to be a few angstroms (1 angstrom = 10^{-10} meter). It is often useful to work in terms of the so-called neutron wave vector, $\mathbf{k}$, which is a vector of magnitude $k = 2\pi/\lambda$ that points along the neutron's trajectory. The vectors $\mathbf{k}$ and $\mathbf{v}$ are collinear and related by the equation

$$\frac{h\mathbf{k}}{2\pi} = m\mathbf{v}, \qquad (1)$$

where h is Planck's constant, m is the mass of the neutron (1.67495 × 10^{-27} kilogram), and $m\mathbf{v}$ is the momentum of the neutron.

The scattering of a neutron by a single nucleus can be described in terms of a *cross section* σ, measured in barns (1 barn = 10^{-28} square meter), that is equivalent to the effective area presented by the nucleus to the passing neutron. If the neutron hits this area, it is scattered isotropically, that is, with equal probability in any direction. Why isotropically? The range of the nuclear potential is tiny compared to the wavelength of the neutron, and so the nucleus is effectively a *point scatterer*. (X rays, on the other hand, do not scatter isotropically because the electron clouds around the atom scattering the x rays are comparable in size to the wavelength of the x rays.)

Suppose that at an instant in time we represent neutrons incident on a fixed nucleus by a wave function $e^{i\mathbf{k}\cdot\mathbf{r}}$, which is a plane wave of unit amplitude expressed in terms of the position vector $\mathbf{r}$. Note that the square modulus of this wave function is unity, which means the neutron has the same probability of being found anywhere in space but has definite momentum $m\mathbf{v} = h\mathbf{k}/2\pi$. The nodes of the wave—that is, the points at which the phase $\mathbf{k} \cdot \mathbf{r}$ is equal to $n\pi$, where n is an integer—are the straight
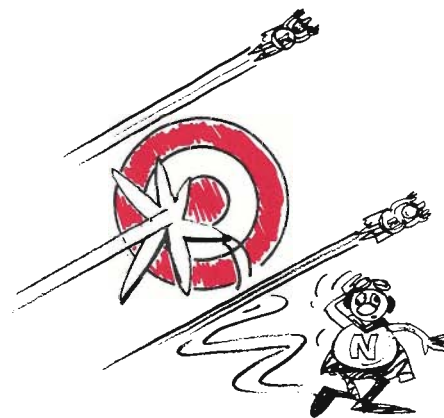
cross section

The effective area presented by a nucleus to an incident neutron. One unit for cross section is the barn, as in "can't hit the side of a barn!"
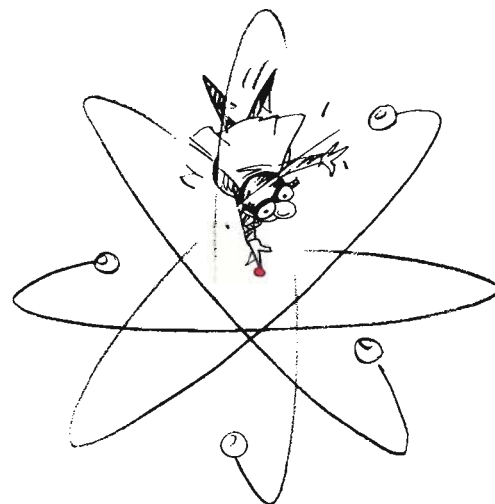
point scatterer

An object that scatters incident radiation isotropically by virtue of being very small compared with the wavelength of the radiation.

NEUTRON SCATTERING FROM A FIXED POINT

Fig. 4. A neutron beam incident on a single scattering center and traveling in the x direction can be represented as a plane wave e^{ikx} with unit amplitude. Because the neutron sees the scattering center (a nucleus) as a point, the scattering will be isotropic. As a result, the scattered neutron beam spreads out in spherical wavefronts (here drawn as circles) of amplitude b/r. The $1/r$ part of this amplitude factor, when squared to get intensity, accounts for the $1/r^2$ decrease in intensity with distance that occurs as the scattered wavefront grows in size. Because we have here taken the scattering center to be rigidly fixed, the interaction is elastic, there is no exchange of energy, and the incident and scattered wave vectors both have magnitude k. (To be rigorous, we should have included the time dependence $e^{i\omega t}$. But since the scattering is elastic, this factor is the same for the incident and scattered waves and cancels out of relative expressions, such as the one for the cross section.)

scattering length

A measure of the strength of the neutron-nucleus interaction, denoted by b and related to the cross section σ by $\sigma = 4\pi b^2$.

isotopic labeling

A technique that takes advantage of the considerable variation in scattering cross section among isotopes. By substituting one isotope for another (of either the same or a different element), the scattering from those constituents containing the substitute may be varied to reveal their positions relative to other constituents.

y

k

Scattered Circular Wave $\frac{-b}{r} e^{ikr}$

r

k

x

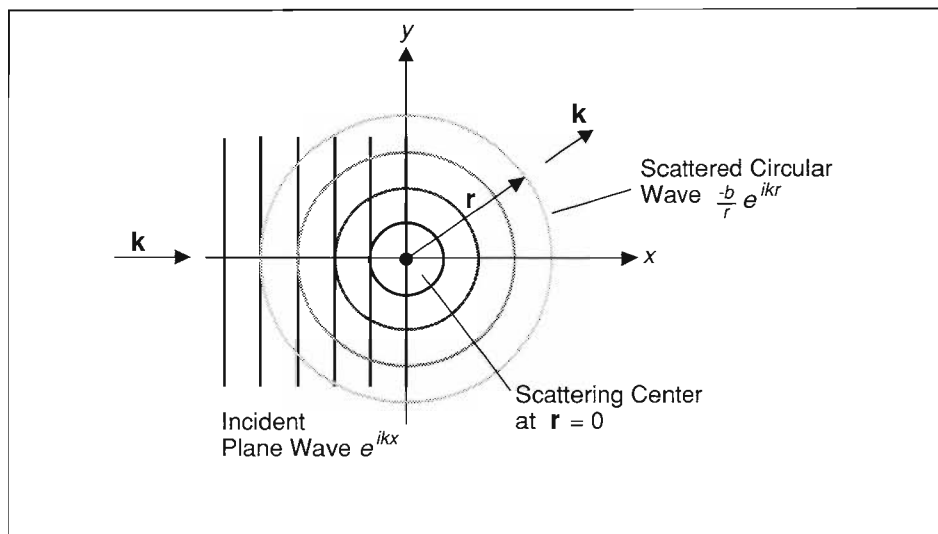Scattering Center at $\mathbf{r} = 0$

Incident Plane Wave e^{ikx}

wavefronts shown in Fig. 4 (for a wave traveling in the x direction). In light of our earlier discussion, we ought to choose the amplitude of the neutron wave function (the constant multiplying the exponential) so that the amplitude squared gives a probability of finding a neutron at a position $\mathbf{r}$ that is consistent with the number of neutrons in the beam we are using. However, since we shall be interested only in the ratio of the amplitudes of the incident and scattered neutron waves, we can set the amplitude of the incident wave to unity for the moment.

What is the amplitude of the neutron wave scattered by the nucleus? That depends on the strength of the interaction between the neutron and the nucleus. Because the scattered neutron wave is isotropic, its wave function can be written as $(-b/r)e^{ikr}$ if the scattering nucleus is at the origin of our coordinate system. The spherical wavefronts of the scattered neutron are represented by the circles spreading out from the nucleus in Fig. 4. The factor $(1/r)$ in the wave function of the scattered neutron takes care of the inverse square law that applies to all wave motions: the intensity of the neutron beam, given by the square of the amplitude of the

wave function, decreases as the inverse square of the distance from the source. In this case, the source is the scattering nucleus. The constant b, referred to as the *scattering length* of the nucleus, measures the strength of the interaction between the neutron and the scattering nucleus. The minus sign in the wave function means that b is a positive number for a repulsive interaction between neutron and nucleus.

For the type of collision being imagined here, the energy of the neutron is too small to change the internal energy of the scattering nucleus, and because we imagine the nucleus to be fixed, the neutron cannot impart kinetic energy. Thus, the scattering occurs without any change of the neutron's energy and is said to be *elastic*. Because the neutron energy is unchanged by a nuclear collision, the magnitude of its velocity and thus of its wave vector is also unchanged, and the same k appears in the wave function of the incident and the scattered neutrons.

What is the relationship between scattering length, b, and the cross section, σ, both of which are a measure of strength of the scattering interaction? The cross section, σ, an area, is related to b, a length, by the simple relation

6
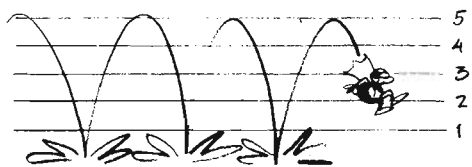
BOING! BOING! BOING!

elastic scattering

Scattering with no change in the energy of the incident neutron; or, in terms of the wave vector of the neutron, scattering in which the direction of the vector changes but not its magnitude.

$\sigma = 4\pi b^2$—as if the scattering length were half the radius of the nucleus as seen by the neutron.

For a few nuclei the scattering length, b, varies with the energy of the incident neutrons because compound nuclei with energies close to those of excited nuclear states are formed during the scattering process. This *resonance* phenomenon gives rise to complex values of b: the real part corresponds to scattering of the neutrons, whereas the imaginary part corresponds to absorption of the neutron by a nucleus. Usually, such resonant effects occur at neutron energies greater than those used to probe the structure of matter. In the majority of cases of interest to scientists doing neutron scattering, b is a real and energy-independent quantity. However, b has to be determined experimentally for each nuclear isotope because, unlike the equivalent quantity for x rays, the scattering length for neutrons cannot be calculated reliably in terms of fundamental constants.

Also unlike x rays, neutrons interact with atoms of an element in a manner that does not seem correlated with the atomic number of the element (as is evident in Fig. 1). In fact, the neutron's interaction with a nucleus of an atom varies from one isotope to another. For example, hydrogen and deuterium, both of which interact weakly with x rays, have neutron scattering lengths that are relatively large and quite different. The differences in scattering lengths from one isotope to another can be used in various *isotopic-labeling* techniques to increase the amount of information available from a particular neutron-scattering experiment. We shall discuss isotopic labeling in more detail in the section on small-angle scattering.

Scattering of Neutrons by Matter

To work out how neutrons are scattered by matter, we need to add up the scattering from each of the individual nuclei. This is a lengthy and not particularly instructive quantum-mechanical calculation. Fortunately, the details of the calculation are not very important. The result is, however, both simple and appealing.

When neutrons are scattered by matter, the process can alter both the momentum and the energy of the neutrons and the matter. The scattering is not necessarily elastic as it is for a single, rigidly fixed nucleus because atoms in matter are free to move to some extent. They can therefore recoil during a collision with a neutron, or if they are moving when the neutron arrives, they can impart or absorb energy just as a baseball bat does.

As is usual in a collision, the total momentum and energy are conserved: when a neutron is scattered by matter, the energy lost by the neutron, ϵ, is gained by the sample. From Eq. 1 it is easy to see that the amount of momentum given up by the neutron during its collision, the *momentum transfer*, is $\frac{h}{2\pi}\mathbf{Q} = \frac{h}{2\pi}(\mathbf{k} - \mathbf{k}')$, where $\mathbf{k}$ is the wave vector of the incident neutrons and $\mathbf{k}'$ is that of the scattered neutrons. The quantity $\mathbf{Q} = \mathbf{k} - \mathbf{k}'$ is known as the *scattering vector*, and the vector relationship between $\mathbf{Q}$, $\mathbf{k}$, and $\mathbf{k}'$ can be

inelastic scattering

Scattering in which exchange of energy and momentum between the incident neutron and the sample causes both the direction and the magnitude of the neutron's wave vector to change.

SCATTERING TRIANGLES

Fig. 5. Scattering triangles are depicted here for both (a) an elastic scattering event in which the neutron is deflected but the neutron does not lose or gain energy (so that $k' = k$) and (b) inelastic scattering events in which the neutron either loses energy ($k' < k$) or gains energy ($k' > k$) during the interaction. In both elastic and inelastic scattering events, the neutron is scattered through the angle 2θ, and the scattering vector is given by the vector relationship $\mathbf{Q} = \mathbf{k} - \mathbf{k'}$. For elastic scattering, a little trigonometry shows (lower triangle in (a)) that $Q = 4\pi \sin\theta/\lambda$.

(a) Elastic Scattering ($k' = k$)

$$\sin\theta = \frac{Q/2}{k}$$

$$Q = 2k\sin\theta = \frac{4\pi\sin\theta}{\lambda}$$

(b) Inelastic Scattering ($k' \neq k$)

Neutron Loses Energy
($k' < k$)

Neutron Gains Energy
($k' > k$)
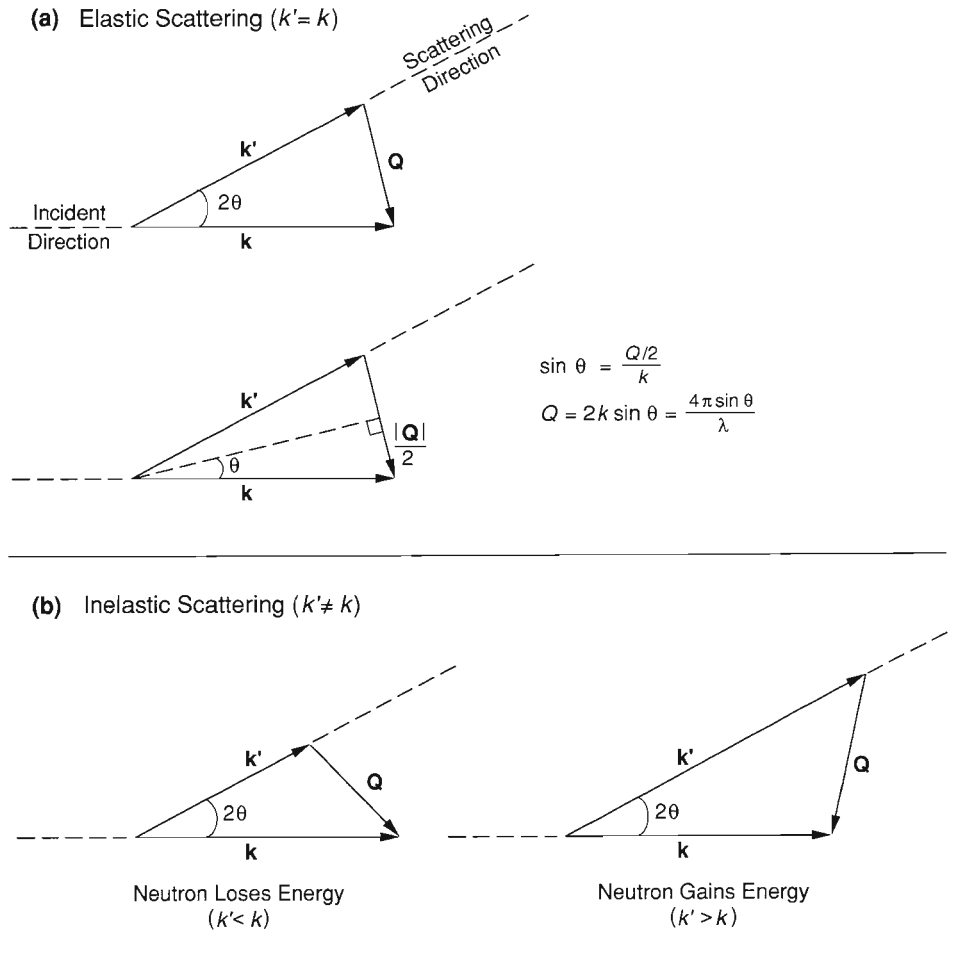
neutron-scattering law

The intensity of scattered neutrons as a function of the momentum and energy transferred to the sample during the scattering. The scattering law is written as $I(Q,\epsilon)$, where $\frac{h}{2\pi}Q$ is the momentum transfer, and ϵ is the energy transfer (see Eq. 3 in "The Mathematical Foundations of Neutron Scattering").

displayed pictorially in the so-called *scattering triangle* (Fig. 5). This triangle also emphasizes that the magnitude and direction of $\mathbf{Q}$ is determined by the the magnitudes of the wave vectors for the incident and scattered neutrons and the angle 2θ through which a neutron is deflected during the scattering process. Generally, 2θ is referred to as the *scattering angle*. For elastic scattering (Fig. 5a) $k = k'$, so $\epsilon = 0$, and a little trigonometry applied to the scattering triangle shows that $Q = 4\pi \sin\theta/\lambda$.

In *all* neutron-scattering experiments, scientists measure the intensity of neu-

trons scattered by matter (per incident neutron) as a function of the variables $\mathbf{Q}$ and ϵ. This scattered intensity, denoted $I(\mathbf{Q}, \epsilon)$, is often referred to colloquially as the *neutron-scattering law* for the sample.

In a complete and elegant analysis, Van Hove showed in 1954 that the scattering law can be written exactly in terms of time-dependent correlations between the positions of *pairs* of atoms in the sample (see "The Mathematical Foundations of Neutron Scattering" for a more detailed discussion). Van Hove's result implies that $I(\mathbf{Q}, \epsilon)$ is simply pro-

portional to the Fourier transform of a function that gives the probability of finding two atoms a certain distance apart. It is the simplicity of this result that is responsible for the power of neutron scattering. If nature had been unkind and included correlations between triplets or quadruplets of atoms in the expression for the scattering law, neutron scattering could never have been used to probe directly the structure of materials.

Of course, we have not yet explained *how* one may measure the intensity of scattered neutrons as a function of Q and ϵ, but *if* we can do that, Van Hove's result provides a way of relating the intensity of the scattered neutrons to the relative positions and the relative motions of atoms in matter. In fact, Van Hove's formalism can be manipulated (see "The Mathematical Foundations of Neutron Scattering") to reveal scattering effects of two types. The first is *coherent scattering* in which the neutron wave interacts with the whole sample as a unit so that the scattered waves from *different* nuclei interfere with each other. This type of scattering depends on the relative distances between the constituent atoms and thus gives information about the structure of materials. *Elastic* coherent scattering tells us about the equilibrium structure, whereas inelastic coherent scattering (with $\epsilon \neq 0$) provides information about the collective motions of the atoms, such as those that produce vibrational waves in a crystalline lattice. In the second type of scattering, *incoherent scattering*, the neutron wave interacts independently with each nucleus in the sample so that the scattered waves from different nuclei *don't* interfere. Rather the intensities from each nucleus just add up. Incoherent scattering may, for example, be due to the interaction of a neutron wave with the *same* atom but at different positions and different times, thus providing information about atomic diffusion.

Diffraction, or Bragg Scattering

The simplest type of coherent neutron scattering to understand is diffraction. Suppose that atoms are arranged at fixed positions on a lattice (such as the two-dimensional portion of the lattice shown in Fig. 6) and a beam of neutrons is fired at that lattice. We imagine that all of the neutrons move on parallel paths and have the same velocity, so that there is only one value for the incident wave vector, $\mathbf{k}$. Because the atoms and their associated nuclei are imagined to be fixed, there is no change in the neutron's energy during the scattering process; that is, the scattering is elastic and $k' = k$.

As the incident neutron wave arrives at each atom, the atomic site becomes the center of a scattered spherical wave that has a definite phase relative to all other scattered waves. In two dimensions, it is as if a handful of pebbles have been thrown into a calm pond. At the point where each pebble strikes the pond (the atomic site), a circular wave spreads outwards. Because the waves from each site overlap there will be places where the disturbances from different waves reinforce one another and other places where they cancel out. This is the phenomenon of interference.

As the waves spread out from a regular array of sites, the individual disturbances will reinforce each other only in particular directions. In other words, if we observe the wave motion at some distance from the lattice, we will see waves (scattered neutrons) traveling in well-defined directions (Fig. 6). These directions are closely related to the symmetry and spacing (or lattice) of the scattering sites—a hexagonal grid will generate a different set of wavefronts than a square grid. Consequently, one may use a knowledge of the directions in which various incident waves are scattered to deduce both the symmetry of the lattice and the distances between

coherent scattering

Scattering in which an incident neutron wave interacts with all the nuclei in a sample in a coordinated fashion; that is, the scattered waves from all the nuclei have definite relative phases and can thus interfere with each other.

incoherent scattering

Scattering in which an incident neutron wave interacts independently with each nucleus in the sample; that is, the scattered waves from different nuclei have random, or indeterminate, relative phases and thus cannot interfere with each other.

diffraction

A type of scattering in which coherently scattered waves interfere.

DIFFRACTION FROM A LATTICE

Fig. 6. As a neutron (represented in the figure by a plane wave) passes through a lattice of regularly spaced scattering centers (rather than the single scattering center of Fig. 4), the spherical wavefronts that represent scattered neutrons will overlap and interfere with each other. In those directions in which the interference is constructive, scattered neutrons may be measured. The figure depicts such constructive interference in *two* dimensions with planar wavefronts represented as lines, spherical wavefronts as colored circles, and the scattering centers as small circles. To simplify the diagram, the scattering is shown only for four centers (solid black) in each of the two rows of scattering planes. Also, color is used to relate each incident wavefront to the scattered wavefronts that have so far been generated by it. Thus, the incident red wavefront has passed over and scattered from four scattering centers in Scattering Plane 1; the orange wavefront has passed over and scattered from these scattering centers *plus* the leftmost scattering center in Scattering Plane 2; the yellow wavefront has passed over all eight scattering centers in both planes. For constructive interference to take place, Q must be perpendicular to the two scattering planes, and the condition $\mathbf{Q} \cdot (\mathbf{r}_j - \mathbf{r}_k) = Qd = 2\pi n$ must be satisfied, where $|\mathbf{r}_j - \mathbf{r}_k| = d$ is the distance between the two scattering planes and n is an integer. Combining this condition with $Q = 4\pi \sin \theta / \lambda$ (from Fig. 5a) yields Bragg's law: $n\lambda = 2d \sin \theta$.
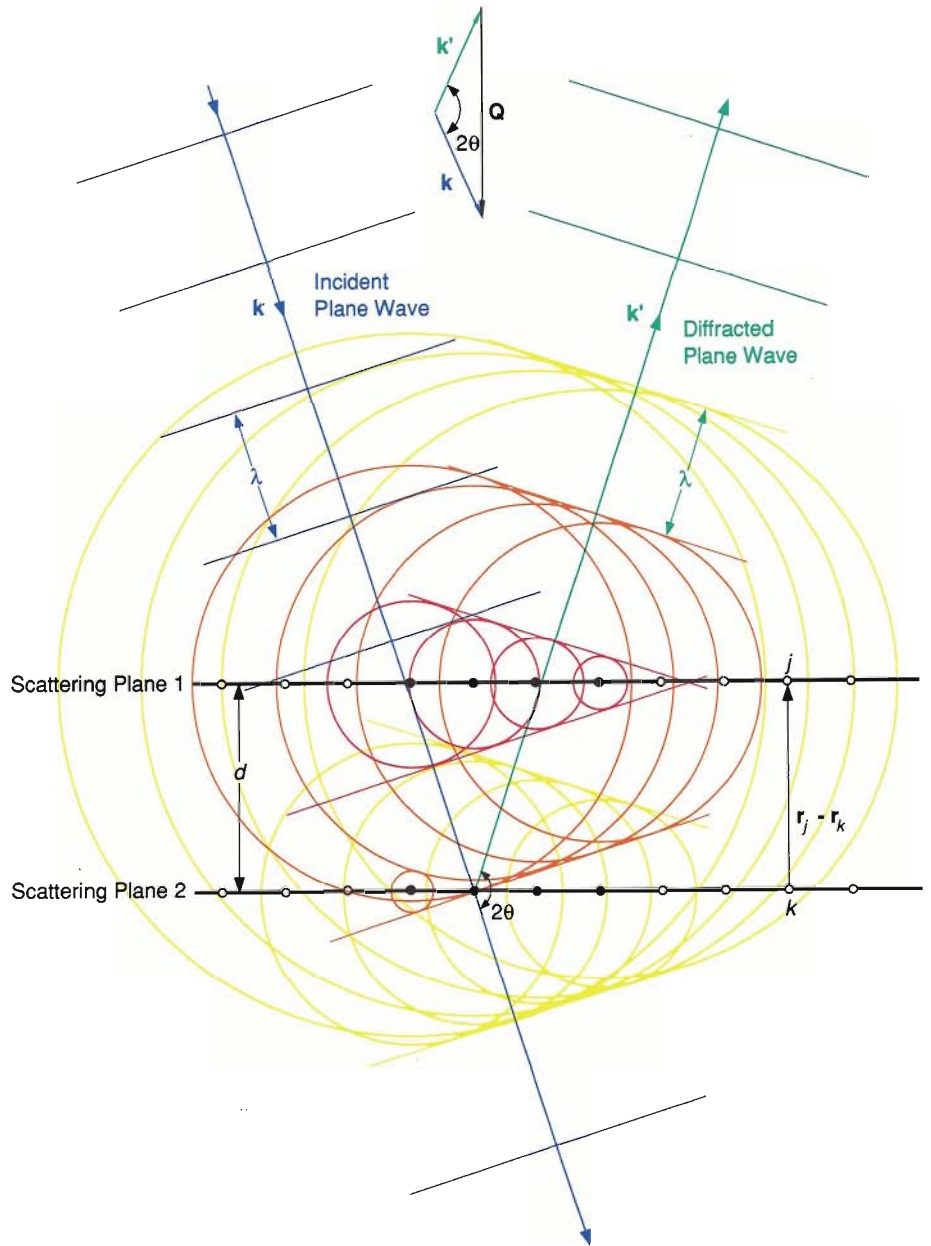
atoms. The type of scattering we have just described is called *diffraction*.

Because diffraction is an elastic, coherent scattering process, Van Hove's formulation of the scattering law reduces to a simple form. For a three-dimensional lattice with one isotope, the scattering law can be written (see "The Mathematical Foundations of Neutron Scattering") as

$$I(\mathbf{Q}) = \sum_{j,k} b_{\text{coh}}^2 e^{i\mathbf{Q} \cdot (\mathbf{r}_j - \mathbf{r}_k)}$$

$$= \left| \sum_j b_{\text{coh}} e^{i\mathbf{Q} \cdot \mathbf{r}_j} \right|^2 \quad (2)$$

$$\equiv S(\mathbf{Q}),$$

where $\mathbf{r}_j$ and $\mathbf{r}_k$ represent the positions of atoms labeled j and k in the lattice and b_{coh} is the coherent scattering length of those atoms.

Equation 2 is the scattered intensity that would be measured in a neutron-diffraction experiment with a real crystal, and is often called the *structure factor*, $S(\mathbf{Q})$. As we count through the atoms of a lattice performing the sum in Eq. 2, the real and imaginary parts of the exponential function both take values that are distributed essentially at random between plus and minus one. Because many atoms are involved, the sum usually averages to zero, except at certain unique values of $\mathbf{Q}$.

Obviously, the values of $\mathbf{Q}$ for which the structure factor, $S(\mathbf{Q})$, is nonzero are rather special, and it is easy to imagine that not many values of $\mathbf{Q}$ satisfy this condition. Further, those values are intimately related to the structure of the crystal because the vectors $\mathbf{r}_j - \mathbf{r}_k$ in Eq. 2 represent the set of distances between different atoms in the crystal.

We can determine the values of $\mathbf{Q}$ at which $S(\mathbf{Q})$ is nonzero and at which diffraction occurs by consulting Fig. 6. Suppose $\mathbf{Q}$ is perpendicular to a plane of atoms such as Scattering Plane 1 in this figure. If the value of $\mathbf{Q}$ is any integral multiple of $2\pi/d$, where d is the distance between parallel, neighboring planes of atoms (Scattering Planes 1 and 2 in Fig. 6), then $\mathbf{Q} \cdot (\mathbf{r}_j - \mathbf{r}_k)$ is a multiple of 2π and $S(\mathbf{Q})$ is nonzero because each exponential term in the sum in Eq. 2 is unity. Thus, $\mathbf{Q}$ must be perpendicular to *planes* of atoms in the lattice and its value must be an integral multiple of $2\pi/d$. For values of $\mathbf{Q}$ that do not satisfy this condition, $S(\mathbf{Q}) = 0$, and there is no scattering.

The values of $\mathbf{Q}$ at which neutron diffraction occurs are governed by the same law that was discovered for x rays in 1912 by William and Lawrence Bragg, father and son. To see this, we apply the condition described above
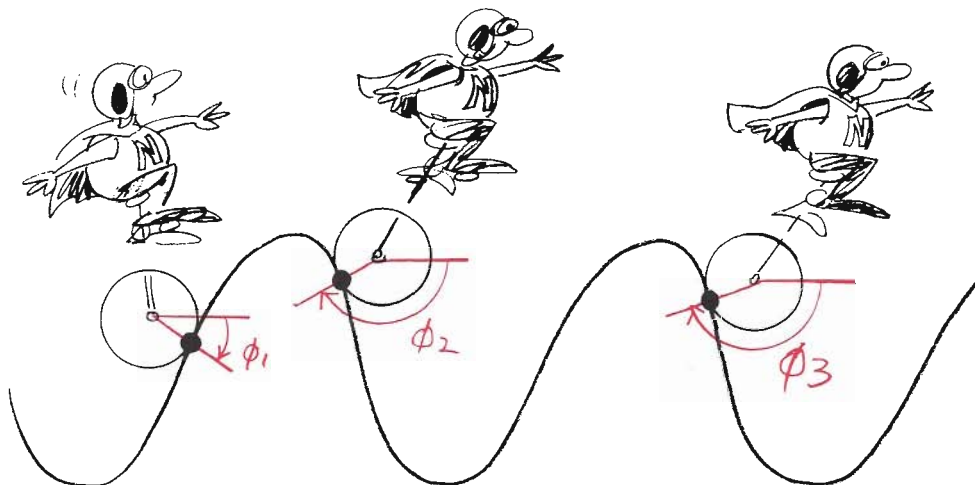
$(Q = n(2\pi/d)$, where n is an integer) to the scattering triangle for elastic scattering. Then using the relationship between Q, θ, and λ shown in Fig. 5, the condition can be rewritten as

$$n\lambda = 2d \sin\theta. \qquad (3)$$

This equation, called *Bragg's law*, relates the scattering angle, 2θ, to the interplanar spacing in a crystalline sample.

Bragg's law can also be understood in terms of the path-length difference between waves scattered from neighboring planes of atoms (Fig. 7). For constructive interference to occur between waves scattered from adjacent planes, the path-length difference must be a multiple of λ, the wavelength. Applying this condition to Fig. 7 immediately yields Bragg's law in the form given in Eq. 3. Many of the results described in the articles in this issue will fall back on this point of view.

Diffraction, or Bragg scattering, as it is sometimes called, may occur for any set of atomic planes that we can imagine in a crystal, provided the wavelength, λ, and the angle, θ, between the incident neutron beam and the planes are chosen to satisfy Eq. 3. Bragg scattering from a particular set of atomic planes resembles reflection from a mir-

relative phase

The phase angle $\phi = \mathbf{k} \cdot \mathbf{r}$ (mod 2π) appears in the expression $Ae^{i\mathbf{k} \cdot \mathbf{r}}$ describing a plane wave of amplitude A. For a plane wave traveling in the x direction with wave vector $k_x = 2\pi/\lambda$, we can write the phase as $\phi = \frac{2\pi x}{\lambda}$ (mod 2π). Thus as the wave travels a distance λ, its phase changes by 2π. When two waves with wave vector $\mathbf{k}$ and equal amplitude A are in phase, their phases at any point in space are the same and the waves add constructively to yield an intensity of $4A^2$. When the relative phase $(\phi_1 - \phi_2)$ of two waves is nonzero, the waves will interfere with each other so that their resulting intensity fluctuates in space and is always less than $4A^2$. Incoherent scattering produces random changes in the phase of the incident wave so that the relative phases of the scattered waves are indeterminate, the waves do not interfere with each other, and the intensity of each wave is added separately to yield the total intensity.

THE PATH-DIFFERENCE
APPROACH TO BRAGG'S LAW

Fig. 7. Constructive interference occurs when the waves reflected from adjacent scattering planes remain in phase. This happens when the difference in distance traveled by waves reflected from adjacent planes is an integral multiple of the wavelength. The figure shows that the extra distance (shown in red) traveled by the wave reflected from Scattering Plane 2 is $2d \sin \theta$. When that distance is set equal to $n\lambda$, the result is, once again, Bragg's law: $n\lambda = 2d \sin \theta$. Primary scattering occurs when $n = 1$, but higher-order Bragg peaks are also observed for other values of n.
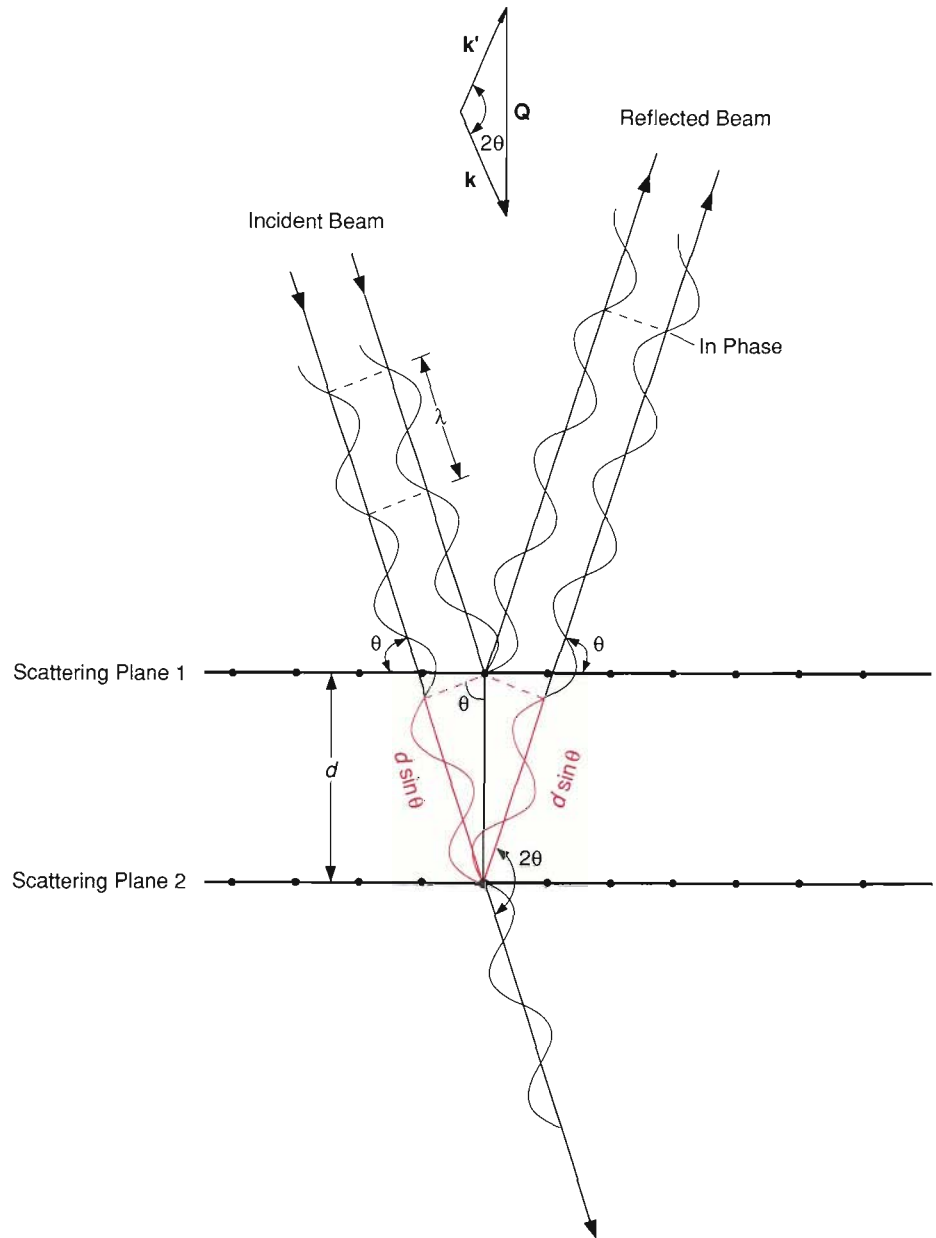
ror parallel to those planes: the angle between the incident beam and the plane of atoms equals the angle between the scattered beam and the plane (Fig. 7). If a beam of neutrons of a particular wavelength is incident on a single crystal, there will, in general, be no diffraction. To obtain diffraction for a set of planes, the crystal must be rotated to the correct orientation so that Bragg's law is satisfied—much as a mirror is adjusted

to reflect the sun at someone's face. The signal thus observed by a neutron detector at a particular scattering angle is called a Bragg peak because as we rotate the crystal to obtain diffraction we observe a peak in the signal being recorded.

According to Eq. 2, the intensity of the scattered neutrons is proportional to the square of the density of atoms in the atomic planes responsible for the

scattering. We can see this by noting that as the summation is carried out for each atom j in one plane, unit exponential factors are added for all the atoms k in another plane. And the more closely the atoms are spaced within a reflecting plane, the more unit factors will be summed per unit area. Thus, an observation of Bragg peaks allows us to deduce both the spacing of planes (from Bragg's law) and the density of the atoms in the planes. To measure Bragg peaks corresponding to many different atomic planes with neutrons of a particular wavelength, we have to vary both the scattering angle, 2θ, and the crystal orientation. First we choose the detector position so that the scattering angle satisfies Bragg's law, then we rotate the crystal until a Bragg diffracted beam falls on the detector.

To this point we have been discussing a simple type of crystal that can be built from *unit cells*, or building blocks, that each contain only one atom. In this case, each of the exponential factors that contribute to $S(\mathbf{Q})$ in Eq. 2 is unity, and the structure is easily deduced from the intensities of the Bragg peaks and the scattering angles at which Bragg diffraction occurs. However, the unit cells of materials of interest to chemists or biologists almost invariably have more complicated shapes and contain many different types of atoms distributed throughout their volumes. Those atoms, of course, are not positioned randomly in the unit cell but are arranged in a geometric pattern determined by the way they bond together. Nevertheless, it may not be trivial to deduce the atomic positions from an observation of Bragg scattering because some of the exponential factors that contribute to $S(\mathbf{Q})$ are now complex and the phases of these quantities cannot be obtained directly from a measurement of Bragg diffraction. Deducing the structure of a complex material may take several months and a great deal of ingenuity.

In diffraction experiments with single crystals, the sample must be correctly oriented with respect to the neutron beam to obtain Bragg scattering. Furthermore, if neutrons of a single wavelength are used, the detector must also be positioned at the appropriate scattering angle for the atomic planes causing the scattering. On the other hand, polycrystalline powders, which consist of many randomly oriented single-crystal grains, will diffract neutrons whatever the orientation of the sample relative to the incident beam of neutrons. There will always be grains in the powder that are correctly oriented to diffract. Thus, whenever the scattering angle, 2θ, and the neutron wavelength, λ, satisfy the Bragg equation (Eq. 3) for a set of planes, a reflection will be detected, independent of the sample orientation. This observation is the basis of a widely used technique known as *powder diffraction*, which is implemented in slightly different ways depending on the nature of the neutron source. Before describing powder diffraction in greater detail, we digress to consider the different techniques that may be used to produce neutrons for scattering experiments.

Neutron Production

Neutron-scattering facilities throughout the world generate neutrons either with nuclear reactors or with high-energy particle accelerators. The neutrons produced have energies up to tens or even hundreds of mega-electron volts (MeV), and the corresponding neutron wavelengths are far too short for investigating condensed matter. Furthermore, neutrons whose energies are above a few electron volts tend to damage solids by knocking atoms out of their official positions, producing vacancies and interstitials. For this reason, neutrons must be "cooled down" before being used for scattering experiments.
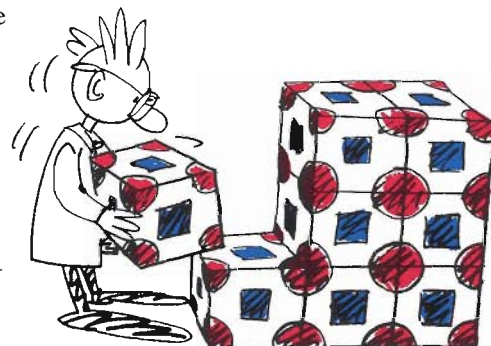
unit cell

The repeating unit of a crystal.
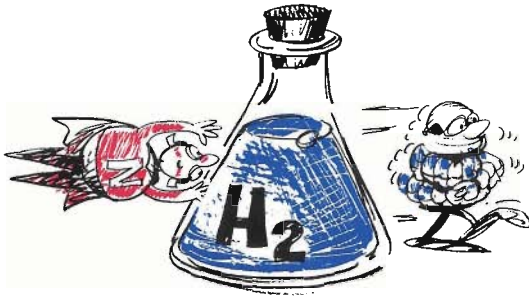
cold neutrons

Neutrons whose energies have been reduced below about 5 meV by inelastic scattering in a cold material such as liquid hydrogen or deuterium. Researchers use such longer-wavelength neutrons to conduct experiments at larger length scales.

Such cooling is done by bringing the neutrons into thermal equilibrium with a "moderating" material—a material with a large scattering cross section, such as water or liquid hydrogen. The moderator, whose volume may vary from a deciliter to several tens of liters, is placed close to the neutron source. Neutrons enter the moderator and, in a series of collisions in the material, lose energy to recoiling moderator atoms. After a few tens of collisions, the energies of the neutrons are similar to those of the atoms of the moderator. Thus, *thermal* neutrons are emitted from the moderator surface with a spectrum of energies around an average value determined by the moderator temperature. The average energy of neutrons from a water moderator at ambient temperature is about 25 thousandths of an electron volt (25 meV), and the average energy from a liquid-hydrogen moderator at 20 kelvins is around 5 meV. The wavelength of a 25-meV neutron is 1.8 angstroms (1.8×10^{-10} meter), which is of the same order as typical interatomic distances and, therefore, is quite suitable for diffraction experiments.

Reactor Sources. Neutrons are produced in a nuclear reactor by the fissioning of atoms in the reactor fuel, which, for research reactors, is invariably uranium. The neutrons are moderated in the manner described above and allowed to emerge from the reactor in a continuous stream with an energy spectrum similar to the curves of Fig. 8a.

For most scattering experiments at reactors, the neutrons emerging from the moderator must be reduced to a *monochromatic* beam; that is, only those neutrons in a single, narrow energy band are selected from the spectrum. This selection is usually accomplished by Bragg reflection from a large single crystal of a highly reflective material, such as pyrolytic graphite, germanium, or copper. A crystal monochromator

(a) Reactor Neutrons

(b) Spallation Neutrons

Energy (meV)

REACTOR AND SPALLATION NEUTRONS

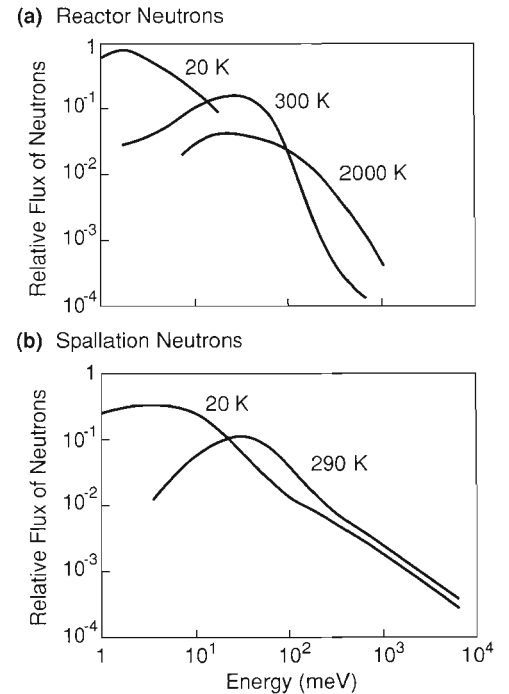Fig. 8. (a) The relative flux of neutrons as a function of energy for the high-flux reactor at the Institut Laue-Langevin in Grenoble, France. The curves show the distribution of neutrons emerging from moderators at temperatures of 20, 300, and 2000 kelvins. (b) Similar distribution curves for neutrons generated at the Manuel Lujan, Jr. Neutron Scattering Center at Los Alamos (LANSCE) by moderators at temperatures of 20 and 290 kelvins.

works because, even though the incident beam contains neutrons of many wavelengths, the spacing of the reflecting planes of atoms, d, and the scattering angle, 2θ, are chosen so that only those neutrons with a wavelength satisfying the Bragg equation are transmitted in the direction of the experiment. The wavelength of the neutrons used for experiments can then be controlled by changing the scattering angle at the monochromator.
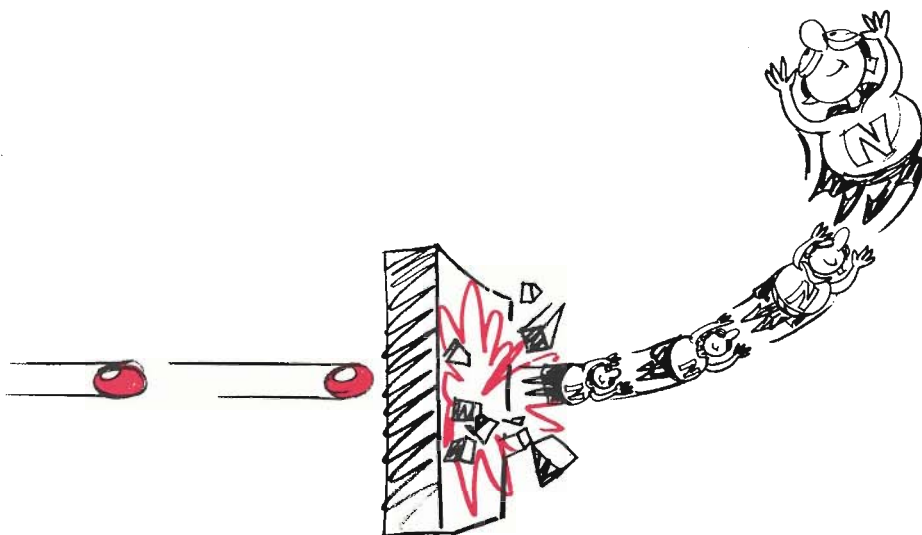
Spallation Sources. Other neutron facilities, such as the one at the Manuel Lujan, Jr. Neutron Scattering Center at Los Alamos (LANSCE), use accelerators to produce *spallation* neutrons. This is done by allowing high-energy protons (or, less effectively, electrons) to collide with a heavy-metal target, such as tungsten or uranium, driving neutrons from the nuclei of the target. The protons are produced by the accelerators—in this case, LAMPF (the Los Alamos Meson Physics Facility) coupled with a proton storage ring—in bursts that last for less than a microsecond. At LANSCE there are 20 such bursts of 800-MeV protons per second. Each proton in the burst then generates about 20 neutrons.

One of the advantages of a spallation source is that only a small amount of energy—about 27 MeV per neutron—is deposited in the spallation target by the protons. Nuclear fission produces about four or five times as much energy in generating each of its neutrons. However, the cost of producing the high-energy protons—the electricity bill of the accelerators—is not cheap.

The moderated neutrons that finally emerge into the experimental area from a spallation source have a spectrum resembling the curves of Fig. 8b. Clearly, this spectrum is quite different from

that produced by a reactor (Fig. 8a) because there is a greater percentage of high-energy neutrons. However, the spectrum is not the only difference between the two types of neutron sources. Neutrons from a spallation source arrive in pulses rather than continuously as they do at a reactor. This fact means that the monochromator crystal needed at reactors can here be avoided and *all* the neutrons can be used (rather than only those in a narrow energy band).

The trick that allows the use of all neutrons from a spallation source relies on the measurement of the time it takes for each detected neutron to traverse the distance between the moderator and the detector. From this *time of flight*, the neutron velocity can be determined, and Eq. 1 gives its wavelength. Generating a monochromatic beam is therefore unnecessary.

A thermal neutron with an energy of 25 meV travels at a speed of about 2.2 kilometers per second, or about Mach 7. A typical neutron spectrometer is about 10 meters long, so the neutron travels from the moderator to the detector in about 5 milliseconds. Because the duration of the neutron pulse emerging from the moderator of a pulsed source is typically a few tens of microseconds, the time of flight of the neutron can be determined with high relative precision.
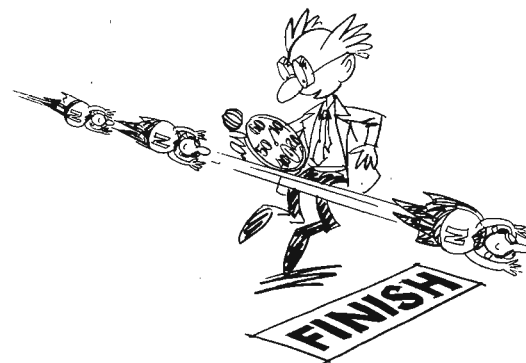
Powder Diffraction

Now let's return to powder diffraction. In a powder-diffraction instrument at a *reactor* source (Fig. 9), a monochromatic beam of neutrons impinges on a powdered sample, and the neutrons scattered from the sample are recorded as a function of the angle 2θ. Each Bragg peak in a typical scattering pattern (Fig. 10) corresponds to diffraction from atomic planes with a different interplanar spacing, or value of d. Many peaks can be recorded simultaneously by placing detectors at a variety of scattering angles (such as the sixty-four helium-3 detectors in Fig. 9).

In a powder diffraction instrument at a *spallation* source (Fig. 11), the sample is irradiated with a pulsed beam of neutrons having a wide spectrum of energies. Scattered neutrons are recorded in banks of detectors located at different scattering angles, and the time at which each scattered neutron arrives at the detector is also recorded. At a particular scattering angle, the result is a diffraction pattern very similar to that measured at a reactor, but now the independent variable is the neutron's time of flight rather than the scattering angle. Because the neutron's time of flight is proportional to its wavelength and, for constant scattering angle, wavelength is proportional to the spacing between atomic planes (Eq. 3), the measured neutron scattering can be plotted against either time of flight, λ, or d-spacing (Fig. 12). (The resemblance between Figs. 10 and 12 is obvious. The patterns are equivalent ways of probing Bragg's law, and in fact, diffraction data obtained at reactors and spallation sources can be plotted on the same scale by simply using $Q = 4\pi \sin\theta/\lambda$ as the independent variable.)

As in the reactor case, detectors at a spallation source can be placed at different scattering angles, allowing many patterns to be measured simultaneously.
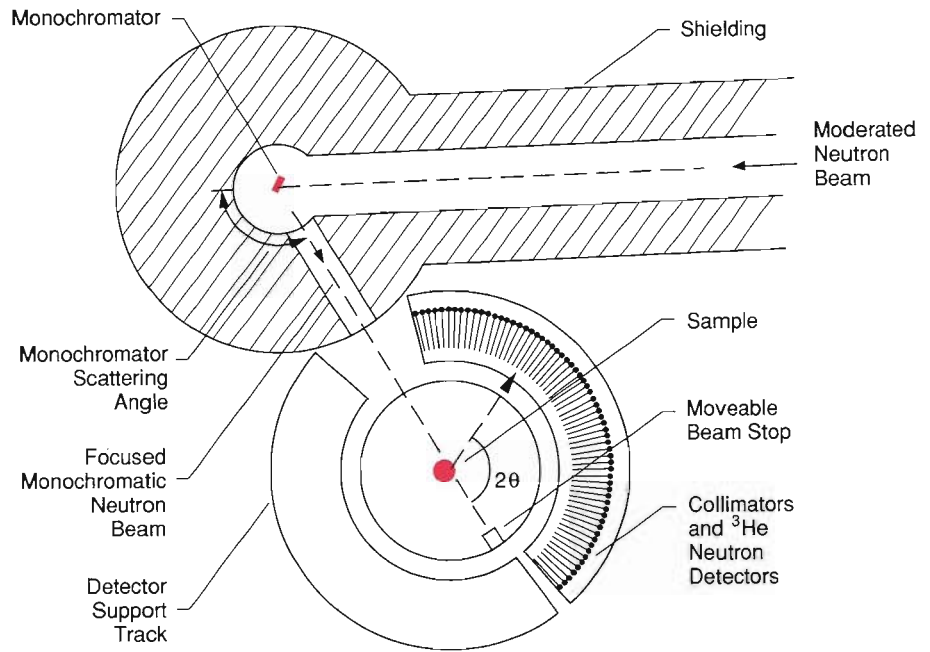
POWDER DIFFRACTION AT A REACTOR SOURCE

Fig. 9. An essential component of a powder diffractometer at a high-flux reactor is a very large crystal whose reflecting surface may be as large as 200 square centimeters. The crystal acts as a monochromator by scattering neutrons of a given energy toward the sample. To help focus the beam of neutrons, the crystal may also be curved, effectively acting as a concave mirror. A second scattering occurs at the powder sample, which scatters the monoenergetic focused beam toward a set of detectors (here, 64 helium-3 neutron detectors). These detectors are here shown positioned along an arc on one side of the sample, but the whole array can be moved to other positions along the circular support track. The distance between the monochromator and the sample is typically about 2 meters.
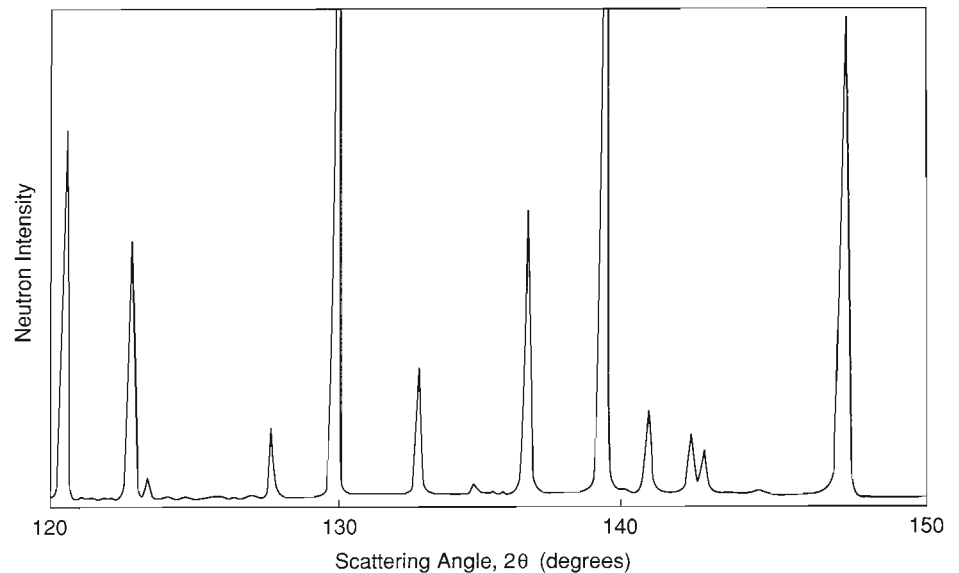
A POWDER DIFFRACTION PATTERN RECORDED AT A REACTOR

Fig. 10. A typical powder diffraction pattern obtained at a reactor source gives intensity, or numbers of neutrons, as a function of the scattering angle 2θ. Each peak represents neutrons that have been scattered from a particular set of atomic planes in the crystalline lattice.
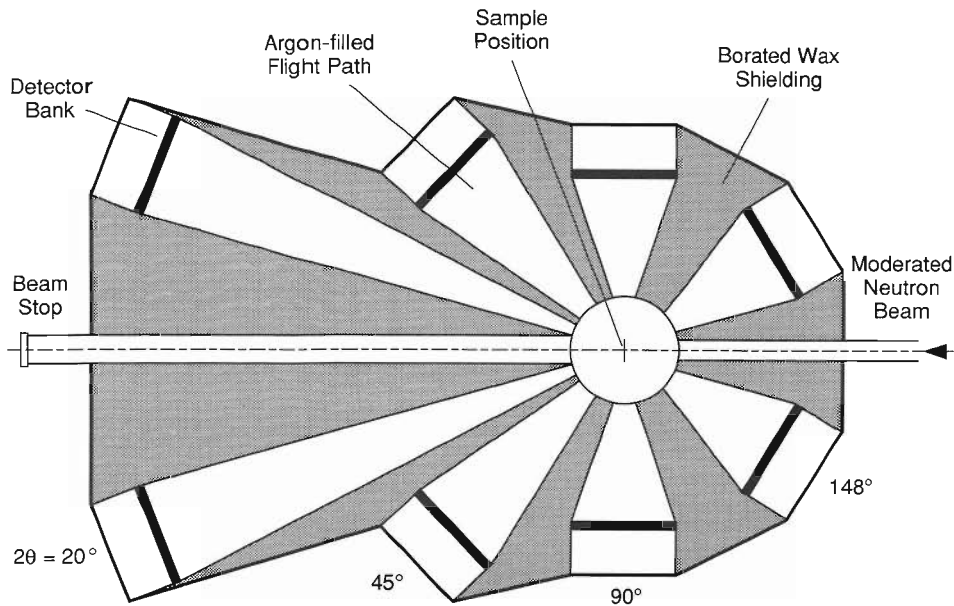
POWDER DIFFRACTION AT A SPALLATION SOURCE

Fig. 11. The Neutron Powder Diffractometer (NPD) at LANSCE (see photograph on page 54). The incident beam of neutrons, having been moderated with water chilled to $10°C$, is directed onto the target in a large evacuated chamber. Surrounding this chamber are eight banks of detectors positioned at fixed scattering angles. Each bank consists of sixteen helium-3 detectors, and the d-spacing that can be measured ranges from about 1.2 to 33.6 angstroms at the $20°$ detector bank to about 0.25 to 5.2 angstroms at the $148°$ detector bank. The distance between the sample and the detectors at the $90°$ scattering angle is about 2 meters, so the whole spectrometer is very much larger than the equivalent instrument at a reactor.
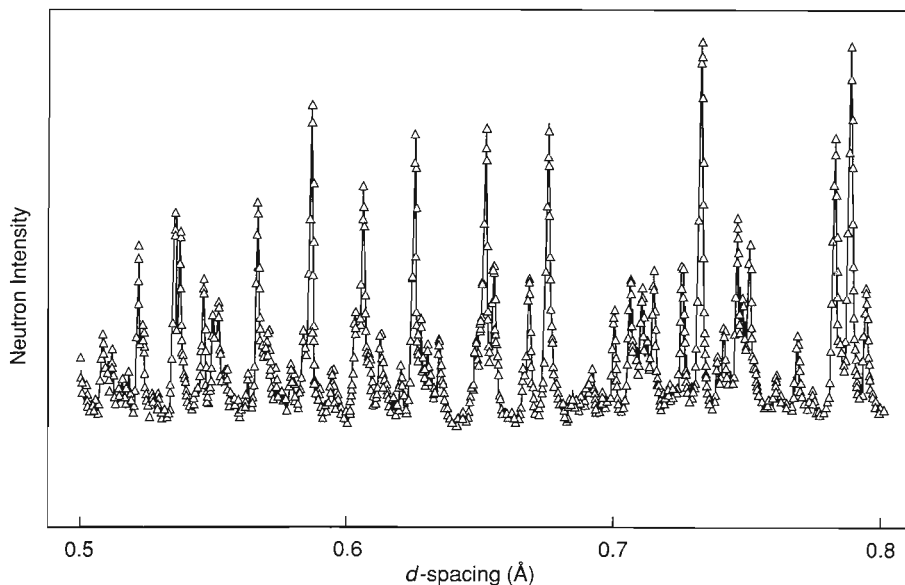
A POWDER DIFFRACTION PATTERN RECORDED AT A SPALLATION SOURCE

Fig. 12. A typical powder diffraction pattern obtained at a spallation source ("fat garnet" measured at one of the $148°$ bank of detectors in the diffractometer of Fig. 11). As in Fig. 10, the vertical coordinate is the intensity, or number of neutrons, but the horizontal coordinate is the d-spacing between atomic planes. The horizontal variable could as well be λ (via Bragg's law, Eq. 3) or the neutron time of flight (via Eqs. 1 and 3).

Detectors at small scattering angles provide information about widely spaced atomic planes, whereas those at larger angles record data relevant to small spacings. There is usually some overlap of information provided by the different detectors.

Using patterns like those of Figs. 10 and 12, the atomic structure of a polycrystalline sample may be deduced from Eq. 2. In practice, however, one guesses the atomic positions, evaluates Eq. 2, and from a comparison of the calculated and measured diffraction patterns, refines the atomic coordinates. This type of procedure is described in detail in the article "X-Ray and Neutron Crystallography—A Powerful Combination" by Robert Von Dreele.

Probing Larger Structures

Another way of thinking about coherent elastic neutron scattering is shown in Fig. 13. One can imagine the incident and scattered neutron waves setting up a "probe wave" in the sample—much as two misaligned picket fences generate a set of moiré fringes. One can alter the wavelength of the probe wave, λ_{probe}, by changing the angle between the ingoing and outgoing waves (that is, the scattering angle) or by increasing or decreasing the wavelength of the neutrons used. To obtain information about structures by coherent elastic scattering, λ_{probe} must be chosen to be approximately the same as the size of the structure. For crystallography this means that λ_{probe} needs to be of the same order as interatomic spacings. We already know this from Bragg's law. A little trigonometry applied to Fig. 13 will show that $\lambda_{probe} = \lambda_{neutron}/2\sin\theta$, so that when λ_{probe} equals the distance between two adjacent scattering planes, Bragg's law is satisfied.

The probe-wave idea shows us how we can measure structures that are larger than typical interatomic distances. We

THE PROBE-WAVE VIEW
OF NEUTRON SCATTERING

Fig. 13. Another way to view neutron scattering is to imagine that the incident neutron wave (In) and the scattered neutron wave (Out) form a secondary "probe wave" (here seen as a moiré pattern in both examples) that must match the average periodicity of the structure in the scattering sample. Because the average periodicity of the top sample is larger than that of the lower one, the wavelength of the probe wave, λ_{probe}, must also be larger, which in turn means that the scattering angle, 2θ, must be smaller (here $3°$). Another way to vary λ_{probe} is to change the wavelength of the neutron, $\lambda_{neutron}$.
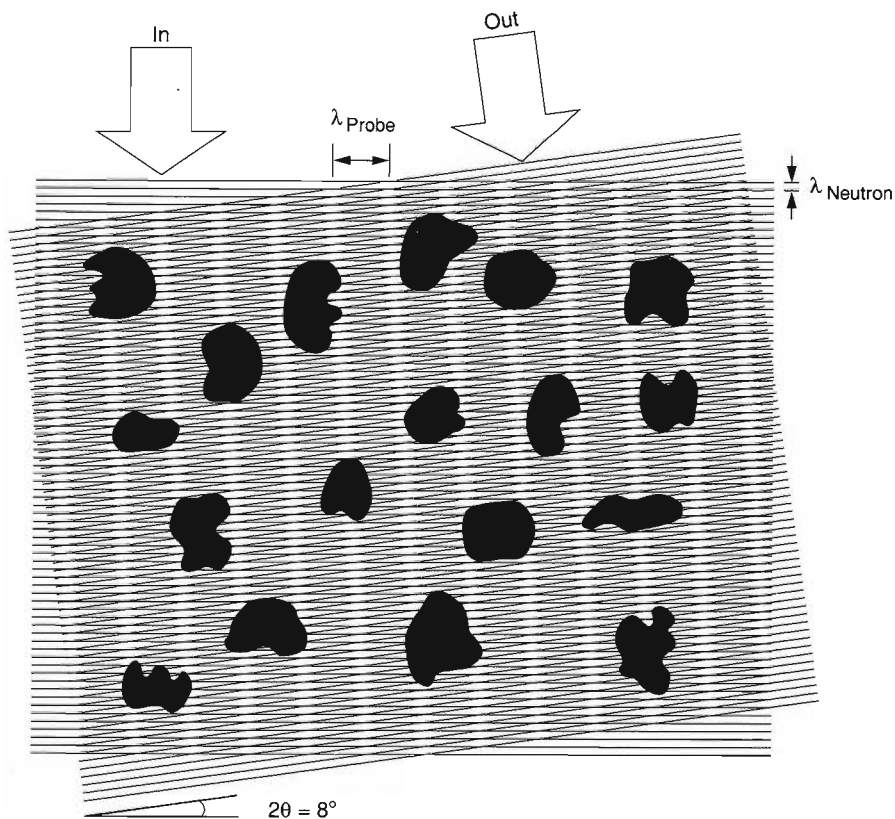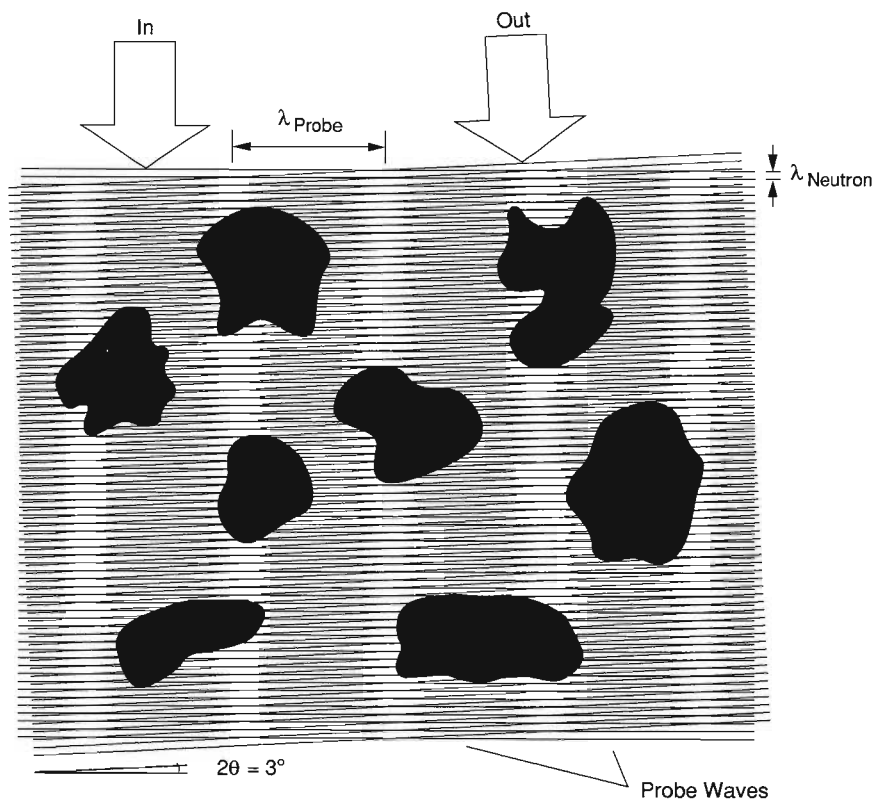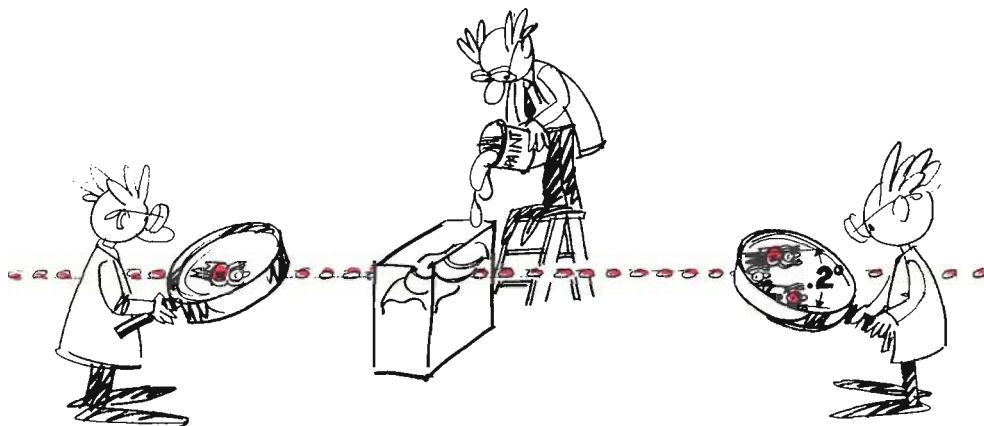
simply arrange for λ_{probe} to be large, either by decreasing the scattering angle or by increasing the neutron wavelength. In practice, to examine some of the larger structures displayed in Fig. 3—polymers, colloids, or viruses, for example—we need to use neutron wavelengths greater than 5 angstroms and scattering angles less than 1 degree. Because of the latter constraint, this technique is known as *small-angle neutron scattering*, or SANS.

The Van Hove formulation for neutron scattering may be manipulated (see "The Mathematical Foundations of Neutron Scattering") to provide the following equation for the intensity of neutrons scattered at small angles (that is, for small values of Q):

$$I(\mathbf{Q}) = \left| \int b(\mathbf{r})e^{i\mathbf{Q}\cdot\mathbf{r}} \, d^3r \right|^2, \qquad (4)$$

where the integral extends over the entire scattering sample and $b(\mathbf{r})$, the *scattering-length density*, is calculated by summing the coherent scattering lengths of all the atoms over a small volume and dividing by that volume.

In many cases, samples measured by SANS consist of particles with a uniform scattering-length density b_p that are dispersed in a uniform matrix with a scattering-length density b_m. Examples include pores in rock, colloidal dispersions, biological macromolecules in

water, and many more. The integral in Eq. 4 can, in this case, be separated into a uniform integral over the whole sample and a term that depends on the difference, $b_p - b_m$, between the scattering length of the particles and that of the matrix. This difference is called the *contrast factor*. If all the particles are identical and their positions are uncorrelated, Eq. 4 becomes

$$I(\mathbf{Q}) = \left(b_p - b_m\right)^2 N_p \left| \int_{V_p} e^{i\mathbf{Q}\cdot\mathbf{r}} d^3r \right|^2, \qquad (5)$$

where the integral is now over the volume V_p of one of the particles and N_p is the number of such particles in the sample.

The integral above of the phase factor $e^{i\mathbf{Q}\cdot\mathbf{r}}$ over a particle is called the *form factor* for that particle. For many simple particle shapes, the form factor can be evaluated without difficulty: the expression for spherical objects was first derived by Lord Rayleigh in 1911.

Equation 5 allows us to understand an important technique used in small-angle scattering known as *contrast matching*. The total scattering is proportional to the square of the scattering contrast between a particle and the matrix in which it is embedded. If we embed the particle in a medium whose scattering length is equal to that of the particle,
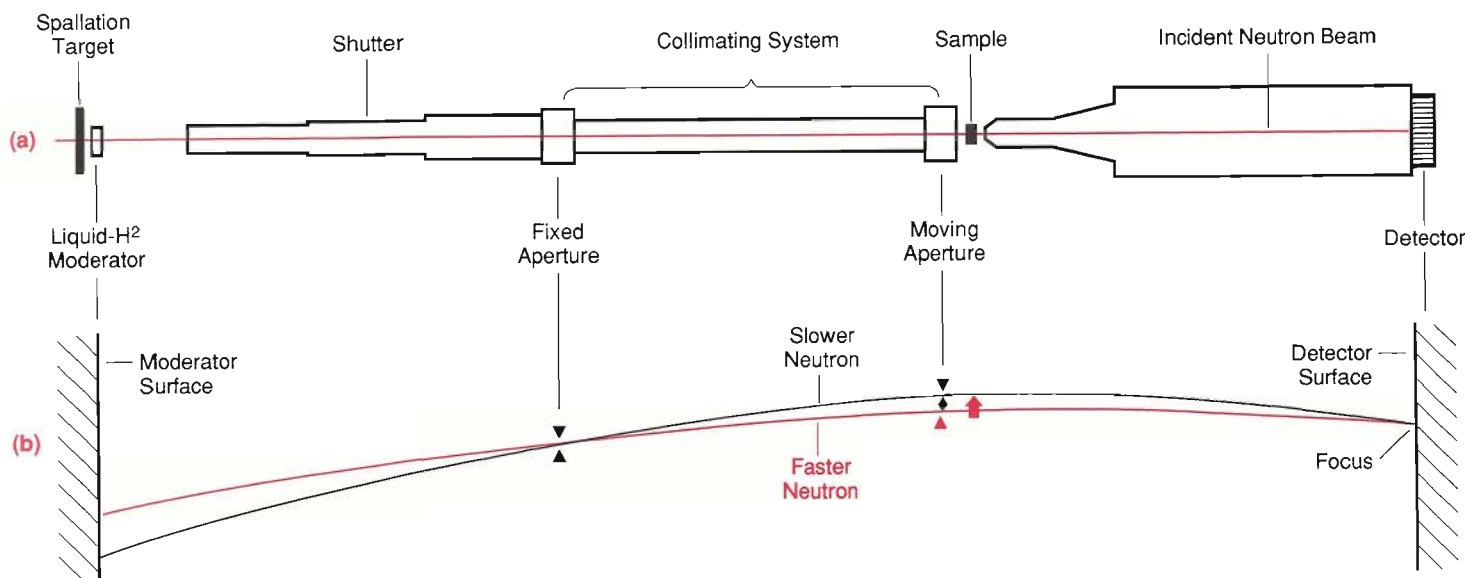
A SMALL-ANGLE NEUTRON-SCATTERING SPECTROMETER

Fig. 14. (a) The spectrometer illustrated here, the Low-Q Diffractometer (LQD) at LANSCE, measures neutron scattering at small angles. The neutrons are first moderated in liquid hydrogen to increase the percentage of very cool, long-wavelength neutrons in the beam that hits the sample. The moderated beam then passes through a collimating system that is more than 7 meters long before impinging on the sample. To increase the accuracy with which the small scattering angles can be measured, the large position-sensitive detector is placed far from the sample (about 4 meters). (b) Neutrons from a spallation source have a range of speeds and are thus under the influence of gravity for different amounts of time, an effect that smears the signal at the detector. However, the beam can be "focused" by placing a fixed aperture at the beginning of the collimator and a moveable aperture at the end of the collimator and accelerating the latter aperture upward during the pulse of neutrons. Such an arrangement selects only those neutrons with parabolic trajectories that end at the center, or focus, of the detector. Small-angle scattering is suitable for studying structures with dimensions in the range of 10 to 1000 angstroms.

the latter will be invisible. (This technique is used by the manufacturers of gel toothpaste—there really is gritty material in there to clean your teeth, but you can't see it because the grit and the gel have similar refractive indices!)

Suppose the particles we are interested in are spherical eggs rather than uniform spheres: they have a core (the yolk) with one scattering length and a covering (the white) of a different scattering length. If such particles are immersed in a medium whose scattering length is equal to that of the egg white, then a neutron-scattering experiment will only "see" the yolk. The form factor will be evaluated by integrating over this central region only. On the other hand, if our particles are suspended in a medium whose scattering length is the same as that of the yolk, only the egg white will be visible; the form factor will correspond to that of a thick, hollow shell. The scattering pattern will be different in the two cases, and from two experiments, we will discover the structures of both the covering and the core of the particle.

Variation of the scattering-length density of the matrix is often achieved by choosing a matrix that contains hydrogen (such as water). By replac-

ing different fractions of the hydrogen atoms with deuterium atoms, a large range of scattering-length densities can be achieved for the matrix. This contrast-matching technique works, as we pointed out earlier, because of the significantly different scattering-length densities of hydrogen and deuterium, and it is one of the main reasons for the successful application of neutron scattering to problems in biology. Both DNA and protein can be contrast matched by water containing different fractions of deuterium. Several problems in structural biology that have been studied by contrast matching are described in "Biology on the Scale of Neglected Dimensions" by Jill Trewhella.

Small-angle scattering is perhaps the easiest neutron-scattering technique to realize in practice. Like diffraction experiments, SANS experiments at a reactor source require a monochromator, whereas at a spallation source measurement of times of flight determine the wavelengths of the incident and scattered neutrons.

The Low-Q Diffractometer at the LANSCE facility (Fig. 14a) is an example of a SANS spectrometer at a spallation source. One essential component of the instrument is a large position-

sensitive neutron detector located behind the sample directly in line with the incident beam. Another important component (invented by Phil Seeger at LANSCE) is the gravity focuser (Fig. 14b), which accounts for the fact that neutrons fall under the influence of gravity. If the aperture at the exit of the collimator that defines the trajectory of the incident neutron beam was fixed, neutrons of different velocities could only pass through that slit if they were following parabolic paths that fell on the detector at different heights, blurring the image produced there. To avoid this blurring, the exit aperture of the collimator is moved upward during each neutron pulse. Slower neutrons then have to go through an opening that is higher relative to the center of the detector. The position of the aperture at each instant is chosen so that all neutrons, independent of their speed, arrive at the center of the detector, if they are not scattered by the sample. The whole thing is a little like a stone-throwing contest: weak throwers have to throw stones on a higher trajectory to hit the target.

Inelastic Scattering

In reality, atoms are not frozen in fixed positions in a crystal. Thermal energy causes them to oscillate about their lattice sites and to move around inside a small volume with the lattice site at its center. Since an atom can fully contribute to the constructive interference of Bragg scattering only when it is located *exactly* at its official position in the lattice, this scattering becomes weaker the more the atoms vibrate and the less time they spend at their official positions.

When a crystal structure is determined from single-crystal or powder diffraction, the extent of the thermal motion of the atoms is found at the same time as the atomic positions. Often, the thermal motions are anisotropic, indicating that it is easier for an atom to move

in particular directions away from its equilibrium position. Sometimes this information can be related to other properties, such as structural changes that occur at a phase transition or elastic anisotropy.

Although such weakening of the scattering signal is the only effect of the thermal motion of atoms on *elastic* Bragg scattering, it is not the only way to use neutrons to observe atomic motion. In fact, one of the great advantages of neutrons as a probe of condensed matter is that they can be used to measure the details of atomic and molecular motions by measuring *inelastic* scattering. In other words, when the neutron bounces off a molecular framework that is *not* totally rigid, we can have an inelastic interaction with an exchange of energy between neutrons and the lattice.

To explain this, we begin with another simple analogy. If one end of a rope is tied to a fixed point and the other end is jerked up and down, a wave can be observed traveling along the rope. A discontinuous version of this effect can be obtained with a chorus line (for this analogy I am indebted to a colleague who once choreographed it for a midwestern television station). If each member of the line swings a leg but starts the swing slightly after his or her nearest neighbor to one side, the net effect is the appearance of a wave traveling along the line. The thermal motion of atoms in a crystal can be described in terms of a superposition of waves of this sort. One may imagine the atoms to be the feet of the members of the chorus line.

The analogy, if not the image, can be improved by replacing the swinging legs with rigid pendulums with weights at their extremities. Rather than watching for a neighbor to swing a leg, we achieve coupling by attaching identical springs between each pendulum and its two nearest neighbors. Now, if we dis-

place one pendulum, the springs tend to cause the neighboring pendulums to move as well, and a wave starts passing up and down the line, just as it did for the chorus. The frequency of motion depends on the mass of the pendulums and the stiffness of the springs that connect them.

Waves similar to those in the chain of pendulums pass through a lattice of atoms connected by the binding forces that are responsible for the cohesion of matter. The whole effect is much more difficult to visualize in this case, however, because it happens in three dimensions. Nevertheless, it is possible to prove that any atomic motion in a crystal can be described by a superposition of waves of different frequencies and wavelengths traveling in different directions. In other words, the thermal motion of the atoms about their lattice sites can be described as a superposition of waves moving through the lattice, and these waves are known as *phonons*. Their energies are quantized so that each phonon has an energy $h\nu$, where ν is the frequency of atomic motion associated with that phonon. Just as in the pendulum analogy, the frequency of a phonon depends on the wavelength of the distortion, the masses of the atoms, and the stiffness of the "springs," or binding forces, that connect them.

When a neutron is scattered by a crystalline solid, it can absorb or emit an amount of energy equal to a quantum of phonon energy, $h\nu$. This gives rise to inelastic coherent scattering of neutrons in which the neutron energy before and after the scattering event differ by an amount ϵ equal to the phonon energy. In most solids ν is a few times 10^{12} hertz, and the corresponding phonon energy is a few meV (10^{12} hertz corresponds to an energy of 4.18 meV). Because the thermal neutrons used for scattering experiments also have energies in the meV range, scattering by a phonon causes an appreciable fractional

phonons

Fundamental vibrational waves in a crystal in which nuclei oscillate in a coordinated manner about their "official" positions.

change in the neutron energy. This allows an accurate measurement of the energy change and makes neutrons an ideal tool for measuring phonon frequencies and hence for obtaining information about the forces that hold matter together.

For inelastic scattering—from phonons, for example—a neutron has different velocities, and thus different wave vectors, before and after it interacts with the sample; so the corresponding sides of the scattering triangle ($\mathbf{k}$ and $\mathbf{k}'$ in Fig. 5b) are of unequal lengths. To determine the phonon energy and the scattering vector, $\mathbf{Q}$, we need to determine the neutron wave vector before and after the scattering event. At a reactor we may resort to the method already discussed—Bragg scattering from single crystals. A first crystal, the monochromator, directs neutrons of a given energy at the sample (as was done for the powder diffractometer shown in Fig. 9). After the sample scatters these neutrons in various directions, a second crystal—positioned at a well-defined scattering angle and called the *analyzer*—Bragg reflects only those neutrons that have a particular energy into a suitably placed detector. This type of instrument is called a *three-axis spectrometer* (Fig. 15) because there are three centers (monochromator, sample, and analyzer) at which the scattering angles can be altered. Such instruments are the workhorses for the measurement of phonons at reactors.

Three-axis spectrometers have contributed prolifically to the various scientific problems studied by neutron scattering, probably because they are so *inefficient*. At each setting of the spectrometer—corresponding to particular scattering angles at the monochromator, sample, and analyzer—a measurement is made for a single scattering vector, $\mathbf{Q}$, and energy transfer, ϵ. Each measurement usually takes several minutes; a complete scan at a series of values of ϵ may take hours or even days. This inefficiency has advantages, though—it allows the experimenter to concentrate on measuring particular excitations at particular values of $\mathbf{Q}$ and ϵ, and it gives that person time to plan each new measurement in light of the data already accumulated.

The success of three-axis spectrometers leads to an interesting philosophical dilemma. Does materials science by its very nature require for its study an instrument such as a three-axis spectrometer? That is, is there some reason to believe that a majority of interesting and important effects occur, like Bragg scattering, only in a restricted range of values of $\mathbf{Q}$ and ϵ? Or has our understanding of materials actually been hampered because three-axis spectrometers have been so popular and prolific? Have we seen only a part of the truth because three-axis spectrometers can only probe a single scattering vector and energy transfer at one time? Would we learn more if we could make measurements for a wide range of values of $\mathbf{Q}$ and ϵ simultaneously? Of course only the extensive use of alternative types of spectrometers can answer this question. Many of the instruments that are best suited to surveys of neutron scattering for large ranges of scattering vector and energy transfer are located at spallation sources such as the one at LANSCE.

There is no real equivalent of the
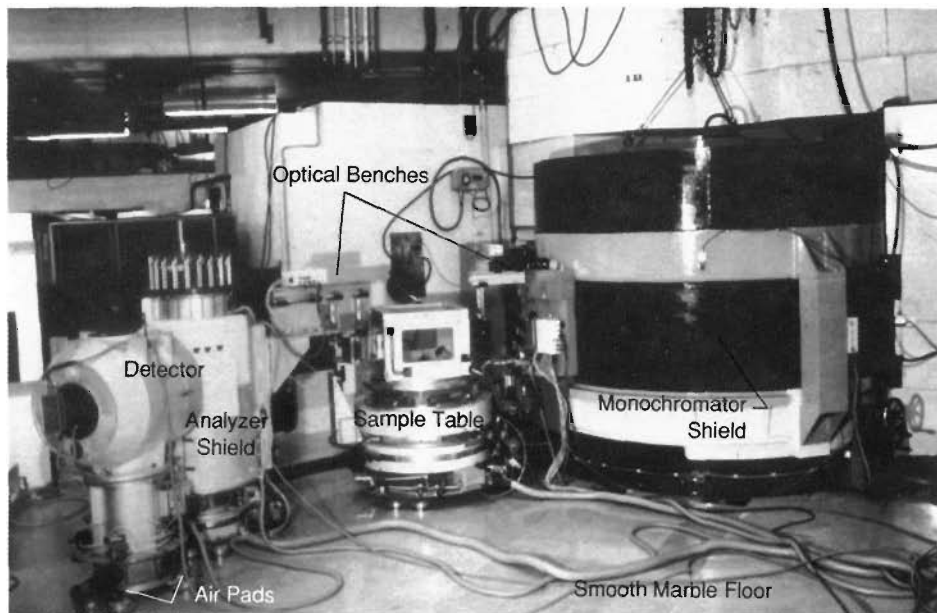
Fig. 15. A three-axis spectrometer built by the author at the Institut Laue Langevin in Grenoble, France. The scattering angles at the monochromator, sample, and analyzer can be varied by moving these connected units on the air pads seen in the photograph. This spectrometer is equipped for polarization analysis. The hollow box-like object on the sample table has current-carrying wires along each edge that can produce a field of about 100 oersteds at any direction on a sample placed at the center of the box. Various spin flippers, diaphragms, and filters are mounted on the optical benches before and after the sample position.

three-axis spectrometer that can be built at a spallation source. Inelastic scattering can, however, be measured in a variety of ways. Perhaps the simplest is to place an analyzing crystal in the scattered neutron beam just as one does with the three-axis machine. This crystal determines the final energy of the neutrons scattered by the sample. Once this energy and the total time of flight from the moderator to the detector are known, the incident energy can also be deduced.

Another method of measuring inelastic scattering at a pulsed spallation source has been used to obtain some of the data discussed by Juergen Eckert and Phil Vergamini (see "Neutrons and Catalysis"). This method uses a filter rather than an analyzing crystal in the scattered neutron beam. The filter allows only neutrons whose energy is less than a certain cutoff value to pass through to a detector behind the filter. Filters of this type can be made, for example, from a block of cooled polycrystalline beryllium that is several centimeters thick. When neutrons impinge on the block, they are scattered just as

they would be from any polycrystalline material. But there is a maximum value of the neutron wavelength beyond which Bragg scattering cannot occur because there are no atomic planes spaced far enough apart to diffract these long-wavelength neutrons. Neutrons with wavelengths greater than the cut-off therefore pass through the filter without being scattered out of the beam. In the case of beryllium, neutrons with wavelengths greater than about 4 angstroms (energies less than about 5 meV) are transmitted. In the Filter-Difference Spectrometer at LANSCE, two filters are used, beryllium and beryllium oxide. The latter material transmits neutrons with energies below 3.7 meV. By subtracting data obtained with the BeO filter from that obtained with the Be filter, we obtain a result that includes only those neutrons with final energies in the narrow window between 3.7 meV and 5 meV, the two filtering energies. This technique allows the energy of the scattered neutrons to be determined accurately. As usual, the total time of flight lets us deduce the incident energy of the neutrons.

The Filter Difference Spectrometer is not well suited for measurements of phonons because the geometry of the instrument makes it inherently difficult to determine the scattering vector, $\mathbf{Q}$, to a high degree of accuracy. This is an advantage when one is measuring incoherent inelastic scattering, however, because the energy transfer ϵ is often independent of $\mathbf{Q}$, and one may sum scattered intensities for many values of $\mathbf{Q}$, thereby increasing the statistical accuracy of the data obtained. This summation is accomplished automatically with the Filter Difference Spectrometer at LANSCE.

The final method of measuring inelastic scattering at a spallation source—a method that *does* determine the scattering vector accurately—makes use of a so-called chopper spectrometer. The chopper, which can be thought of as a short (20-centimeter) pipe rotating about an axis perpendicular to its length, is placed in the neutron beam ahead of the scattering sample. If the pipe is rotating at a frequency that is an integral multiple of that of the pulsed neutron source, it briefly becomes aligned with

the neutron beam at the same time during each neutron pulse from the moderator. Because the chopper is usually several meters from the neutron moderator, the fast neutrons in each pulse arrive at the chopper ahead of their slower brethren. Only those neutrons that arrive at the chopper when it is open—that is, aligned with the beam—get through. Thus, the chopper selects neutrons in a small band of velocities and allows them to impinge on the sample. Neutrons outside this band will arrive either too late or too early at the chopper and will be stopped. The chopper thus determines the wave vector of the neutrons incident on the sample, whereas a measurement of the total time of flight allows the wave vector of the scattered neutrons to be calculated as well.

A great advantage of chopper spectrometers is that neutron detectors can be placed at many different scattering angles simultaneously, allowing scattering to be recorded at many values of Q and ϵ. The disadvantage is that the experimenter is inundated with data and must rely heavily on computers to reduce the massive array of numbers to something comprehensible.

In fact, massive amounts of data are the norm for spectrometers at spallation sources. At each detector, a series of Q and ϵ values are measured that correspond to a full range of differing flight times of the detected neutrons. One automatically obtains these values whether one wants the flood of data or not. In short, a three-axis spectrometer at a reactor source is a rifle, whereas its equivalent at a spallation source is a shotgun. Which source is more efficient for a given experiment really depends on what type of information one wants—a single bull's-eye or a barn door full of interesting holes! More seriously, we can obtain a detailed knowledge of the scattering law for a few values of Q and ϵ at a reactor source and a more extended picture covering a wide range of these variables at a spallation source.

Magnetic Scattering

So far we have discussed only the interaction between neutrons and atomic nuclei. But there is another interaction between neutrons and matter—one that results from the fact that a neutron has a magnetic moment (Fig. 2). Just as two bar magnets either attract or repel one another, the neutron experiences a force of magnetic origin whenever it passes close to another magnetic particle, such as an electron in matter.

Most electrons in atoms or in matter are paired so that the magnetic moment of one electron cancels that of its partner. Occasionally, however, not all the outer, or binding, electrons are paired in a particular compound, and neutrons are scattered by the resulting magnetic moments. Diffraction experiments, similar to those described earlier, can be used to measure the density of such unpaired electrons between the atoms of a solid.

Ferromagnetic materials, such as iron, are magnetic because the moments of their unpaired electrons tend to align spontaneously. For many purposes, such materials behave as if a small magnetic moment were located at each atomic site with all the moments pointed in the same direction. These moments give rise to Bragg diffraction of neutrons in the same manner as the nuclear interaction. Because nuclear and magnetic interactions experienced by the neutron are of similar magnitude, the corresponding Bragg reflections are also of comparable intensity.

One difference between the two types of scattering, however, is that magnetic scattering, unlike its nuclear counterpart, is not isotropic. The magnetic interaction has a dipolar nature, which can easily be observed by bringing two bar magnets close to one another. Suppose the two magnets are parallel with their north poles pointing upward. If one magnet is above the other, unlike poles will be close, and the magnets will attract; if they are side by side, like poles will be close, and the magnets will repel. For neutrons, the dipolar nature of magnetic interaction means that only the component of the sample's magnetization that is perpendicular to the scattering vector, Q, is effective in scattering neutrons. Neutron scattering is therefore sensitive to the direction of magnetization in a material as well as to its spatial distribution.

The anisotropic nature of the magnetic interaction can be used to separate nuclear and magnetic Bragg peaks in ferromagnets, for which both types of Bragg peaks occur at the same values of Q. If the electronic moments can be aligned by an applied magnetic field, magnetic Bragg peaks for which Q is parallel to the induced magnetization vanish, leaving only the nuclear component. On the other hand, an equivalent Bragg peak for which the scattering vector is perpendicular to the field will manifest both nuclear and magnetic contributions.

In an antiferromagnet (a material with unpaired electrons that have an *alternating*, or antiparallel, arrangement), the repeat distance between planes of magnetic moments is twice that of the spacing between corresponding planes of atoms. As a result, Bragg's law is satisfied at scattering angles whose sines are half those for normal Bragg scattering, as well as at the normal angles. Half the magnetic Bragg peaks fall between their nuclear counterparts, and the problem of separating magnetic and nuclear contributions does not arise. Nevertheless, the dipolar character of the magnetic interaction again allows the electronic spin directions to be established. A recent example of this is to be found in the superconducting cuprates—the so-called high-temperature superconductors—some of which are an-

(a)

(b)
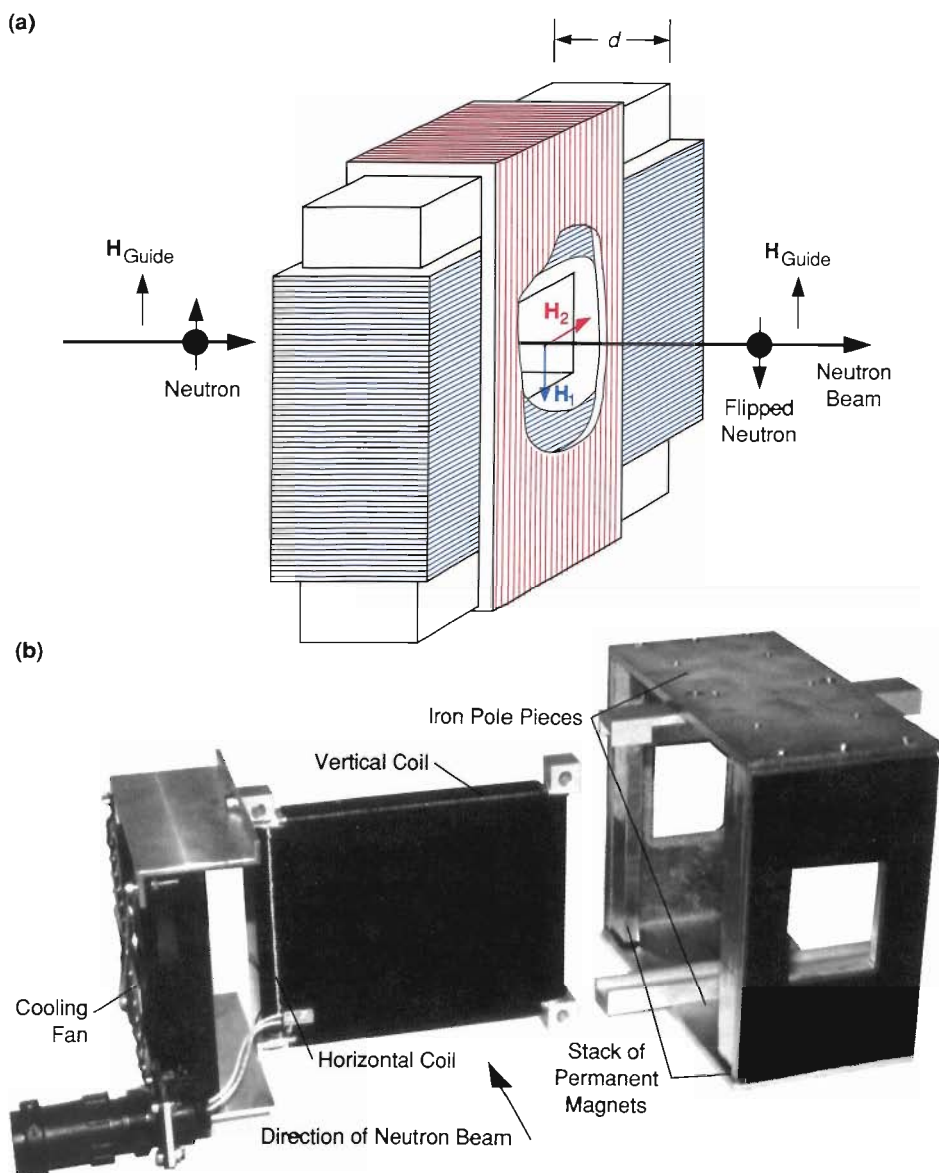
A FLAT-COIL NEUTRON-SPIN FLIPPER

Fig. 16. (a) Schematic diagram of one type of neutron flipper. A direct current in the horizontal coil of aluminum wires (blue) produces a field H_1 inside the device that is equal but opposite to the neutron guide field H_{guide}, effectively canceling that component. The vertical coil (red) produces a field H_2 that is at right angles to the guide field and thus to the moment of the neutron, causing it to precess. The strength of this field and the thickness d of the flipper are chosen so that the neutron precesses exactly 180 degrees during its passage. (b) Photograph of disassembled neutron flipper. The penetrating power of neutrons is apparent in the fact that there is no "window" in the two coils of wire; the neutrons pass on through the aluminum wire unimpeded. The component on the right produces a vertical guide field of about 40 oersteds.

tiferromagnetic when oxygen deficient.

Polarized Neutrons. Usually, a neutron beam contains neutrons with magnetic moments pointing in all directions. If we could measure the number of neutrons with moments parallel and antiparallel to a particular direction—say an applied magnetic field—we would find equal populations. However, various special techniques can generate a *polarized* beam, that is, one with a large fraction of its neutron moments in the same direction. The polarization of such a beam can be maintained by applying a modest magnetic field (a few tens of oersteds) all along the beam. Such a field is called the *guide field*.

There are several ways to polarize neutron beams: Bragg diffraction from suitable magnetized crystals, reflection from magnetized mirrors made of cobalt and iron (CoFe), and transmission through polarized helium-3, for example. Each of these methods aligns the neutron moments parallel or antiparallel to an applied magnetic field. If the neutron moments are parallel to the field, they are said to be *up*; if the moments are antiparallel, they are said to be *down*. An 'up' polarizer will not transmit 'down' neutrons, just as a 'down' polarizer blocks 'up' neutrons. Thus, by placing an 'up' polarizer before and after a scattering sample, the neutron scattering law can be measured for those scattering processes in which the direction of the neutron moment is not changed. To measure the other combinations—such as 'up' neutrons being flipped to 'down' neutrons—requires either a variety of different 'up' and 'down' polarizers or a device called a *flipper*. Because polarizers tend to be expensive, flippers are the practical choice.

A flipper is a device that can change the direction of a neutron moment from up to down or vice versa. This can be done in one of two ways. Either the guide-field direction can be inverted without changing the direction of the neutron moment in space, or the neutron moment can be inverted without altering the direction of the guide field. In either case, the direction of the neutron moment with respect to the field (which is all that counts) has been changed.

An example of the second type of flipper is shown in Fig. 16. It consists of two flat coils of wire wrapped one on top of the other. One of the coils produces a field inside the flipper that is equal and opposite to the guide field, effectively canceling that component, and the other coil produces a field perpendicular to the guide field. Thus, when a neutron enters the flipper, it suddenly experiences a magnetic field that is at right angles to the direction of its mag-

netic moment. In this situation the classical equations that describe the motion of the neutron moment are similar to those of a rotating top that has been pushed by a force from the side and so begins precessing about its original axis of rotation. The neutron does the same thing—its moment starts to precess about the local field direction at a rate known as the *Larmor frequency*, which depends on the magnitude of the field inside the flipper. By choosing the thickness of the flipper and the strength of the field in the second coil appropriately, one can arrange for the neutron moment to rotate precisely 180 degrees during its passage through the flipper. Clearly, if a neutron's moment was up before the flipper, it will be down after the flipper, and vice versa.

Now suppose we have a spectrometer with polarizers before and after the scattering sample. If flippers are inserted on either side of the sample, we can measure all of the neutron scattering laws—up to down, up to up, and so forth—simply by turning the appropriate flipper on or off. This technique, known as *polarization analysis*, is useful because some scattering processes flip the neutron's moment whereas others do not.

Scattering from a sample that is magnetized provides a good example. Magnetic scattering will flip the neutron's moment if the magnetization responsible for the scattering is perpendicular to the guide field used to maintain the neutron polarization. If the magnetization is parallel to the guide field, no flipping occurs. Thus, like the dipolar interaction described earlier, polarization analysis is a technique that helps determine the direction of electronic moments in matter.

Incoherent scattering that arises from the random distribution of nuclear spin states in materials provides another example of the use of polarization analysis. Most isotopes have several spin states, and the scattering cross section

for a nucleus varies with spin state. The random distribution of nuclear spins in the sample gives rise to incoherent scattering of neutrons. It turns out that two-thirds of the neutrons scattered by this incoherent process have their moments flipped, whereas the moments of the remaining third are unaffected. This result is independent of the isotope that is responsible for the scattering and of the direction of the guide field. Although incoherent scattering can also arise if a sample contains a mixture of isotopes of a particular element, neither this second type of incoherent scattering nor coherent nuclear scattering flip the neutron's moment. Polarization analysis thus becomes an essential tool for sorting out these different types of scattering, allowing nuclear coherent scattering to be distinguished from magnetic scattering and spin-incoherent scattering.

Polarization analysis has been particularly useful in the study of magnetic phenomena because it has helped to determine the directions of the magnetic fluctuations responsible for scattering. Without this technique, many of the elegant experiments that have provided confirmation for ideas about nonlinear physics (see "Nonlinear Science—From Paradigms to Practicalities" by David K. Campbell, *Los Alamos Science* No. 15, 1987) could not have been performed. The three-axis spectrometer of Fig. 15, for example, is equipped for polarization analysis.

Magnons. Another important aspect of magnetized materials is the fact that the directions of the atomic moments in a material such as iron can oscillate like the pendulums considered earlier for lattice vibrations. Here again, there is a coupling between magnetization at different atomic sites, and a wave of magnetic oscillations can pass through the material. These magnetic excitations, or *magnons*, are the magnetic analogue of the phonon displacement

waves described earlier. Not surprisingly, magnon frequencies can be measured by inelastic neutron scattering in the same way as phonon frequencies. Since the magnetic oscillations that make up the magnons are perpendicular to the equilibrium direction of the atomic moments, the scattering causes the magnetic moment of the neutrons to be flipped, provided the neutron guide field is parallel to the equilibrium direction of the atomic moments. This, of course, allows one to distinguish between phonons and magnons.

Surface Structure

So far we have described only experiments in which the structure of bulk matter is probed. One may ask whether neutrons can provide any information about the structure of the surfaces of materials. At first sight, one might expect the answer to be a resounding "No!" After all, one of the advantages of neutrons is that they can penetrate deeply into matter without being affected by the surface. Furthermore, because neutrons interact only weakly with matter, large samples are generally required. Because there are far fewer atoms on the surface of a sample than in its interior, it seems unreasonable to expect neutron scattering to be sensitive to surface structure.

In spite of these objections, it turns out that neutrons *are* sensitive to surface structure when they impinge on the surface at sufficiently low angles. In fact, for smooth surfaces, perfect reflection of neutrons occurs for almost all materials at angles of incidence (the angle between the incident beam and the surface) less than a *critical angle*, denoted γ_c. This angle is proportional to the coherent scattering-length density of the material and the neutron wavelength. For a good reflector, such as nickel, the critical angle measured in degrees is about one-tenth of the neutron
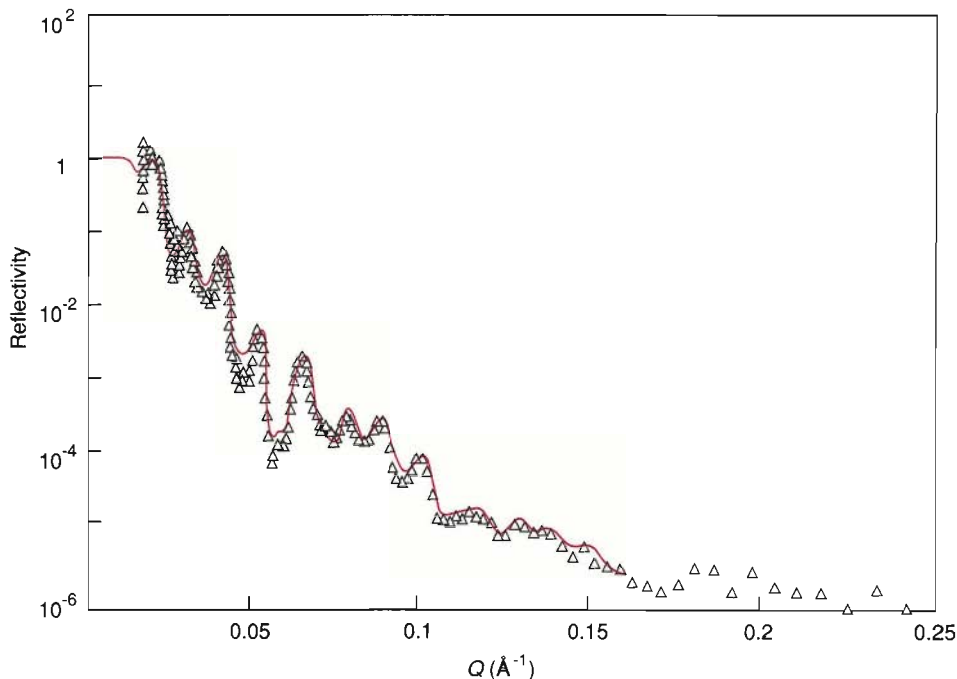
SURFACE REFLECTIVITY MEASUREMENTS

Fig. 17. Neutron reflectivity as a function of Q (= $4\pi \sin\theta/\lambda$) for a 1500-angstrom thick diblock copolymer (polystyrene-poly-methylmethacrylate) multilayer deposited on a silicon substrate. The solid line represents calculated reflectivity for the data shown. The calculation was performed by Tom Russell, IBM Almaden Research Labs.

wavelength measured in angstroms— it is well under a degree for thermal neutrons. As the angle of incidence increases above the critical angle, less and less of the incident neutrons are reflected by the surface. In fact, reflectivity, which measures the fraction of neutrons reflected from the surface, obeys the same law, discovered by Fresnel, that applies to the reflection of light: reflectivity decreases as the fourth power of the angle of incidence at sufficiently large grazing angles.

However, Fresnel's law applies to reflection of radiation from the smooth, flat surface of a homogeneous material. If the material is inhomogeneous and there is a variation of the scattering-length density perpendicular to the surface, the neutron reflectivity, measured as a function of the angle of incidence, shows a more complicated behavior. By keeping the reflection angle, θ, small, neutron reflectometry can be used to probe density variations in the surface to depths of a few thousand angstroms with a resolution of a few angstroms.

Most of today's technical gadgets

are either painted or coated in some fashion to prevent corrosion or wear. Reflectometry can often provide useful information about such protective layers. Figure 17, for example, shows the reflectivity, measured on the LANSCE Surface Profile Analysis Reflectometer (SPEAR), from a 1500-angstrom layer of diblock copolymer (polystyrene-polymethylmethacrylate) multilayer deposited on a silicon substrate. The spacing of the undulations in this result provides a direct measure of the average thickness of the polymer layers in the film. When the detailed shape of the reflectivity profile is compared with theoretical predictions, the density and thickness of the polymer layers, as well as the thickness of the interface between layers, can be deduced.

Neutron reflectometry is a relatively new technique. It is also one ideally suited to spallation sources. In the next few years I expect the method to provide new information on subjects as diverse as the recycling of polymers, magnetic recording media, and the cleanup of oil spills. For someone like me who

has been associated with neutron scattering for more than twenty years, the birth of this new technique is a happy event. It means that there are still qualitatively new ways in which neutrons can help unravel the complex structures of the materials on which we depend. ∎

The Mathematical Foundations of Neutron Scattering

In 1954 Van Hove derived a general expression for the intensity, $I(\mathbf{Q}, \epsilon)$, of neutrons scattered by any assembly of nuclei. His result makes use of Fermi's observation that the actual interaction between a neutron and a nucleus may be replaced by an effective potential that is much weaker than the actual interaction. This *pseudo*-potential causes the same scattering as the actual interaction but is weak enough to be used in the perturbation expansion derived by Max Born. The Born approximation says that the probability of an incident plane wave of wave vector $\mathbf{k}$ being scattered by a weak potential $V(\mathbf{r})$ to become an outgoing plane wave with wave vector $\mathbf{k}'$ is proportional to

$$\left| \int e^{i\mathbf{k}\cdot\mathbf{r}} V(\mathbf{r}) e^{-i\mathbf{k}'\cdot\mathbf{r}}\, d^3r \right|^2 = \left| \int e^{i\mathbf{Q}\cdot\mathbf{r}} V(\mathbf{r})\, d^3r \right|^2, \tag{1}$$

where the integration is over the volume of the scattering sample. (We should note that even though individual nuclei scatter spherically, $V(\mathbf{r})$ represents the potential due to the entire sample, and the resulting disturbance for the assembly of atoms is a plane wave.)

The potential to be used in Eq. 1 is Fermi's pseudo-potential, which, for a single nucleus, is given by $b_j \delta(\mathbf{r} - \mathbf{r}_j)$, where b_j is the scattering length of a nucleus labeled j located at position $\mathbf{r}_j$ and δ is a Dirac delta function that is zero unless the position vector $\mathbf{r}$ coincides with $\mathbf{r}_j$. Thus, for an assembly of nuclei, such as a crystal, the potential $V(\mathbf{r})$ is the sum of individual neutron-nuclei interactions:

$$V(\mathbf{r}) = \sum_j b_j \delta(\mathbf{r} - \mathbf{r}_j), \tag{2}$$

where the summation is over all the nuclear sites in the crystal.

Using Eqs. 1 and 2, Van Hove was able to show that the scattering law—that is, the number of neutrons scattered per incident neutron—can be written as

$$I(\mathbf{Q}, \epsilon) = \frac{1}{h} \frac{k'}{k} \sum_{j,k} b_j b_k \int_{-\infty}^{\infty} \langle e^{-i\mathbf{Q}\cdot\mathbf{r}_k(0)} e^{i\mathbf{Q}\cdot\mathbf{r}_j(t)} \rangle e^{-i\epsilon t}\, dt. \tag{3}$$

Note that the sum here is over *pairs* of nuclei j and k and that the nucleus labeled j is at position $\mathbf{r}_j(t)$ at time t, whereas the nucleus labeled k is at position $\mathbf{r}_k(0)$ at time $t = 0$. The angular brackets $\langle \ldots \rangle$ denote an average over all possible starting times for observations of the system, which is equivalent to an average over all the possible thermodynamic states of the sample.

The position vectors $\mathbf{r}_j$ in Eq. 3 are quantum-mechanical operators that have to be manipulated carefully. Nevertheless, it is instructive to ignore this subtlety and treat the equation as if it described a system obeying classical mechanics because such an approach clarifies the physical meaning of the equation. The sum over atomic sites in Eq. 3 can then be rewritten as

$$\sum_{j,k} b_j b_k \langle e^{-i\mathbf{Q}\cdot[\mathbf{r}_k(0) - \mathbf{r}_j(t)]} \rangle = \sum_{j,k} b_j b_k \int_{-\infty}^{\infty} \delta(\langle \mathbf{r} - [\mathbf{r}_k(0) - \mathbf{r}_j(t)] \rangle) e^{-i\mathbf{Q}\cdot\mathbf{r}}\, d^3r, \tag{4}$$

in which the Dirac delta function appears again, this time in terms of $\mathbf{r}$ and a difference vector between the position of nucleus j at time t and that of nucleus k at time zero.

Let us suppose for the moment that the scattering lengths of all the atoms in our sample are the same ($b_j = b_k = b$). In this case, the scattering lengths in Eq. 4 can be removed from the summation, and the right side becomes

$$Nb^2 \int_{-\infty}^{\infty} G(\mathbf{r}, t) e^{-i\mathbf{Q}\cdot\mathbf{r}} \, d^3r, \tag{5a}$$

where

$$G(\mathbf{r}, t) = \frac{1}{N} \sum_{j,k} \langle \delta(\mathbf{r} - [\mathbf{r}_k(0) - \mathbf{r}_j(t)]) \rangle \tag{5b}$$

and N is the number of atoms in the sample. The delta function in the definition of $G(\mathbf{r}, t)$ is zero except when the position of an atom k at time zero and the position of atom j at time t are separated by the vector $\mathbf{r}$. Because the delta functions are summed over all possible pairs of atoms to obtain $G(\mathbf{r}, t)$, this function is equal to the probability of an atom being at the origin of a coordinate system at time zero *and* an atom being at position $\mathbf{r}$ at time t. $G(\mathbf{r}, t)$ is generally referred to as the *time-dependent pair-correlation function* because it describes how the correlation between two particles evolves with time.

Van Hove's neutron-scattering law (Eq. 3) can now be written as

$$I(\mathbf{Q}, \epsilon) = \frac{Nb^2}{h} \frac{k'}{k} \int_{-\infty}^{\infty} G(\mathbf{r}, t) e^{-i\mathbf{Q}\cdot\mathbf{r}} e^{-i\epsilon t} \, d^3r \, dt, \tag{6}$$

which allows us see that $I(\mathbf{Q}, \epsilon)$ is proportional to the space and time Fourier transforms of the time-dependent pair-correlation function. This general result gave a unified description for all neutron-scattering experiments and thus provided the framework for defining neutron scattering as a field.

As discussed in the text of the main article, this fact—that $I(\mathbf{Q}, \epsilon)$ is simply the Fourier transform of a function that gives the probability of finding two atoms a certain distance apart—is responsible for the power of neutron scattering. By inverting Eq. 6, information about both structure and dynamics of condensed matter may be obtained from the scattering law.

Coherent and Incoherent Scattering

Even for a sample made up of a single isotope, all of the scattering lengths that appear in Eq. 3 will not be equal. This is because the scattering length of a nucleus depends on its spin state, and most isotopes have several spin states. Generally, however, there is no correlation between the spin of a nucleus and its position in a sample of matter. For this reason, the scattering lengths that appear in Eq. 3 can be averaged over the nuclear spin states without affecting the thermodynamic average (denoted by the angular brackets).

Two spin averages come into play: the average value of b ($\overline{b}$) and the average value of b^2 ($\overline{b^2}$). In terms of these quantities, the sum in Eq. 3 can be averaged over the nuclear spins to give

$$\sum_{j,k} \overline{b_j b_k} A_{jk} = \sum_{j,k} (\overline{b})^2 A_{jk} + \sum_{j} (\overline{b^2} - (\overline{b})^2) A_{jj}, \tag{7}$$

where A_{jk} is shorthand for the integral in Eq. 3. The first term on the right side of Eq. 7 represents the so-called *coherent scattering*, whereas the second represents

the *incoherent scattering*. Thus, we can define the coherent and incoherent scattering lengths as

$$b_{\text{coh}} = \overline{b} \text{ and}$$
$$b_{\text{inc}} = \sqrt{\overline{b^2} - (\overline{b})^2}. \tag{8}$$

The expression for the coherent scattering law is a sum over both j and k and thus involves correlations between the position of an atom j at time zero and the position of a second atom k at time t. Although j and k are occasionally the same atom, in general they are not the same because the number N of nuclei in the sample is large. We can thus say that coherent scattering essentially describes interference between waves produced by the scattering of a single neutron from all the nuclei in a sample. The intensity for this type of scattering varies strongly with the scattering angle.

Incoherent scattering, on the other hand, involves correlations between the position of an atom j at time zero and the position of the *same* atom at time t. Thus, in incoherent scattering, the scattered waves from different nuclei do not interfere with each other. For this reason, incoherent scattering provides a good method of examining processes in which atoms diffuse. In most situations, the incoherent scattering intensity is isotropic; that is, it is the same for any scattering angle. This effect often allows incoherent scattering to be ignored when observing coherent scattering because the incoherent effects just add intensity to a structureless background.

The values of the coherent and incoherent scattering lengths for different elements and isotopes do not vary in any obviously systematic way throughout the periodic table. For example, hydrogen has a large incoherent scattering length (25.18 fermis) and a small coherent scattering length (−3.74 fermis). Deuterium, on the other hand, has a small incoherent scattering length (3.99 fermis) and a relatively large coherent scattering length (6.67 fermis). As mentioned in the main article, the difference between the coherent scattering lengths of hydrogen and deuterium is the basis of an isotopic-labeling technique, called *contrast matching*, that is especially important in applications of neutron scattering to structural biology and polymer science.

Diffraction

One of the important applications of Van Hove's equation (Eq. 3) is the scattering law for diffraction, which we develop here for a crystal containing a single isotope. Even though diffraction is predominantly an elastic scattering process ($\epsilon = 0$), neutron diffractometers actually integrate over the energies of scattered neutrons. Thus, rather than setting $\epsilon = 0$ in Eq. 3 to calculate the diffracted intensity, we integrate the equation over ϵ. This procedure ensures that the effect of atomic vibrations is included in the diffraction cross section. The integral of Eq. 3 over ϵ gives another Dirac delta function, $\delta(t)$, that tells us that the pair correlation function, $G(\mathbf{r}, t)$, has to be evaluated at $t = 0$ for diffraction. The result, for a crystal containing a single isotope, is

$$I(\mathbf{Q}) = b_{\text{coh}}^2 \sum_{j,k} \langle e^{i\mathbf{Q} \cdot (\mathbf{r}_j - \mathbf{r}_k)} \rangle, \tag{9}$$

where the atomic positions $\mathbf{r}_j$ and $\mathbf{r}_k$ are evaluated at the same instant.

If the atoms in a sample were truly stationary, the thermodynamic averaging brackets could be removed from Eq. 9 because $\mathbf{r}_j$ and $\mathbf{r}_k$ would be constant. In reality the atoms oscillate about their equilibrium positions and only spend a fraction of their time at these positions. When this is taken into account, the thermodynamic average introduces another factor, called the Debye-Waller factor, and Eq. 9 then becomes

$$I(\mathbf{Q}) = b_{\text{coh}}^2 \sum_{j,k} e^{i\mathbf{Q} \cdot (\mathbf{r}_j - \mathbf{r}_k)} e^{-\frac{1}{2}Q^2 \langle u^2 \rangle} \equiv S(\mathbf{Q}), \tag{10}$$

where $\langle u^2 \rangle$ is the average of the square of the displacement of an atom from its equilibrium position and diffracted intensity is now also called $S(\mathbf{Q})$, the structure factor. This equation is the basis of any crystallographic analysis of neutron-diffraction data.

Small-Angle Scattering.

An important simplification of Eq. 3 occurs when the scattering angle is small. This approximation leads to the formula for one of the most popular neutron-scattering techniques—SANS, or small-angle neutron scattering.

Although Eq. 3 correctly describes neutron scattering at *any* scattering angle, when the magnitude of $\mathbf{Q}$ is very small compared to a typical interatomic distance, the exponential factors in Eq. 3 do not vary much from atom to atom, and the sum over the atomic sites may be replaced by an integral. As a result, the small-angle scattering law for coherent, elastic scattering from an assembly of "objects" (such as those depicted in Fig. 13 in the main text) can be written

$$I(\mathbf{Q}) = \left| \int b(\mathbf{r})e^{-i\mathbf{Q}\cdot\mathbf{r}}d^3r \right|^2 , \tag{11}$$

where $b(\mathbf{r})$ is the scattering-length density and the integral extends over the entire sample. To calculate $b(\mathbf{r})$ for a large molecule, for example, we simply add up the coherent scattering lengths of the atoms in the molecule and divide by the molecular volume. Equation 11 is essentially a coarse-grained version of the "truth" given by Eq. 3 and is valid only when Q is small. However, it is the basic analytic tool of small-angle scattering. ∎

Roger Pynn was born and educated in England. He received his M. A. from the University of Cambridge in 1966 and his Ph.D. in neutron scattering, also from the University of Cambridge, in 1969. He was a Royal Society European Fellow to Sweden in 1970; he did two years of postdoctoral research in Norway; and then he was an associate physicist for two years at Brookhaven National Laboratory. After spending eleven years at the world's leading center for neutron scattering, the Institut Laue Langevin in Grenoble, France, he was appointed as the Director of the Manuel Lujan, Jr. Neutron Scattering Center at Los Alamos.

Acknowledgments

Further Reading

G. L. Squires. 1978. *Introduction to Thermal Neutron Scattering.* Cambridge: Cambridge University Press.

C. G. Windsor. 1981. *Pulsed Neutron Scattering.* London: Taylor and Francis.

Physics Today, January 1985. A special issue on neutron scattering.

The Mathematical Foundations
of Neutron Scattering

In 1954 Van Hove derived a general expression for the intensity, $I(\mathbf{Q}, c)$, of neutrons scattered by any assembly of nuclei. His result makes use of Fermi observation that the actual interaction between a neutron and a nucleus may be replaced by an effective potential that is much weaker than the actual interaction. This pseudo-potential causes the same scattering as the actual interaction but is weak enough to be used in the perturbation expansion derived by Max Born. The Born approximation says that the probability of an incident plane wave of wave vector k being scattered by a weak potential V (r) to become an outgoing plane wave with wave vector $\mathbf{k'}$ is proportional to

$$\left| \int e^{i\mathbf{k}\cdot\mathbf{r}} V(\mathbf{r}) e^{-i\mathbf{k'}\cdot\mathbf{r}} d^3r \right|^2 = \left| \int e^{i\mathbf{Q}\cdot\mathbf{r}} V(\mathbf{r}) d^3r \right|^2, \tag{1}$$

where the integration is over the volume of the scattering sample. (We should note that even though individual nuclei scatter spherically, $V(\mathbf{r})$ represents the potential due to the entire sample, and the resulting disturbance for the assembly of' atoms is a plane wave.)

The potential to be used in Eq. 1 is Fermi's pseudo-potential. which, for a single nucleus, is given by $b_j \delta(\mathbf{r} - \mathbf{r}_j)$, where b_j is the scattering length of a nucleus labeled j located at position $\mathbf{r}_j$ and δ is a Dirac delta function that is zero unless the position vector $\mathbf{r}$ coincides with $\mathbf{r}_j$. Thus, for an assembly of nuclei, such as a crystal, the potential V $(\mathbf{r})$ is the sum of individual neutron-nuclei interactions:

$$V(\mathbf{r}) = \sum_j b_j \delta(\mathbf{r} - \mathbf{r}_j). \tag{2}$$

where the summation is over all the nuclear sites in the crystal.

Using Eqs. 1 and 2, Van Hove was able to show that the scattering law—that is, the number of neutrons scattered per incident neutron-can be written as

$$I(\mathbf{Q}, c) = \frac{1}{h}\frac{k'}{k} \sum_{j,k} b_j b_k \int_{-\infty}^{\infty} \langle e^{-i\mathbf{Q}\cdot\mathbf{r}_k(0)} e^{i\mathbf{Q}\cdot\mathbf{r}_j(t)} \rangle e^{-i\epsilon t} \, dt. \tag{3}$$

Note that the sum here is over *pairs* of nuclei j and k and that the nucleus labeled j is at position r;(f) at time t, whereas the nucleus labeled k is at position $\mathbf{r}_k(\mathbf{0})$ at time $t = 0$. The angular brackets (. .) denote an average over all possible starting times for observations of the system, which is equivalent to an average over all the possible thermodynamic states of the sample.

The position vectors $\mathbf{r}_j$ in Eq. 3 are quantum-mechanical operators that have to be manipulated carefully. Nevertheless, it is instructive to ignore this subtlety and treat the equation as if it described a system obeying classical mechanics because such an approach clarifies the physical meaning of the equation. The sum over atomic sites in Eq. 3 can then be rewritten as

$$\sum_{j,k} b_j b_k \langle e^{-i\mathbf{Q}\cdot[\mathbf{r}_k(0)-\mathbf{r}_j(t)]} \rangle = \sum_{j,k} b_j b_k \int_{-\infty}^{\infty} \delta\langle(\mathbf{r} - [\mathbf{r}_k(0) - \mathbf{r}_j(t)])\rangle e^{-i\mathbf{Q}\cdot\mathbf{r}} d^3r, \tag{4}$$

in which the Dirac delta function appears again, this time in terms of **r** and a difference vector between the position of nucleus j at time t and that of nucleus k at time zero.

Let us suppose for the moment that the scattering lengths of all the atoms in our sample are the same $(b_j = b_k = b)$. In this case, the scattering lengths in Eq. 4 can be removed from the summation, and the right side becomes

$$Nb^2 \int_{-\infty}^{\infty} G(\mathbf{r}, t) e^{-i\mathbf{Q} \cdot \mathbf{r}} \, d^3 r, \tag{5a}$$

where

$$G(\mathbf{r}, t) = \frac{1}{N} \sum_{j,k} \langle \delta(\mathbf{r} - [\mathbf{r}_k(0) - \mathbf{r}_j(t)]) \rangle \tag{5b}$$

and N is the number of atoms in the sample. The delta function in the definition of $G(r, t)$ is zero except when the position of an atom k at time zero and the position of atom j at time t are separated by the vector **r**. Because the delta functions are summed over all possible pairs of atoms to obtain G (**r, t),** this function is equal to the probability of an atom being at the origin of a coordinate system at time zero *and* an atom being at position **r** at time t. G (**r**, $t)$ is generally referred to as the *time-dependent pair-correlation function* because it describes how the correlation between two particles evolves with time.

Van Hove's neutron-scattering law (Eq. 3) can now be written as

$$I(\mathbf{Q}, \epsilon) = \frac{Nb^2}{h} \frac{k'}{k} \int_{-\infty}^{\infty} G(\mathbf{r}, t) e^{-i\mathbf{Q} \cdot \mathbf{r}} e^{-i\epsilon t} \, d^3 r \, dt, \tag{6}$$

which allows us see that $I(\mathbf{Q}, \epsilon)$ is proportional to the space and time Fourier transforms of the time-dependent pair-correlation function. This general result gave a unified description for all neutron-scattering experiments and thus provided the framework for defining neutron scattering as a field.

As discussed in the text of the main article, this fact—that $I(\mathbf{Q}, \epsilon)$ is simply the Fourier transform of a function that gives the probability of finding two atoms a certain distance apart-is responsible for the power of neutron scattering. By inverting Eq. 6, information about both structure and dynamics of condensed matter may be obtained from the scattering law.

Coherent and Incoherent Scattering

Even for a sample made up of a single isotope, all of the scattering lengths that appear in Eq. 3 will not be equal. This is because the scattering length of a nucleus depends on its spin state, and most isotopes have several spin states. Generally, however, there is no correlation between the spin of a nucleus and its position in a sample of matter. For this reason, the scattering lengths that appear in Eq. 3 can be averaged over the nuclear spin states without affecting the thermodynamic average (denoted by the angular brackets).

Two spin averages come into play: the average value of b $(\bar{b})$ and the average value of $b^2 (\overline{b^2})$. In terms of these quantities, the sum in Eq. 3 can be averaged over the nuclear spins to give

$$\sum_{j,k} \overline{b_j b_k} A_{jk} = \sum_{j,k} (\bar{b})^2 A_{jk} + \sum_{j} (\overline{b^2} - (\bar{b})^2) A_{jj}. \tag{7}$$

where A_{jk} is shorthand for the integral in Eq. 3. The first term on the right side of Eq. 7 represents the so-called *coherent scattering,* whereas the second represents

the *incoherent scattering.* Thus, we can define the coherent and. incoherent scattering lengths as

$$b_{\text{coh}} = \bar{b} \text{ and}$$

$$b_{\text{inc}} = \sqrt{\overline{b^2} - (\bar{b})^2}.$$ (8)

The expression for the coherent scattering law is a sum over both j and k and thus involves correlations between the position of an atom j at time zero and the position of a second atom k at time t. Although j and k are occasionally the same atom, in general they are not the same because the number N of nuclei in the sample is large. We can thus say that coherent scattering essentially describes interference between waves produced by the scattering of a single neutron from all the nuclei in a sample. The intensity for this type of scattering varies strongly with the scattering angle.

Incoherent scattering, on the other hand, involves correlations between the position of an atom j at time zero and the position of the *same* atom at time t. Thus, in incoherent scattering, the scattered waves from different nuclei do not interfere with each other. For this reason, incoherent scattering provides a good method of examining processes in which atoms diffuse. In most situations, the incoherent scattering intensity is isotropic; that is, it is the same for any scattering angle. This effect often allows incoherent scattering to be ignored when observing coherent scattering because the incoherent effects just add intensity to a structureless background.

The values of the coherent and incoherent scattering lengths for different elements and isotopes do not vary in any obviously systematic way throughout the periodic table. For example, hydrogen has a large incoherent scattering length (25.18 fermis) and a small coherent scattering length (–3.74 fermis). Deuterium, on the other hand, has a small incoherent scattering length (3.99 fermis) and a relatively large coherent scattering length (6.67 fermis). As mentioned in the main article, the difference between the coherent scattering lengths of hydrogen and deuterium is the basis of an isotopic-labeling technique, called *contrast matching,* that is especially important in applications of neutron scattering to structural biology and polymer science.

Diffraction

One of the important applications of Van Hove's equation (Eq. 3) is the scattering law for diffraction, which we develop here for a crystal containing a single isotope. Even though diffraction is predominantly an elastic scattering process ($\epsilon = 0$), neutron diffractometers actually integrate over the energies of scattered neutrons. Thus, rather than setting $\epsilon = 0$ in Eq. 3 to calculate the diffracted intensity, we integrate the equation over ϵ. This procedure ensures that the effect of atomic vibrations is included in the diffraction cross section. The integral of Eq. 3 over ϵ gives another Dirac delta function, $\delta(t)$, that tells us that the pair correlation function, $G(\mathbf{r}, t)$, has to be evaluated at $t = 0$ for diffraction. The result, for a crystal containing a single isotope, is

$$\Gamma(\mathbf{Q}) = b_{\text{coh}}^2 \sum_{j,k} \langle e^{i\mathbf{Q}\cdot(\mathbf{r}_j - \mathbf{r}_k)} \rangle,$$ (9)

where the atomic positions $\mathbf{r}_j$ and $\mathbf{r}_k$ are evaluated at the same instant.

If the atoms in a sample were truly stationary, the thermodynamic averaging brackets could be removed from Eq. 9 because $\mathbf{r}_j$ and $\mathbf{r}_k$ would be constant. In reality the atoms oscillate about their equilibrium positions and only spend a fraction of their time at these positions, When this is taken into account, the thermodynamic average introduces another factor, called the Debye-Wailer factor, and Eq. 9 then becomes

$$\Gamma(\mathbf{Q}) = b_{\text{coh}}^2 \sum_{j,k} e^{i\mathbf{Q}\cdot(\mathbf{r}_j - \mathbf{r}_k)} e^{-\frac{1}{2}Q^2 \langle u^2 \rangle} \equiv S(\mathbf{Q}),$$ (10)

where $\langle u^2 \rangle$ is the average of the square of the displacement of an atom from its equilibrium position and diffracted intensity is now also called $S\,(Q)$, the structure factor. This equation is the basis of any crystallographic analysis of neutron-diffraction data.

Small-Angle Scattering.

An important simplification of Eq. 3 occurs when the scattering angle is small. This approximation leads to the formula for one of the most popular neutron-scattering techniques—SANS, or small-angle neutron scattering.

Although Eq. 3 correctly describes neutron scattering at *any* scattering angle, when the magnitude of Q is very small compared to a typical interatomic distance, the exponential factors in Eq. 3 do not vary much from atom to atom, and the sum over the atomic sites may be replaced by an integral. As a result, the small-angle scattering law for coherent, elastic scattering from an assembly of "objects" (such as those depicted in Fig. 13 in the main text) can be written

$$ I(\mathbf{Q}) = \left| \int b(\mathbf{r}) e^{-i\mathbf{Q}\cdot\mathbf{r}} d^3 r \right|^2 , \qquad (1\ ) $$

where $b\,(\mathbf{r})$ is the scattering-length density and the integral extends over the entire sample. To calculate $b\,(\mathbf{r})$ for a large molecule, for example, we simply add up the coherent scattering lengths of the atoms in the molecule and divide by the molecular volume. Equation 11 is essentially a coarse-grained version of the "truth" given by Eq. 3 and is valid only when Q is small. However, it is the basic analytic tool of small-angle scattering. ∎

Roger Pynn was born and educated in England. He received his M.A. from the UnIversity of Cambridge in 1966 and his Ph.D. in neutron scatterIng, also from the University of Cambridge, in 1969. He was a Royal Society European Fellow to Sweden in 1970; he did two years of postdoctoral research in Norway; and then he was an associate physicist for two years at Brookhaven National Laboratory. After spending eleven years at the world's leading center for neutron scattering, the Institut Laue Langevin in Grenoble, France, he was appointed as the Director of the Manuel Lujan, Jr. Neutron Scattering Center at Los Alamos.

Further Reading

G. L. Squires. 1978. *Introduction to Thermal Neutron Scattering*. Cambridgc: Cambridge University Press.

C. G. Windsor. 1981. *Pulsed Neutron Scattering*. London: Taylor and Francis.

Physics Today, January 1985. A special issue on neutron Scattering.

Roger Pynn describes the scattering of low-energy neutrons (red) as they pass near atoms (blue) on a cubic lattice.
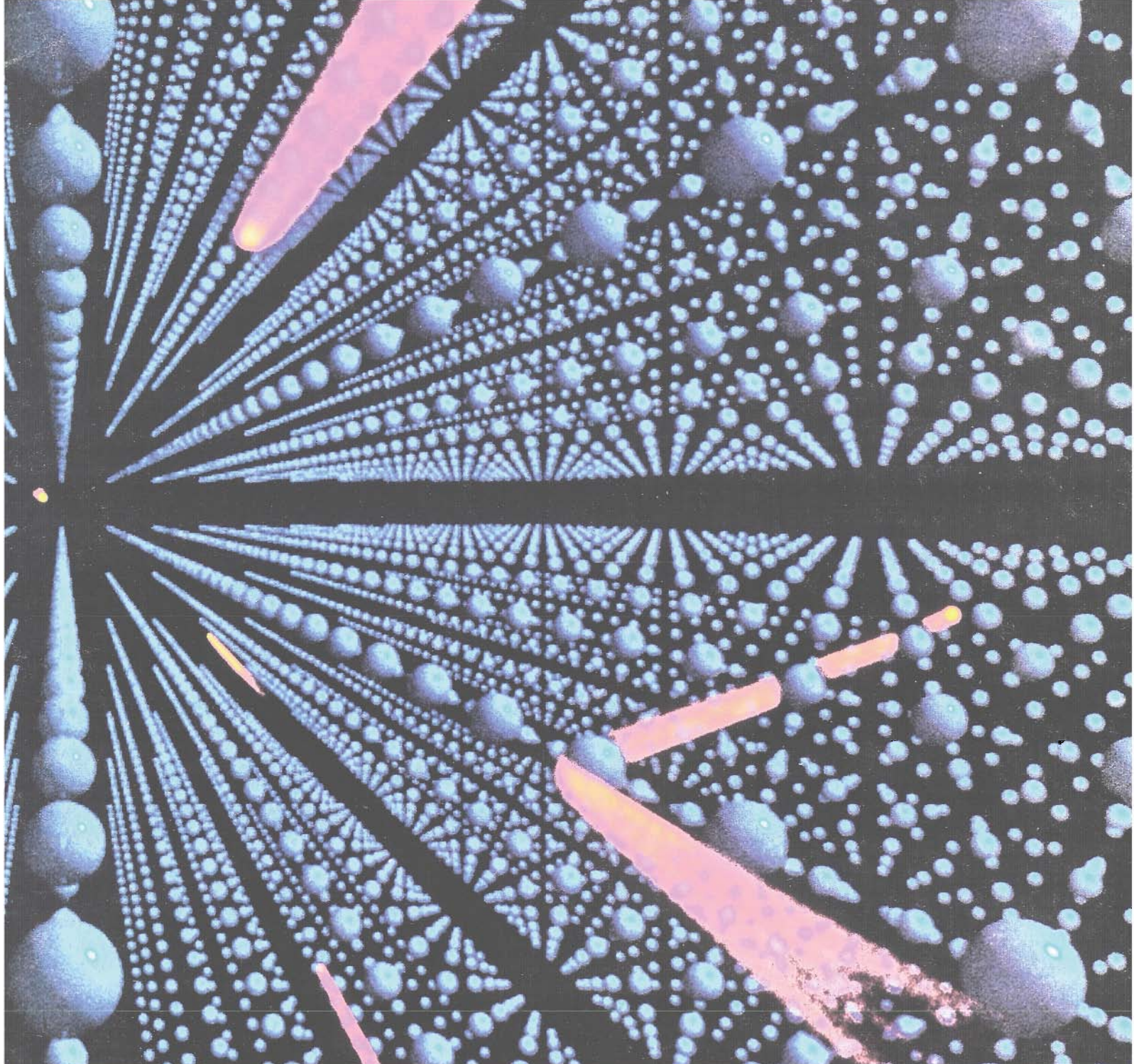
Putting Neutrons in Perspective

An Interview with Roger Pynn

*I*magine alchemists struggling in the dark with materials they do not understand; see modern industrialists trying to control the properties of polymers, colloids, and gels; think of a theorist grasping for the essentials of complicated nonlinear phenomena. Do you feel you need some solid ground? Neutrons traveling through matter see mostly empty space, but their few interactions with nuclei begin to show us the structures of materials—concrete information that can crystallize our questions and theories into a framework of future discovery. Relating the basic properties of materials to their structures transforms alchemy to technology and binds industrial research to basic science. This is the world of neutron scattering.

As director of the Los Alamos Neutron Scattering Center, Roger Pynn is caught up in every aspect of this world. When the user rooms at LANSCE are jammed with biologists, chemists, physicists, and industrial researchers, he jumps from question to question and field to field in only the time it takes him to hurry through the halls. (There were times when we thought he might write all the articles for this issue instead of only three.) Even his own ongoing research projects range from phase transitions and surface phenomena to instrument design and data analysis.

Given the present state of neutron scattering in the United States, however, Roger's most important work may be as LANSCE's ambassador to the larger, more political world. In the last weeks of 1989, one hundred participants in a condensed-matter physics conference petitioned Presidential Science Adviser Allan Bromley for a new commitment to funding for neutron-scattering research. Citing aging facilities and a lack of young scientists entering the field, these researchers warned that the United States neutron-scattering effort may lose irretrievable ground to the thriving community in Europe. We inter-

viewed Roger to find out where LAN-SCE fits in this picture, and he gave us a unique perspective on neutron scattering's history and on the most pressing problems facing the field today.

Science: Last winter, a petition to secure funding for neutron scattering's future was sent to Presidential Science Adviser Allan Bromley. Can you explain why?

Pynn: The petition actually relates to the present as much as the future. Right now, the two most powerful nuclear reactors used for neutron scattering in this country are closed to address safety concerns. Brookhaven has been closed for nearly a year, and Oak Ridge has been closed for over three years. In addition, the National Institute of Standards and Technology reactor was closed during 1989 while neutron guides were being added. It is bad enough that these facilities haven't been available for research, but a larger and larger share of the neutron-scattering budget has also been spent on their safety studies and repairs—and that drains the budgets of the remaining facilities. For example, this year at LANSCE we'll be able to run our beam less than half the year, which means we won't accommodate nearly as many experiments as we'd like.

Science: Is the situation likely to improve in the near future?

Pynn: Well to make matters worse, these reactors are all around twenty years old, and their neutron-scattering instruments are antiquated. If funds were available, we could improve experiments at these facilities by a factor of five to ten simply by modernizing equipment. For a while, we wouldn't even need to worry about getting higher beam fluxes. I should say, however, that the future of this field depends on building facilities with higher beam fluxes. If the U.S. wants to keep pace with Europe and Japan, it is critical that we build a next-generation neutron source.

Science: Why is it so critical to keep pace? The petition to Bromley mentioned economic growth.

Pynn: As far as that goes, neutron scattering has all sorts of technological applications. We can use neutron diffraction to do nondestructive testing of residual stresses and strains in a wide variety of industrial products. Small-angle scattering can examine the structures of polymers and colloids, which

The future of this field depends on building facilities with higher beam fluxes. If the U.S. wants to keep pace with Europe and Japan, it is critical that we build a next-generation neutron source.

are the basic ingredients of many modern materials. Neutron reflectometry can look at the structures of protective coatings and lubricants. Look what came across my desk today: "A Neutron Scattering Study of Diffusion and Permeation Processes through Pores in Clay." You can imagine applying that to underground waste disposal or oil mining. Really, the industrial uses of neutron scattering are endless.

Science: So you argue that we should fund neutron scattering because it is crucial to the industrial future of the United States?

Pynn: Partly. I am very interested in promoting industrial uses of neutron scattering, but that's not the only reason to promote the technique. We should develop materials-research techniques and do science with them whether their applications are instantly apparent or not. For example, when new materials like high-temperature superconductors

come along, you want to understand them and you use every resource you have—electron microscopy, nuclear magnetic resonance, neutron scattering, x rays, or whatever. In the case of high-temperature superconductors, neutron scattering gives unique information about structure because it can locate light elements and also look at magnetic properties. Researchers at Brookhaven were doing valuable experiments of this kind when the reactor was closed down. For other problems, other methods might be more valuable, but it is impossible to predict which ones. So we should maintain a capability in each technique if we want to have an effective materials research program. Furthermore, techniques need to be explored because they open new areas of basic interest. As a matter of fact, the development of neutron scattering illustrates this point quite well. Back in the forties, scientists shot neutrons into samples because they wanted to find out about the fundamental properties of this new elementary particle. Today, neutron scattering provides useful information about the samples themselves to physicists, chemists, and biologists in addition to all the industries I mentioned.

Science: Go back and tell us a little more of the early history.

Pynn: Soon after Chadwick's discovery of neutrons in 1932, researchers began trying to understand the properties of these particles by sending them through various materials. Theorists knew from quantum mechanics that neutrons would behave like waves, and that low-energy neutrons would produce interference or diffraction patterns much like x rays. But there were many theories and disagreements about specifics. For example, people wondered what details of the interaction between a neutron and a nucleus could be determined from a scattering experiment, and they wondered whether the neutron's expected magnetic field would resemble that of

a bar magnet, a current loop, or something in between. Fermi resolved the first problem in 1936 by proving theoretically that the neutron-nucleus interaction is described by only one meaningful parameter—what we call today the scattering length. The second problem grew into a fertile debate between Bloch and Schwinger. In 1939, Halpern and Johnson suggested experiments to settle this argument and to test a number of other theories. But of course in the early forties, pressure from the Manhattan Project for critical neutron data was dictating most of the neutron research.

Science: Were these early neutron-scattering experiments much the same as those done today?

Pynn: Until the first reactors were built, researchers couldn't generate enough flux for anything but transmission experiments, but after the war Fermi and Marshall began experiments comparable to ours today. They measured neutron diffraction from a crystal whose structure was known from x-ray experiments—some simple thing like rock salt—and then they drew conclusions about the interaction of neutrons with different nuclei by comparing the neutron data with the x-ray data. At the same time, Wollan and Shull were using a similar method to study neutron diffraction from crystalline powders and getting some very important results. In 1947 they demonstrated that neutrons can see hydrogen atoms in a crystal, and that unique ability has become neutron scattering's great contribution to the study of biological systems. Another seminal experiment was Hughes and Burgy's verification of Schwinger's current-loop hypothesis. They produced the first fully polarized neutron beam by using magnetic mirrors—precursors to the reflection technique that we use today in neutron guide tubes. I should say that in all these experiments people were still primarily interested in understanding the neutron.
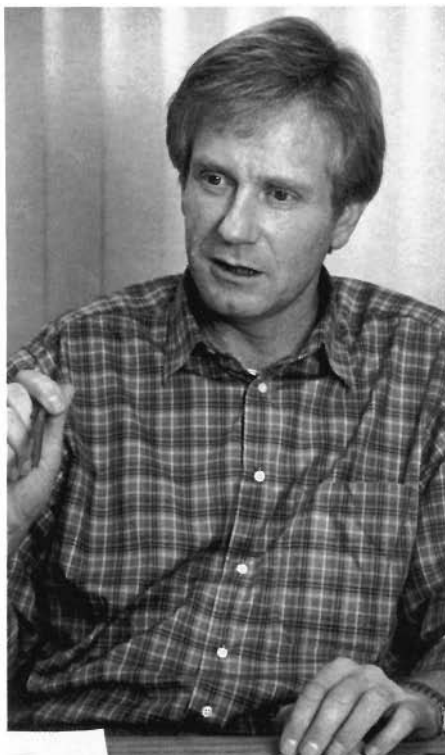
The great breakthrough for neutron scattering came in 1952 with the first [measurements of] the internal dynamics of condensed-matter samples.

It was the opposite of most research today, where we assume we know about the neutron and draw conclusions about the sample from the scattering.

Science: Even so, it sounds like the early researchers conceived of most of the techniques in use today.

Pynn: To a large extent that is true. Most of the gains made in neutron scattering have been technological. For example, if you want to make guide tubes that transport neutrons with minimum losses, you have to make glass which is optically flat enough and you have to learn how to vapor-deposit nickel. We have made many improvements like that. The neutron spin-echo technique that came along in the early seventies was a genuinely new method of getting high resolution without losing much intensity, but for the most part the basic experimental concepts go way back to the beginning.

Science: When did researchers start using neutrons to probe materials?

Pynn: Once the properties of the neutron were understood, it was natural to turn scattering experiments into a means of studying static crystalline structures—a change which occurred in the early to mid fifties. This wasn't a completely new research field, however; essentially it extended the x-ray crystallography work that had been going on for some forty years. The great breakthrough for neutron scattering came in 1952 with the first inelastic-scattering experiments, which investigated the internal dynamics of condensed-matter samples.

In an inelastic-scattering experiment, you measure the energy and momentum a neutron transfers to the atoms of a solid—energy which can then vibrate throughout the sample as a collective excitation called a phonon. Theorists had already described phonons as vibrational waves whose frequencies relate to the interatomic forces in solids, but neutron scattering was the first way to make measurements in real samples. Groups in France and the U.S. began measuring phonons using time-of-flight spectrometers, and in Canada Brockhouse began using what he called a constant-Q scan on his triple-axis spectrometer. Such a machine measures only at well-defined scattering angles and energy transfers, which makes for very precise, very focussed inelastic-scattering data. Brockhouse used to say something like you never get more data than you need from a triple-axis spectrometer, and you always get it at a rate that lets you figure out exactly what to measure next. As I said, many people were working to develop instruments for measuring inelastic scattering processes, but the three-axis spectrometer became the prevalent tool. For whatever reason, Brockhouse was able to apply it to a wider range of materials and problems than anyone else. Given the technology at the time, he made some spectacular measurements

of phonons in semiconductors, and metals, and ionic crystals, and everything you can think of. He just knocked them off one after another.

The study of phonons in all sorts of materials became a major focus of neutron-scattering research throughout much of the sixties. As the experimental part of my doctoral thesis, for example, I studied the phonon spectrum of magnesium. I had this huge single crystal of magnesium—four inches long and an inch and a half across—and I measured phonons in the damn thing. Today people would say, "So what? Why would you bother to measure phonons in a single crystal of magnesium?" The answer is that we were just beginning to learn how to calculate phonon frequencies in metals from first principles. People would propose models of the different bonding forces in metals, calculate what the phonon frequencies should be, and then compare them with neutron-scattering measurements. If they didn't see agreement, they would go back and fool around with their models or come up with better ones. Efforts like those gave us a much better qualitative and quantitative understanding of what holds metals together.

Science: What were some other major discoveries from the fifties and sixties?

Pynn: Perhaps the most striking was Van Hove's elegant formulation of the neutron-scattering law in 1954. Before Van Hove people used neutrons to study structure and dynamics in a variety of ways, but they didn't understand how these different techniques related to each other. Van Hove's analysis unified the whole field of research. It brought together in one simple equation the static structure factor, which we measure in diffraction experiments, and the collective excitations, which we measure in inelastic-scattering experiments. Before Van Hove no one had really demonstrated the simplicity and power of neutron scattering as a research tool.

There were also neutron-scattering results which affected materials physics as a whole. In 1951 Cliff Shull used neutron diffraction to verify Neel's theory of antiferromagnetic structure, which said that in certain materials the magnetic moments of electrons line up in alternating sequences. There had been no way of proving the existence of such antiferromagnets until neutron scattering came along. Once this structure had been verified, people used inelastic neutron scattering to discover the collective excitations in these antiferromagnets and then developed theories to explain them. In 1961 neutron scattering was also used to observe rotons, a type of collective excitation in superfluid helium which Landau had predicted on the basis of God knows what genius.

Science: It sounds like neutron scattering research was beginning to broaden its scope. How did that happen?

Pynn: After a while it becomes tiresome to measure a sample of tedium boride for the seventy-fourth time just to find out what the interatomic forces are at some other temperature. In the mid sixties people gradually started to do experiments that involved some special physical phenomena; for example, if a material went through a ferroelectric transition, they would use neutron scattering to see whether the phonons had played a role in this transition. The motivation became, "Well, here is this phenomenon called a phase transition. What can we learn about it with neutrons?" rather than, "Here we have this piece of solid garbage on the shelf. Let's measure it."

Science: What is a phase transition?

Pynn: A phase transition is a disappearance of order in a sample of matter, brought on by a change in some external factor like temperature or pressure. Everyone is familiar with ice changing to water or water to steam, transitions in which the structural order as well as the bulk properties of the mat-

[Van Hove] brought together in one simple equation the static structure factor, which we measure in diffraction experiments, and the collective excitations, which we measure in inelastic-scattering experiments. Before [that], no one had really demonstrated the simplicity and power of neutron scattering as a research tool.

When I look at the theory of critical phenomena, it seems clear that the universality classes could not have been identified without the clues provided by [neutron scattering] experiments.

ter change suddenly; but there are also phase transitions in which order changes gradually. The order can be structural, or magnetic, or ferroelectric, or whatever. Since the early sixties, there has been a huge intellectual effort to describe the gradual changes in order as different systems approach the so-called critical point where order disappears altogether. By the early seventies, theorists succeeded in organizing these continuous phase transitions into universality classes defined by the way the order changes. They discovered that the symmetries of the system, and not the particular details of the forces responsible for creating the order, determine the universality class of the phase transition. Some very high-powered people have worked on this theory of critical phenomena, and a Nobel prize was given a few years ago for work in this field.

Science: How did neutron scattering contribute to this study?

Pynn: Let's take the specific example of a magnetic phase transition. As you heat a magnetic sample, its spatially averaged magnetism will decrease continuously until you reach the Curie temperature [critical point] where the magnetism disappears completely. This might seem very simple, but it is not. Although the average magnetism goes through a smooth decrease, the magnetism at any point in the sample fluctuates more and more about the average as the sample nears its critical point. In addition, the fluctuations in magnetization at one point in the sample are correlated with fluctuations at a nearby point. As the Curie temperature is approached, the spatial extent of these correlations becomes very large and the fluctuations slow down. The way in which the correlation length increases and the fluctuations slow down—that is, the dependence on temperature— characterizes the universality class of the transition. Neutrons can measure both of these quantities, as well as the

average magnetic order. Without all the neutron scattering experiments, I don't think theorists would have been able to understand these immensely complicated phenomena. Now when I look at the theory of critical phenomena, it seems clear that the universality classes could not have been identified without the clues provided by the experiments. So neutron scattering had a large role in that development.

Science: Was neutron scattering a well-recognized research field at that time?

Pynn: It was beginning to become one. Until the early seventies, the people doing neutron scattering were actually condensed-matter physicists who had become interested in the technique. More important than that, however, was the lack of dedicated research facilities. In the fifties and much of the sixties, neutron scattering was just a parasitic operation at research reactors that had been built to study things like isotope production and radiation damage. Those experiments had first priority at all the facilities because the people doing neutron scattering didn't decide the politics of reactor use. Instead, they hung around the edges, borrowing beam lines and setting up spectrometers when they could. The first reactor built exclusively for neutron scattering was the Brookhaven High Flux Reactor, which came on line in the mid sixties.

Science: Did the Brookhaven reactor begin the field as we know it today?

Pynn: I'm not sure I can define a beginning. Certainly the advent of the Brookhaven reactor gave neutron scattering a dedicated tool, but it didn't change things that drastically. Even though this new reactor was dedicated to neutron scattering, it was also dedicated, in a sense, to the few people who were employed at Brookhaven. Essentially, you still had a small neutron-scattering group working by themselves. The field as we know it today—with scientists from all over doing their re-

search at large central facilities—started during the early seventies in Europe.

Science: Was there a definable beginning to this change?

Pynn: The first real user facility—and incidentally still the pre-eminent neutron scattering facility in the world today—is the Institut Laue-Langevin in Grenoble, France. Sometime after the war, a German professor of physics by the name of Maier-Leibnitz proposed building a large research reactor as part of the Franco-German cooperative effort, and this reactor became the ILL. It was born of politics, not because people said, "We have to do neutron scattering." It was born because the Germans and the French wanted to get together for scientific and cultural exchanges. As the rumor goes, it was born because Maier-Leibnitz had a relative who was close to Adenauer, but that may not be true. At any rate, Maier-Leibnitz persuaded the politicians that a reactor dedicated to neutron scattering was something they needed as well as a scientific need. Next, he toured the United States and Canada, which were strong in neutron scattering at that time, asking for advice about designing a first-rate research program. The advice was, "Well, first you build your three-axis spectrometers and get a program established, and then you think about doing something else." Triple-axis machines were very popular at that time, especially in the United States, and many of the questions asked in inelastic neutron scattering were dictated by the machine's characteristics. Even today that is true to a certain extent. Anyway, Maier-Leibnitz said, "Thank you very much, but I will not build a single three-axis machine at my institute." And at first he didn't. Instead, he hired a bunch of young people, many of whom knew nothing whatsoever about neutron scattering, and set them working on some of his own bright ideas. They were happy to try anything because they didn't know what

LANSCE operations manager Richard Woods with Roger Pynn and Dianne Hyer

was impossible. They invented things and incorporated ideas from prototypical instruments at smaller reactors in France and Germany, and they wound up with all sorts of novel instrumentation—and only one three-axis machine when the institute became operational.

Science: What were some of Maier-Leibnitz's bright ideas?

Pynn: There were several. One of the great successes at the ILL has been the use of long-wavelength neutrons—what we call cold neutrons. There were other cold-neutron sources in operation at the time the ILL was built, but they were created by putting moderators on existing beams. Maier-Leibnitz proposed building the moderator in next to the reactor core so cold neutrons would be generated in copious quantities whenever the reactor was running. At the time, many people said the ILL people were crazy to tie the operation of the reactor and the cold source together in this way. Next, Maier-Leibnitz decided to use hundreds of meters of optically flat, nickel-coated glass tubes to transport neutrons away from the reactor and into a huge new guide hall where the background radiation would be low.

In the fifties and much of the sixties, neutron scattering was ... a parasitic operation at research reactors.... The people doing neutron scattering ... hung around the edges, borrowing beam lines and setting up spectrometers when they could.

The Brookhaven reactor gave neutron scattering a dedicated tool, but ... it was also dedicated, in a sense, to the few people who were employed at Brookhaven.

This was an incredibly courageous decision given that guide tubes of that type had only been benchtop tested and cost several thousand dollars a meter. The ILL researchers also improved the angular resolution of small-angle scattering experiments by putting their detector 40 meters away from the sample—a distance that had never been tried before. There is still not an instrument in the world that comes close to the resolution of that machine.

Maier-Leibnitz also had some wild ideas that didn't work. For example, he wanted to replace triple-axis spectrometers with three separate remotely controlled units that could be moved around on air pads. So the ILL people laid down enormous, smooth marble floors called tanzboden, or dance floors, and built these air-levitated units—sort of nineteen seventies R2D2 units. Those things never worked as planned. In the end someone clamped them together into a traditional three-axis spectrometer that moved on air pads instead of on the naval gun mounts used in the sixties. That technology has now spread almost everywhere.

Science: Were these developments in instrumentation motivated by specific scientific questions?

Pynn: Not really. Maier-Leibnitz had specific ideas about improving the measurement techniques themselves, and he knew this would lead to new and exciting science—an obvious idea that seems to have been largely misunderstood in the United States. Remember, neutron scattering is a signal-limited technique. You can't measure a particular effect unless enough neutrons reach the detector. Let's take a simple example. The original triple-axis spectrometers used big, flat monochromator and analyzer crystals, usually aluminum or copper or something else that grew well in single crystal form. That is analogous to doing optics with rather poor flat mirrors, and it is very inefficient. To maximize the

intensity of a light beam, you usually use good-quality reflecting surfaces and focus the beam with curved mirrors or lenses. People tried all sorts of things to make single crystals more efficient at transmitting neutrons, for instance laying them on a table and beating them with a hammer. That helped a little bit but it wasn't very controlled.

Maier-Leibnitz and his co-workers thought they knew how to improve the flat crystals, and they approached this problem in a systematic way—tailoring new materials and using multiple crystals to achieve a focusing effect. Spectrometers today deliver much better intensity and resolution than the original instruments, mostly due to improvements in individual components rather than to increases in neutron fluxes.

Science: How do you increase the scattering efficiency of crystals?

Pynn: It is a very sophisticated technique that involves well-defined distortions of the crystal. You start with a crystal that has less than a certain density of dislocations, then you cut it to a specific shape, then you squeeze it

Neutron scattering is a signal-limited technique. You can't measure a particular effect unless enough neutrons reach the detector.

along a specific direction usually while heating it to a certain temperature. This technique has really only been pursued at the ILL, which fits with their history of developing instrumentation. Let me say right now that one-third of the initial budget of the ILL was for instrument and spectrometer design and construction. By comparison, the plan for a next-generation source in the U.S. allots only one-fifteenth of the budget to instrumentation, and you know exactly

Maier-Leibnitz had specific ideas about improving the measurement techniques themselves, and he knew this would lead to new and exciting science—an obvious idea that seems to have been largely misunderstood in the United States.

what that will produce—nothing new.

Science: How important has instrument development been to the field?

Pynn: Initially people used diffractometers and triple-axis machines, which limited experiments to a small range of phenomena. Now we use small-angle-scattering instruments, backscattering instruments, diffuse-scattering instruments, time-of-flight instruments, spin-echo instruments, reflectometers, and all sorts of other things—and this variety is very important. Basically, you can imagine that any experiment you want to do exists in a space whose dimensions are momentum transfer, energy transfer, and resolution. Because neutron scattering is limited by the flux of neutrons you have in your beam, you must develop special types of instruments to get you into different corners of that space. A generic instrument simply won't take you ev-

erywhere. So people have built special new instruments to study phenomena involving high momentum transfer, or very high resolution, or low energies, or whatever specific problem they were interested in. The ILL has done a great deal to expand the momentum-energy space to which we have access, and spallation sources like LANSCE also expand it. I should point out that future experiments could fall anywhere in this space at random; so it is difficult to overemphasize instrument development.

Science: When did the ILL people begin inviting outside researchers to come and do experiments?

Pynn: I suppose they must have realized the enormous potential of the place once they began building all the spectrometers and opening all the beam lines. Mössbauer, who succeeded Maier-Leibnitz as director of the ILL, really initiated the user system by encouraging proposals from universities and by setting up a committee to evaluate these proposals and decide who should get beam time. By attracting a variety of users, this system made a great contribution to the expansion of the field. The ILL uses it today, and we are copying it here at LANSCE.

Science: Say more about how neutron scattering came to be used for the widely varying research we see today.

Pynn: For the most part, the ideas came from outside the field, not from the professional neutron scatterers. I know that a very strong group from Oxford drove much of the expansion of chemistry research at ILL, and the push toward uses in polymer science and biology also came from the outside.

Science: How did these outsiders find out that neutron scattering was such a useful tool?

Pynn: I can at least answer that question for the Oxford chemists. Sometime back in the late fifties, Cockcroft, who had been the director of the Atomic Energy Research Establishment at the Har-

well reactor, was made chairman of a committee to get university researchers involved in work at government labs. So he gave the people at the Harwell reactor the equivalent of $3 million in today's money to develop neutron scattering, and they drove up the road to Oxford and started trying to interest people. Among others, they talked to a chemist named John White, a real dynamo, and he started to figure out the experiments you could do with neutron scattering in chemistry. He later became one of the directors of the ILL, establishing strong connections with the chemistry group at Oxford.

Some physicists who began with neutron scattering also helped widen the field, for example Bernard Jacrot. He had been doing scattering experiments since the fifties, studying magnetism and magnetic properties—he did some of the early time-of-flight experiments. Anyway, the story goes that sometime during his stay at the ILL, Jacrot put a dump truck outside his office window, threw everything into it, and started doing biology instead. He was one of the people who showed the power of the contrast-matching technique, which is now almost second nature to biologists.

All these people came together at the ILL. So neutron scattering got a large facility where it could grow with influences from other fields, and Europe got a centralized facility where other fields could use neutron scattering. Today, for example, the U.K. sees the ILL as a training ground for Ph.D.'s in methods of research. No such movement happened in the United States. In the United States, small groups working around their own reactors never really got together, and that is as true today as it was twenty years ago.

Science: Why did that happen?

Pynn: Perhaps it has something to do with the way things are funded here. I'm not a great fan of the peer-review system as it works in the U.S. because

In the United States, small groups working around their own reactors never really got together, and that is as true today as it was twenty years ago.

International Workshop on Cold Neutron Sources, March 1990

it doesn't encourage cooperation or synthesis of ideas. Anyway, the sociology of neutron scattering in the United States has centered the work in small, parochial groups that are very defensive of what they have acquired over the years. In contrast, the ILL became a user facility—people from universities and laboratories came to do science there. Until quite recently the idea of a facility catering to outside users didn't exist in the DOE. Isolated groups of professional neutron scatterers ran every one of the U.S. facilities.

Science: Did outside researchers come and ask to do experiments?

Pynn: The doors to the U.S. facilities certainly weren't wide open. In the scientific sense, the neutron professionals determined everything that happened in neutron scattering. I'm not saying they didn't do good science—they did—but the field was inbred and cut off from new ideas and influences.

Also, because researchers from different facilities did not cooperate and because there was no user group to ask for more facilities, the field got no money beyond what was necessary to keep the small groups going. In fact, the U.S. is missing two generations of scientists in neutron scattering. Almost all the people who were trained as postdocs at Brookhaven or Oak Ridge in the sixties and seventies couldn't get a permanent job in neutron scattering and went on to something else. Despite my air of venerability, I'm young in the American neutron-scattering scene. The number of twenty-five- to thirty-year-olds is essentially zero.

Science: Did your coming here coincide with the DOE's idea of encouraging user groups?

Pynn: I think the idea of a designated user facility came earlier—near the beginning of this decade. A number of DOE panels have looked into neutron scattering. The first one asked what would happen to neutron scattering in

the United States if its budget remained constant. They concluded that neutron scattering in the United States was very healthy, thank you. It was doing better than everybody thought because Americans were smarter and didn't need expensive, modern facilities. This mistake was eventually recognized when sufficient people managed to get out of the United States and see for themselves. After that, various committees were set up to look into the field, and they said, "We have to do better. The Europeans are way ahead of us, and the Japanese are coming." I remember sitting in front of one of these committees during a visit to Brookhaven during 1980, when I was still at the ILL. Various other labs had testified during that hearing, and the Oak Ridge people, for example, had said, "We had one hundred sixty visitors last year." That was a nice introduction to my pie chart showing how all the ILL users broke out into different subjects. There were sixteen hundred of them—an order-of-magnitude difference!

Science: Has neutron scattering in the United States changed in response to these findings?

Pynn: It is not obvious to me that the sociology of the field has changed much here. There is just no coordinated leadership anywhere in this country. As a small beginning, LANSCE and the pulsed source at Argonne now have the same advisory board passing judgment on experimental proposals, but any talk of a national committee seems to fall on deaf ears. It is very hard to get people from the various facilities to work together at this, but we have to try

and collaborate as we can. I don't think the field will move ahead in the United States unless we do.

Science: Is that why you left the ILL and came here?

Pynn: Not really. You have to realize that the ILL has set the standard in neutron scattering for the last decade and will probably do so for the next, but I had basically done everything I wanted to do there. I had built a polarized-

Until quite recently the idea of a facility catering to outside users didn't exist in the DOE. Isolated groups of professional neutron scatterers ran every one of the U.S. facilities.

neutron spectrometer; I had participated in many exciting experiments; and I had worked throughout the ILL organization in various capacities. I could have stayed and done my own research there, but LANSCE was a challenge to me. After coming here as a consultant, I began to wonder if it could be made better than its European competition. I think we have succeeded in some ways, but we could do a lot better with only 20 percent more money.

Science: Is LANSCE now a state-of-the-art neutron source?

Pynn: Before answering that question we should ask if LANSCE is a state-of-the art spallation source. Most neutron-scattering facilities, and most of the

IPNS/LANSCE External Program Advisory Committee

ones I've mentioned today, use beams of neutrons produced by reactors. Spallation sources, on the other hand, direct a beam of accelerated particles onto a target, which then emits bursts of high energy neutrons. With that said, one way to answer my question is in terms of the reliability of the beam-delivery system.

In 1988, we had an awful time; on average, we had neutrons on the samples in our spectrometers during only 50 percent of the time we scheduled. Since we run a user program with people coming in from out of town, that is just a complete disaster. It really is. Suppose you have some guy come in with samples that last a few hours or days, and the beam is down. There goes $10,000 worth of samples to the wastebasket because they couldn't be run on the machine. So we made reliability our highest priority this year and finished with the beam operating 74 percent of the time. That is an acceptable level for a user program, because a lot of the 26-percent loss is an hour here and an hour there. In fact, that is almost as high a reliability as you can expect from a state-of-the-art accelerator source because they are incredibly complicated beasts. The accelerator itself has all sorts of power supplies and magnets. You need to tune beams to get them into closed orbits—all sorts of complicated things like that. If an accelerator has very high reliability—and some do—its design and performance are probably not at the forefront of technology.

Another way to ask if we're state-of-the-art involves the intensity of the neutron beams on our spectrometers. By the end of the 1989 run cycles, we had a higher peak neutron flux than any other spallation source in the world— and when the proton-storage ring is operating at full capacity, our neutron fluxes will be even higher.

Science: How does LANSCE compare with the best reactor sources?

Pynn: I like to use our small-angle-scattering machine, the Low-Q Diffractometer, as an example, because many people thought spallation sources would not be suited to small-angle-scattering experiments. As always, the standard for any comparison in neutron scattering is the similar machine at the ILL. We have optimized the LQD at LANSCE so that our results are as good or better than the ILL's when we probe length scales up to 500 angstroms or so. At larger length scales the ILL instrument wins, but most experiments fall in the range I just mentioned. In that sense we are competitive. However, I have always taken the view that reactor and spallation sources are complementary— that you need both types if you want a complete neutron-scattering program. You can do a lot of things with each one that you can't do with the other. For example, spallation sources are better for powder diffraction, but a simple three-axis spectrometer at a reactor source still produces inelastic-scattering data that we cannot duplicate.

Science: Are you also competitive with the ILL in the number of users?

Pynn: In 1989, one hundred and eighty-six scientists were involved in experiments at LANSCE, and one hundred and twenty of them actually came and worked. In addition, about three thousand people are now on our mailing list. The IPNS at Argonne, which was the first neutron-scattering facility in this country to have a user program, has done a tremendous job of bringing in users and expanding the user community. We are gradually attracting more and more people from universities and

I have always taken the view that reactor and spallation sources are complementary—that you need both types if you want a complete neutron-scattering program.

An Interview with Roger Pynn

industry, but it is a long uphill battle.
Science: In what sense is it uphill? Do
people have to be very brave to try a
neutron-scattering experiment?
Pynn: If you think of going to a facil-
ity to do something that you have never
done before, just wanting to get the an-
swer to your scientific question and not
knowing exactly what is involved, you
realize it must be quite daunting. Even
going to the lab next door and borrow-
ing a simple piece of equipment can be
extremely difficult, and might even keep
you from an important discovery.
Science: That's human nature isn't it?
Pynn: I guess, but the adventurous peo-
ple who overcome these barriers can
be extremely successful. Lots of peo-
ple in Europe now use neutron scatter-
ing and no other technique. They may
be at a university and have no lab of
their own, but they can rely on doing
neutron-scattering experiments at user
facilities.
Science: They get hooked on the field?
Pynn: Perhaps they get hooked on go-
ing to Grenoble and skiing, either that
or on French wine and cuisine! But I
do think it is very important to over-
come the barriers that prevent scientists
from coming to user facilities, and we
work very hard at it. In some sense the
ILL succeeded because they required
only a one-page experiment proposal
and would pay travel fare and living ex-
penses if they accepted it. So people
had nothing to lose. The ILL is very
unusual in paying those expenses; we
don't have that kind of budget at LAN-
SCE. Including the proton-storage ring,
our budget in 1988 was between $14
million and $15 million dollars per year,
whereas the ILL's was over $50 million.
Science: How do the committees deter-
mine who gets to do experiments?
Pynn: My answer to that depends on
which hat I wear, my user's hat or my
LANSCE director's hat. Recently I
got the results from four proposals I
submitted to the ILL. In that case I'm

the user, so I sometimes think the ILL
committees toss a coin and don't con-
sider scientific merit at all. I had one
proposal out of four turned down, you
see, so I argue that the committee just
didn't understand that proposal. I'm
sure all users have that attitude, but we
try, in principle, to get together a group
of people who can judge. That is ex-
tremely hard to do. It comes back to
the question whether theory should lead
experiment or not. If I propose to you

*Lots of people in Europe
now use neutron scattering
and no other technique...
They can rely on doing neu-
tron-scattering experiments
at user facilities.*

an experiment to look for Landau's ro-
ton, you will probably give me beam
time, provided you understand the the-
ory of the roton and understand that
neutrons can find it. It is an essential
experiment; I could end up verifying
Landau's theory or demolishing it. But
suppose I said to you, "There is an ex-
tremely good theory of magnetic ex-
citations in a material called TMMC.
Among other things it predicts four
modes, and I want to see whether there
really are four." You would probably
tell me to do something rude, right? In
fact, I took part in that experiment, and
we happened to identify five modes in-
stead of four. We got beam time only
because we had proposed something
else.
Science: Do you think the committees
are too conservative?
Pynn: Quite often, yes. That is one
of the great disadvantages of the user
system, and I assume it must be the
same for a grant system unless you get
somebody who says, "Let's risk it. This
looks like wild stuff but it just may pay

off." Even so, I'm sure an open system
of proposals and reviews is better than a
private party where a few people control
and use the beam time.
Science: Who at your facility chooses
the experiments?
Pynn: I mentioned that we have a joint
program advisory committee with the
facility at Argonne National Lab. Peo-
ple submit proposals to both facili-
ties, and the committee breaks up into
three subcommittees: one that looks at
diffraction, one that looks at small-angle
scattering and reflectometry, and a third
that looks at inelastic scattering. Un-
fortunately, these are technique-oriented
groups, so there may be only one per-
son in each group who is an expert on
a particular type of science. It is hard
to be sure that you always get the best
decisions out of a committee like that.
But to have a wider scientific debate
in each subcommittee, you need more
members, and that costs money. The
way to solve the problem, in my view,
is to involve other scattering centers in
the same committee and share the costs.
Science: Is there a collaborative ar-
rangement for the people who come
here to do science?
Pynn: Users aren't required to have
any experience with neutron scattering
to come do an experiment at LANSCE.
Instead, we generally set it up so each
user has a local contact, some card-
carrying member of the neutron Mafia
who doesn't mind what he shoots his
neutrons at.
Science: Let's talk about your connec-
tion with the rest of the Laboratory.
Pynn: Well, we have a number of peo-
ple from different divisions who work
more or less permanently at LANSCE.
We have people from the Life Sciences
Division, someone from Materials Sci-
ence and Technology, and someone
from Chemical and Laser Sciences as
well. That's about the extent of it right
now. Even here inside the Lab there are
still a lot of people who don't realize

the useful information they can get with neutron scattering. I know this because I sat on a committee which reviewed proposals for internally supported research [ISRD], and last year a number of them could have used neutrons but didn't propose to do so. They were not intentionally ignoring neutron scattering; they just didn't think of it because they didn't know it was available or what it could do. To a large extent that is our fault—we have to get the information out there. With that in mind, we have been trying to set up a committee with representatives from lots of different divisions to try and acquaint people with the neutron-scattering facility, and to teach them which experiments in their work can use neutrons. We hope they will eventually become advocates of neutron scattering in the Lab. In particular, I think we haven't done a good job of selling to the weapons community. There is a communications problem because we are in an open area, and there are a number of us, including me, who can't hear classified information.

Science: That's right. You're an alien.

Pynn: I'm an alien and if I'm going to discuss weapons-related neutron scattering with people, it has to be in a very generic sense—but that is often good enough. It is only when you are planning the details of the experiments that you need to know details of materials and composition or of shapes and forms and sizes. So I can usually tell someone whether their experiment makes sense without knowing classified details.

The problems I have as an alien are not technical; they're bureaucratic. One thing which seems to have changed since I came to work in Los Alamos is the thicket of rules about foreign nationals at DOE facilities. I can understand the need for security and I respect it, but some of the new rules aren't well considered or in the national interest. For example, there was a draft regulation which would have prevented for-

eign nationals from using almost any computer at a DOE facility—including PCs—for any kind of work, classified or not. That's not very smart. The U.S. needs its scientific contacts with the rest of the world—perhaps now more than in the past. So a sensible compromise has to be found which encourages international science and preserves national security. I have to say, though, I have never experienced any discrimination towards me here at Los Alamos. The people here are always very thoughtful and hospitable, even though the bureaucracy can sometimes give real meaning to the word "alien"—a green thing with horns that spies or worse.

Science: What do you see as the future of neutron scattering?

Pynn: I talked before about the need for a next-generation source. Ideally, we should build both a reactor source and a spallation source—and also keep the older facilities for more patient development of the field. That way we could do the science which needs high-intensity sources, and people would also have places to think, work, try new ideas, and train students without too much pressure. I don't think you can have a very healthy program without such places. However, on a more realistic level, we need to expand our user base in order to generate support for at least one new source. If basic researchers and people from industry combined their support, their voice would be difficult to ignore. So we have to contribute as we can to technological and industrial problems, as well as to basic research. Finally, neutron scatterers need to speak with a unified voice and work together to produce a coherent national policy. At the moment, some of us are trying to create a national steering committee. If we can accomplish that and maintain the present fragile unity, I think the field will be poised to move ahead. ∎

The U.S. needs its scientific contacts with the rest of the world—perhaps now more than in the past. So a sensible compromise has to be found which encourages international science and preserves national security.

LANSCE
a facility for users
by Dianne K. Hyer and Roger Pynn
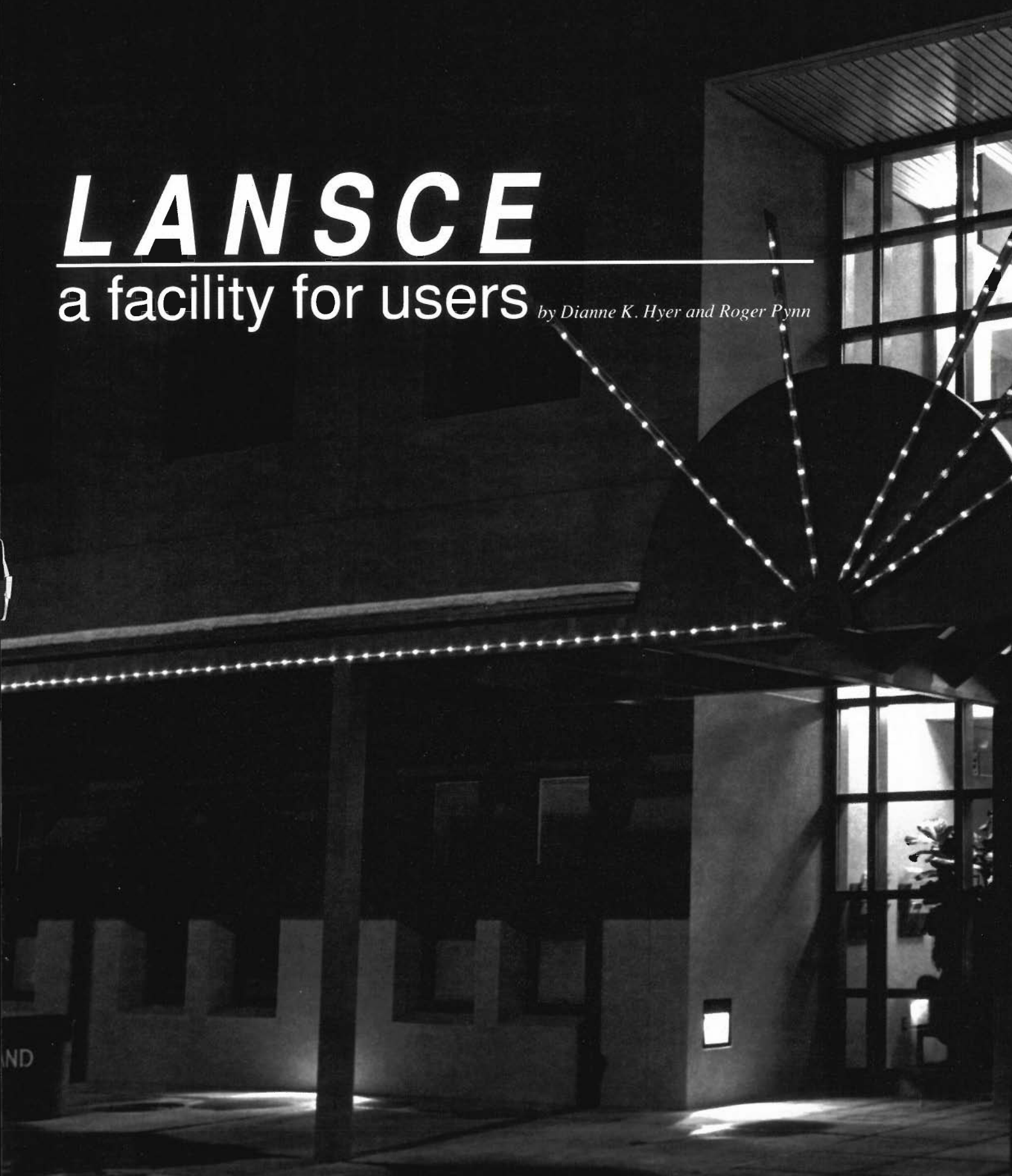
Most facilities for neutron-scattering experiments are based at nuclear reactors, but the Manuel Lujan, Jr. Neutron Scattering Center (or, less formally, LANSCE) is one of a new generation of neutron sources that has become available during the past decade—those based at proton accelerators. The new accelerator sources have several advantages. First, an accelerator source does not involve nuclear fission and therefore poses no threat of runaway operation that might lead to dispersal of hazardous radiation. Safe operation of a reactor requires complex control mechanisms, and, as evidenced by the Chernobyl disaster, accidents involving nuclear reactors can cause widespread harm to people and the environment. The cost of reducing the probability of such accidents to a level society finds acceptable is rapidly becoming prohibitive. A second advantage of an accelerator source is that it produces pulses of neutrons, which can be used much more efficiently than the continuous flow of neutrons provided by a reactor. Thus, although the average neutron flux from a pulsed neutron source may be 1000 times smaller than that from a reactor, the pulsed source is capable of producing neutron-scattering results of the same caliber as the reactor. A lower neutron flux means a lower heat-removal load. For example, only 100 kilowatts of heat have to be removed from the LANSCE pulsed source, whereas 500 times more heat must be removed from a comparable reactor source. Finally, with today's technology we could build an accelerator source that provided 100 times more neutrons than the best source of this type currently operating. In contrast, the best reactor we could build would generate only five to ten times more neutrons than existing reactors; any further improvement would require technology beyond today's dreams. LANSCE
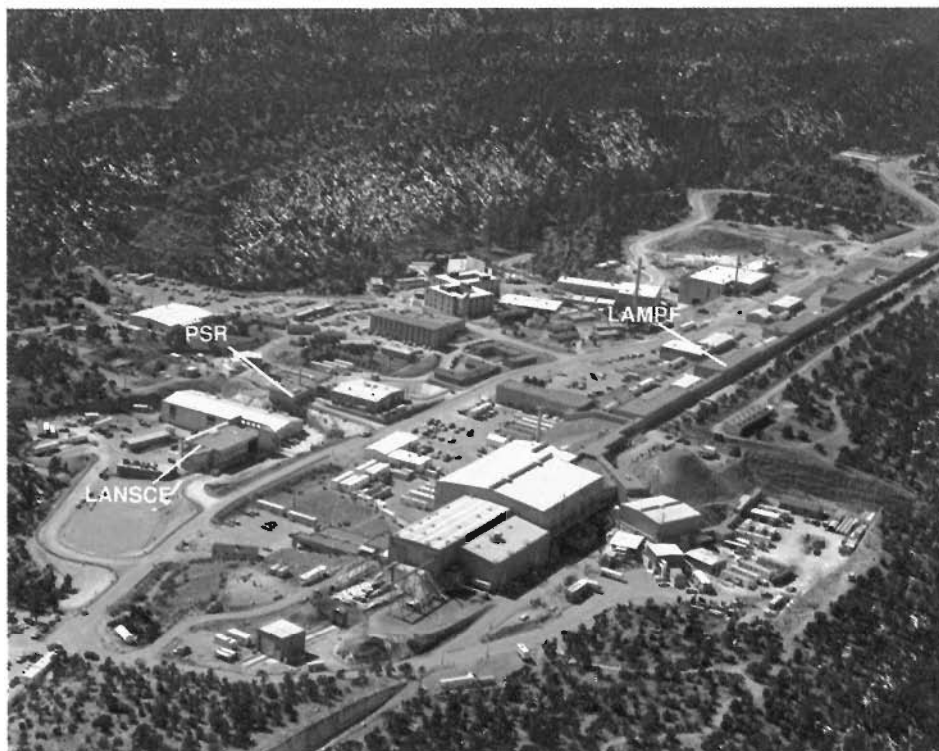
This aerial view of Technical Area 53 shows the proton linear accelerator of the Los Alamos Meson Physics Facility, which produces high-intensity pulses of protons for research in medium-energy physics; the Proton Storage Ring, which converts some of the LAMPF pulses into higher-intensity proton pulses; and LANSCE, where those proton pulses are used to create high-intensity pulses of neutrons for research in materials science.

is thus of interest not only for the tools it currently provides to users but also as an inspiration for future directions in neutron scattering.

History

The origins of the Los Alamos high-intensity pulsed neutron source and user program go back to the 1960s when new accelerator technology culminated in the building of the Los Alamos Meson Physics Facility (LAMPF)—a linear accelerator providing 1-milliampere pulses of 800-MeV protons at a repetition rate of 120 per second. Then in the early 1970s the Weapons Neutron Research (WNR) facility was built as an adjunct to LAMPF. The WNR facility

gave the Laboratory an intense neutron source that could be used to obtain nuclear data needed for weapons design.

The original plan for the WNR facility included a unique feature—a 30-meter-diameter ring, called the Proton Storage Ring (PSR), into which a series of proton pulses from LAMPF could be injected. Each injected pulse was to occupy about three-quarters of the ring's circumference. Several thousand proton pulses were to be "overlaid" within the ring, ejected as one high-intensity pulse several times per second, and used to produce neutrons by spallation—a nuclear reaction in which neutrons are knocked out of heavy nuclei by energetic subatomic particles. The Department of

48

Energy weapons-funding agency accepted a proposal to fund the WNR facility in two phases. The first phase included construction of the proton-beam-transport system and the spallation-target area. The ability of a storage ring to produce high-intensity bursts of protons—which in turn would produce high-intensity bursts of neutrons—was recognized, and plans were made to build the PSR in the second phase.

The late Rex Fluharty recognized that the WNR facility had great potential for materials-science research in addition to the nuclear-physics research that had justified the initial funding. Consequently, it was designed as a multi-user facility to serve both scientific communities—nuclear physics, with neutrons of keV energies, and materials science, with neutrons of meV energies. The WNR facility produced neutrons for the first time in May 1977, and routine operation ensued the following year. Shortly thereafter, the Laboratory administration, spurred by George A. Keyworth (then leader of the Physics Division) sponsored the idea of making the WNR facility a national user facility for materials-science research.

In the early 1980s various national scientific committees advised expansion of the WNR experiment area to facilitate hosting a national user program in neutron scattering. The committees also urged that the nuclear-physics and neutron-scattering programs be provided separate neutron-producing targets so each might be optimized. Those recommendations led, in October 1986, to designation of the existing WNR facility as a national facility for neutron scattering. Shortly thereafter funds were made available for construction of a large addition to the experiment hall, a support building containing laboratories and office space, and several new neutron spectrometers. At the same time a new facility was constucted for nuclear-physics experiments. The expanded
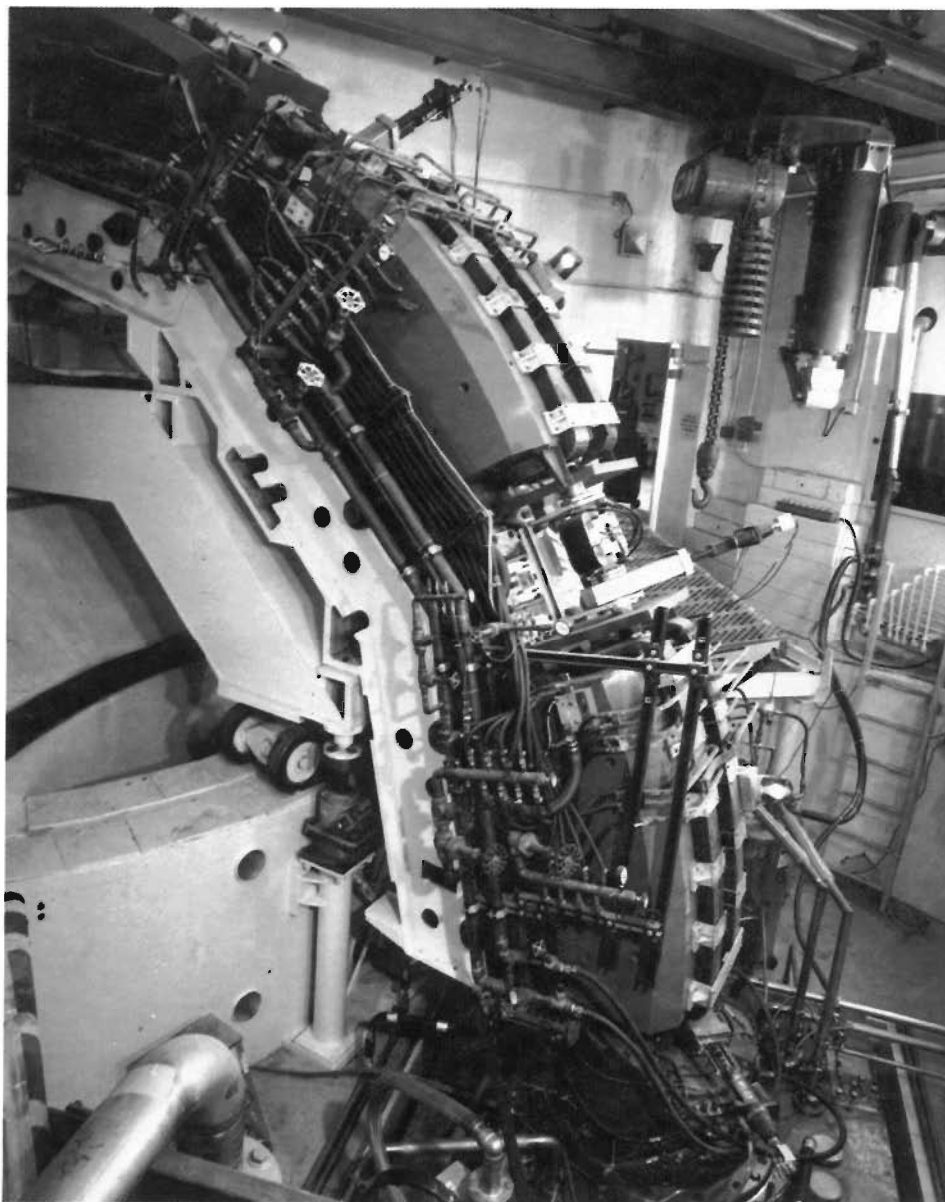
A series of bending magnets guides high-intensity proton pulses from the Proton Storage Ring to the LANSCE spallation target.

materials-science facility was called the Los Alamos Neutron Scattering Center; it was later renamed the Manuel Lujan, Jr. Neutron Scattering Center in honor of long-term New Mexico Congressman Manuel Lujan, Jr.

The basic experimental methods at spallation sources involve the creation of discrete but intense pulses of low-energy neutrons and subsequent measurement of neutron times of flight to determine neutron energies. To understand both the principles of those processes and the technology required for their implementation, we will track a few neutrons from their "birth" in the
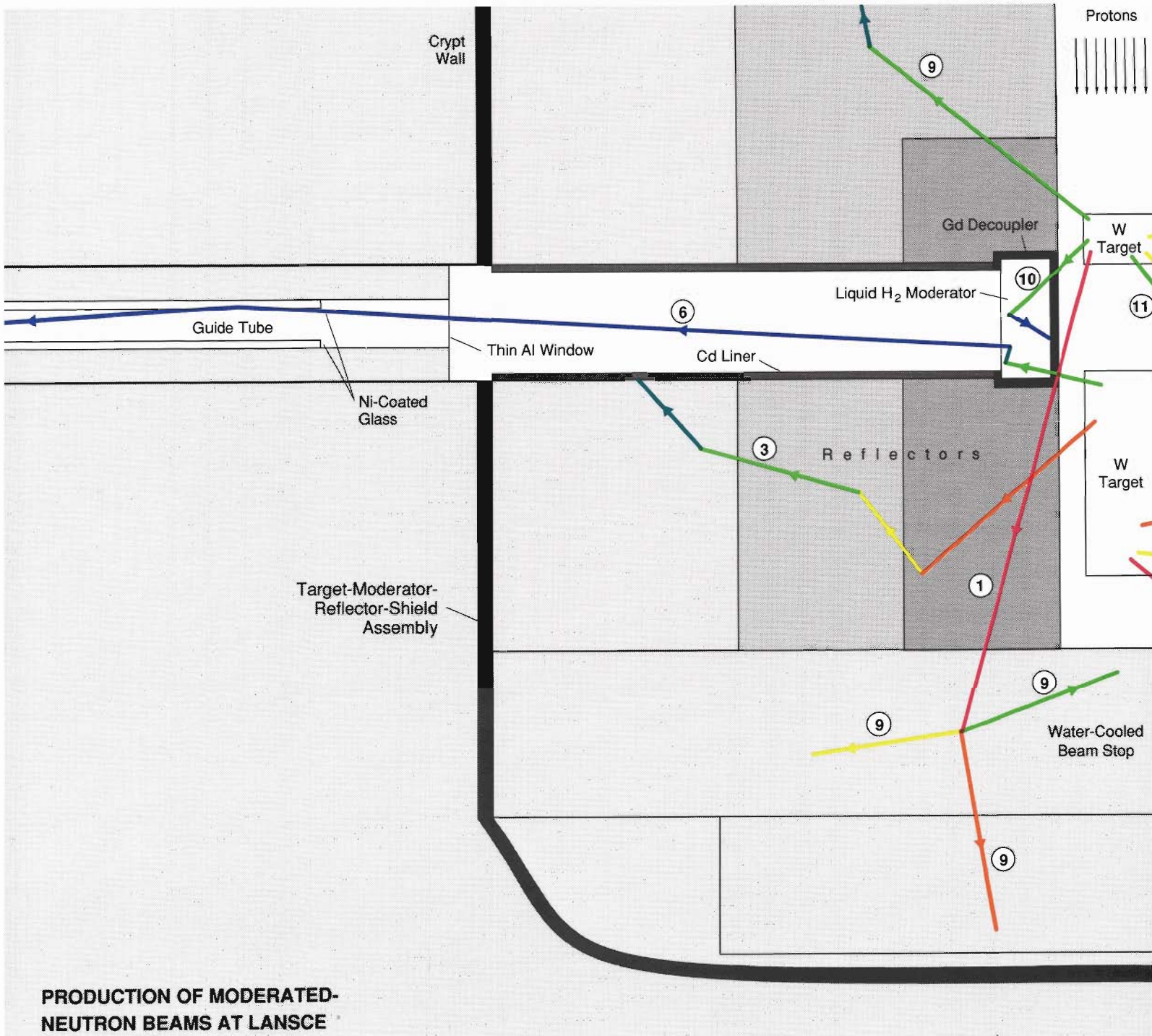
PRODUCTION OF MODERATED-NEUTRON BEAMS AT LANSCE

Fig. 1. This vertical section through the LANSCE target assembly and bulk shield illustrates the hardware and the processes involved in the production of pulsed beams of moderated neutrons. A pulse of spallation neutrons is created when a pulse of 800-MeV protons impinges on the tungsten targets. Some of the spallation neutrons collide with nuclei in moderators or reflectors; their trajectories (arrows) thereby change direction, and their energies (represented by various colors) decrease. The numbered trajectories illustrate points of particular importance.

1. Very-high-energy spallation neutrons leave the target moving roughly in the same direction as the incident proton beam. Such neutrons can generate lower-energy spallation neutrons in the reflectors, the beam stop, and the surrounding shield.

2. Neutrons whose initial trajectories miss a moderator may be scattered into a moderator by a beryllium reflector.

3. Neutrons that reach low energies by wandering around in the reflectors for a long

time are prevented from entering a beam tube by a liner made of cadmium, which absorbs low-energy neutrons.

4. Some thermal neutrons exit the moderator along trajectories that pass through a beam tube and a collimator. Note that the neutron depicted was not absorbed by the cadmium decoupler on the back face of the moderator because of its high initial energy.

5. Thermal neutrons that leave the moderator traveling at large angles relative to the axis

Crypt
Wall

Cd Decoupler

Gd Poison

⑤

④

⑦

Beam Scrapers

Collimator

Beam Tube

Cd Liner

H₂O Moderator

⑧

Reflectors

②

⑨

⑨

1 MeV < E < 800 MeV

10 keV < E < 1 MeV

1 keV < E < 10 keV

1 eV < E < 1 keV

100 meV < E < 1 eV

E < 100 meV

Iron   Nickel   Beryllium
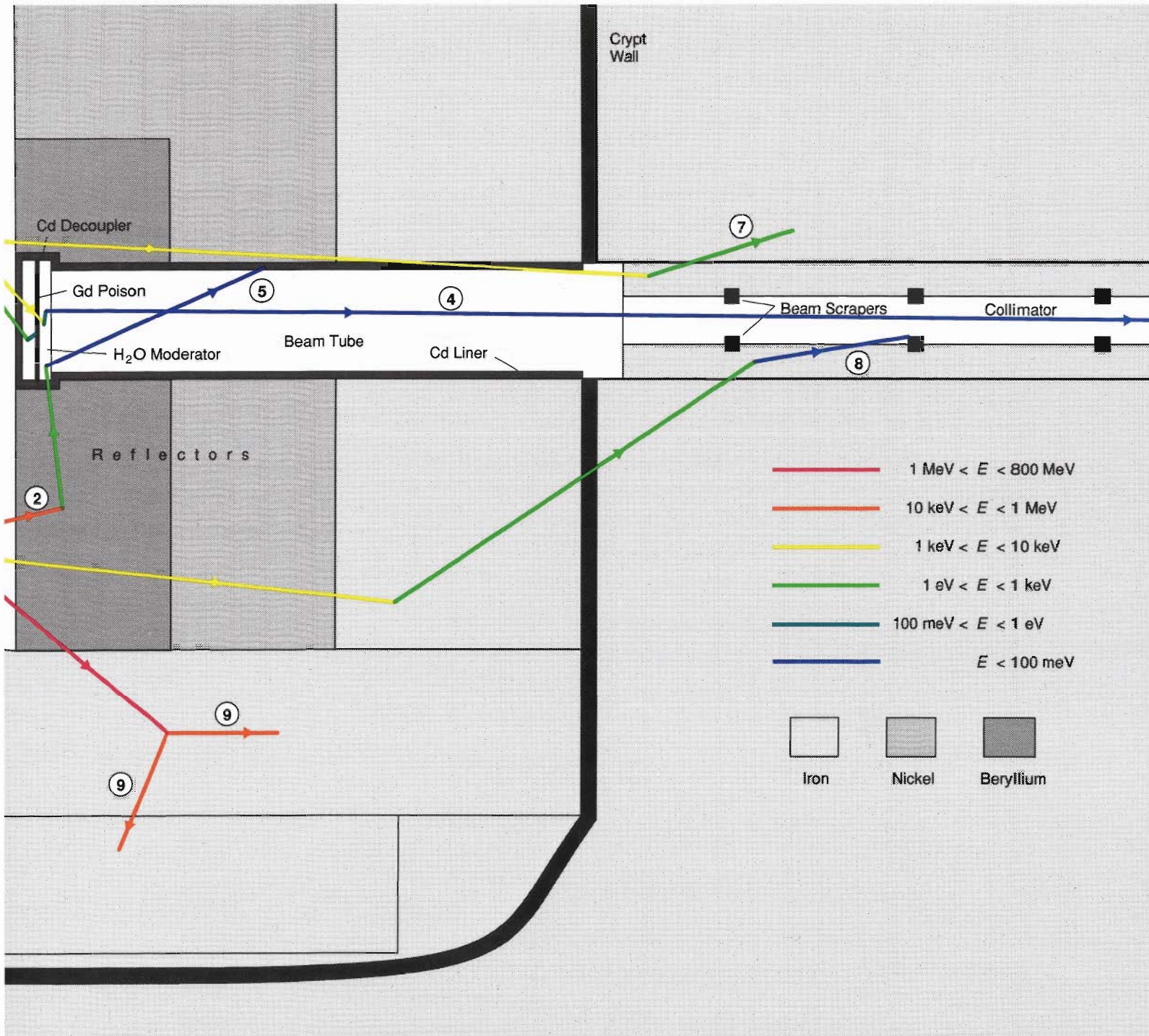
of a beam tube are absorbed by the cadmium walls of the tube.

6. Thermal neutrons that strike a wall of a guide tube at sufficiently small angles are reflected by the tube's nickel coating and therefore proceed down the tube being reflected from opposite sides of the guide tube walls. Such neutrons would have been lost by absorption in the walls of a collimator.

7. The iron walls of the collimator scatter epithermal neutrons and thereby define the

size of the epithermal-neutron beam that exits the collimator.

8. Beam scrapers made of boron carbide absorb thermal neutrons that leak into the collimator and thereby help define the thermal-neutron beam that exits the collimator.

9. Most of the spallation neutrons are never used in an experiment because they are lost in the shielding around the target.

10. Thermal neutrons whose trajectories

would allow them to exit the back face of a moderator are absorbed by the gadolinium decoupler, thus eliminating crosstalk between moderators by ensuring that each moderator feeds only one beam tube.

11. A thin slab of gadolinium poison reduces the thickness of the moderator feeding thermal neutrons to the beam tube and thereby reduces the duration of the thermal-neutron pulse.

spallation target, through their scattering in a sample, to their "death" by absorption in a neutron detector.

The LANSCE Source

Figure 1 shows a vertical cut through the target assembly that generates pulsed neutron beams at LANSCE. The neutron beams emerge into the experiment area from beam tubes joined to holes drilled horizontally through a radiation shield surrounding the crypt—an evacuated steel vessel that holds the target, moderators, and reflectors. Pulses of high-energy (800-MeV) protons from the PSR enter the crypt from above and impinge on the spallation target.

The LANSCE target, a so-called split target, is composed of two 10-centimeter-diameter cylinders of tungsten. The high-mass—and, therefore, neutron-rich—nuclei of tungsten have a high cross section for spallation. Each of the approximately 10^{13} protons per pulse produces an average of 19 spallation neutrons, which have energies ranging from almost zero to nearly 800 MeV. Because neutron-scattering experiments require neutrons with energies below about 1 eV, the spallation neutrons must be reduced in energy, or moderated, before being directed toward a sample. The moderation is accomplished by allowing the neutrons to collide with light nuclei, such as hydrogen, to which they can transfer a substantial fraction of their kinetic energy.

The LANSCE crypt contains four moderators to reduce the energies of the spallation neutrons. One is filled with liquid hydrogen and the others with room-temperature water. (Only two of the LANSCE moderators are shown in Fig. 1; the other two would appear in cross sections other than the one depicted.) The energy spectrum of the neutrons emerging from a moderator depends on the temperature, composition, and thickness of the moderator. For

example, as shown in Fig. 2a, a water moderator yields a spectrum that peaks at about 25 meV, whereas the liquid-hydrogen moderator yields a spectrum that peaks at about 5 meV. The space within the crypt not occupied by the targets and moderators is largely filled with beryllium and nickel reflectors. These materials scatter some neutrons back into the moderators, giving them a second chance to lose energy and to emerge along a beam tube. (Only two of the current sixteen LANSCE beam tubes appear in Fig. 1.)

Figure 1 also shows possible events in the lives of a few spallation neutrons as they are scattered in the moderators and reflectors. When a neutron suffers a collision, both its trajectory and its energy change. Such changes are indicated in the figure by changes in the direction and color of the arrows representing moving neutrons. Most of the neutrons born with energies above 20 MeV start life moving in directions not much different from the trajectory of the proton beam. Less energetic neutrons emerge from the targets more or less isotropically. Some neutrons pass directly into one of the moderators and, after losing energy there, exit from the moderator along a beam tube. Other neutrons are reflected into the moderators by collisions with nuclei in the reflectors. The highest-energy neutrons can induce spallation of nuclei in the reflectors, beam stop, or shielding; that process is indicated in Fig. 1 by emergence of two or more neutrons from a vertex. Some of the secondary spallation neutrons produced by the high-energy neutrons find their way back to the moderators and end up traveling along a beam tube.

The few neutron histories depicted in Fig. 1 clearly show that the moderated neutrons arising from a proton pulse exit a moderator at different times, even though the proton pulse is extremely brief (270 nanoseconds). After all, the paths along which various

neutrons travel differ as do their velocities. Therefore, the duration of the pulse of moderated neutrons is much greater than that of the proton pulse (Fig. 2b). Furthermore, it increases as the average energy of the neutrons decreases. One qualitative feature of the neutron pulse is independent of energy, however: its asymmetry. As Fig. 2b shows, the pulse reaches its peak intensity very quickly and then decays more slowly.

The speed, and hence energy, of an individual neutron can be determined accurately from its time of flight (that is, the time taken by the neutron to travel from a moderator to a detector) only if the time at which the neutron left the moderator is known accurately. Thus, a short neutron pulse allows time of flight to be determined with high relative accuracy. Unfortunately, the pulse duration can be reduced only at the expense of decreased neutron intensity. Thus increasingly accurate time-of-flight measurements are accompanied by increasingly greater statistical errors in the measured neutron scattering law.

What is to prevent neutrons that have been moderated by wandering around for a long time in a reflector from leaking into a beam tube? Such leakage would clearly extend the duration of the neutron pulse and degrade the accuracy of time-of-flight measurements. Closer examination of Fig. 1 shows that some low-energy neutrons stop at the wall of a beam tube, whereas more energetic neutrons keep right on going. That is so because the beam tubes are lined with a material, such as cadmium or boron carbide, that captures low-energy neutrons but does little to stop the more energetic ones. A similar neutron-absorbing material—a "poison"—placed inside a moderator reduces the effective thickness of the moderator for thermal neutrons and thereby reduces the duration of the thermal neutron pulses. The closer the poison is to the exit face of the moderator, the

(a) Energy Spectra

(b) Time Spectra
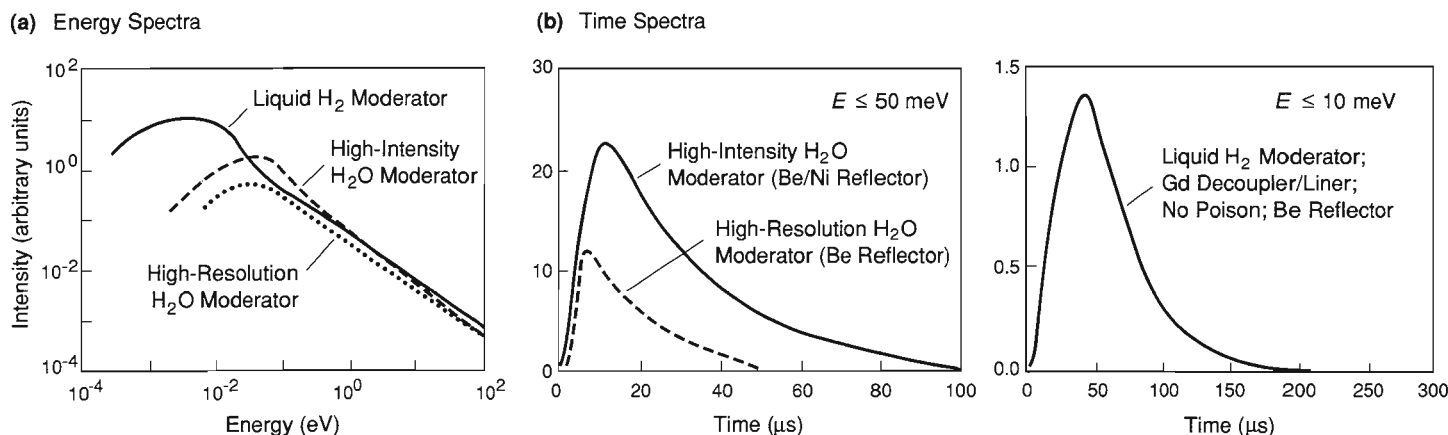
CHARACTERISTICS OF LANSCE NEUTRON PULSES

Fig. 2. (a) The energy spectrum of the neutron pulses emerging from any of the LANSCE moderators is very broad. Details of the spectrum vary with the composition, temperature, and thickness of the moderator. Note the relatively small difference between the energy spectra of neutrons from the high-intensity water moderator and from the high-resolution water moderator. (The primary physical difference between those moderators is the location of the poison.) (b) The time spectrum of the LANSCE neutron pulses is asymmetric and much broader than the time spectrum of the proton pulses, which has a full width at half maximum of 0.27 microseconds. Like the energy spectrum, the time spectrum varies with the composition, temperature, and thickness of the moderator. Location of the poison in the high-resolution water moderator closer to the exit face decreases the FWHM of the time spectrum and thereby increases the accuracy of the time-of-flight (and thus energy) measurements. However, the increased accuracy in energy is bought at the price of lower intensity and hence greater statistical uncertainties in the measured signal.

shorter is the neutron pulse and the lower its intensity. The location of the poison is the feature that primarily distinguishes the two "high-intensity" water moderators at LANSCE from the single "high-resolution" water moderator. (The liquid-hydrogen moderator contains no poison.)

Figure 1 illustrates another important point: Production of moderated neutrons is extremely wasteful. Most of the spallation neutrons wander off into the reflectors and the shielding without ever encountering a moderator or a beam tube. By changing the relative arrangement of the various moderator, reflector, and neutron-absorbing materials, it is possible to increase somewhat the neutron flux emerging from the moderators. For example, the LANSCE split target yields a higher neutron flux than a single target because it simultaneously feeds neutrons into the moderators from above and below. We are fortunate to

have access at the Laboratory to the best computer codes—not to mention a few Crays—to optimize the configuration of the target assembly. The optimization cannot be performed analytically; Monte Carlo computations are the only recourse for tracking neutrons and improving the performance of spallation neutron sources.

The LANSCE split target is unique worldwide, thanks to the conceptual design of Gary Russell and his colleagues. The assembly was installed in August 1985 and has since operated reliably with no target or moderator changes. What makes the LANSCE source so special is that it is very efficient and very "clean": efficient because the moderators are fed with neutrons by both the upper and lower targets, and clean because only a small fraction of the spallation neutrons escape along a beam tube without first being moderated. The latter point is important because unmoder-

ated neutrons, which are not useful for neutron scattering, degrade experimental results by contributing background signals. More important, high-energy neutrons can damage living cells and are better kept within the crypt.

The high-energy spallation neutrons produced at LANSCE would pass right through the biological bulk shield at a reactor, which is typically made of concrete containing a neutron-absorbing material such as boron. The LANSCE crypt is surrounded by a 3.7-meter-thick bulk shield containing a core of iron encircled by a layer of concrete. The bulk shield, in combination with the nickel reflectors within the crypt, reduces the radiation exposure of researchers to very low levels.

Neutron Spectrometers

All neutron-scattering instruments have certain common requirements.

Each needs a method for piping neutrons from the source to a sample and a system for detecting the neutrons that are scattered from the sample. The following sections describe some of the hardware used to achieve those ends.

Collimators. As Fig. 1 shows, neutrons leave the moderator surface in all directions. Only those traveling within a narrow cone of angles along a beam tube can reach a sample outside the bulk shield. To further decrease the size and angular divergence of the beam of neutrons exiting a beam tube, it can be fed through the bulk shield along a collimator. Jutting out from the walls of a typical collimator are boron carbide scrapers, which define a thermal neutron beam by absorbing those thermal neutrons that strike them. Boron carbide does not absorb neutrons with energies above a few keV; those neutrons are attenuated by the iron walls of the collimator. The very few neutrons with energies greater than 20 keV that escape the crypt form a beam somewhat larger than the thermal neutron beam— a halo. Because the halo would create undesirable background, it is absorbed by heavy shielding material outside the bulk shield. It is worth noting that a collimator is designed so that a straight line drawn between any point along its iron walls and the scattering sample passes through at least one of the boron carbide scrapers. Thus the pulse of neutrons incident on the sample is not broadened by thermal neutrons that leak through the collimator walls.

The cross-sectional area and shape of the neutron beam defined by a collimator are chosen to match the requirements of a neutron spectrometer. For example, the samples studied with the Single-Crystal Diffractometer are small. Consequently, the beam-defining apertures in its collimator are small. The reflectometer, on the other hand, uses two narrow beams whose shapes are dictated by the

The Neutron Powder Diffractometer was the first instrument to be installed in the new addition to the experiment hall. The *d*-spacing resolution of this diffractometer is better than that of any other instrument of its type in the United States.

need to reflect them at very small angles from the flat surface of a sample.

Many of the collimators within the LANSCE bulk shield can be flooded with mercury, which acts as a shutter, closing off the beam and allowing access to the sample. An interlock

system ensures that no one has access to a spectrometer unless its shutter is closed. Thwarting of any one interlock prevents delivery of protons to the spallation target and thus in effect turns off all the beams.

Additional profiling of the neutron

beam takes place outside the bulk shield in each instrument according to individual design. For example, the Neutron Powder Diffractometer has five pairs of variable apertures in the incident beam line that allow the user to tailor the beam to the sample size. Neutrons that would miss the sample would produce unwanted background, so it is best to remove them from the beam by placing the apertures as far "up beam" of the sample and detectors as possible. Another instrument, the Low-Q Diffractometer, has a beam scraper that offers a single aperture or multiple apertures. The multiple apertures can be used to increase the neutron intensity when a weakly scattering sample is being studied. The sample must then be sufficiently large to intercept the multiple beams, which converge at the detector rather than at the sample.

Guide Tubes. The astute reader may marvel at the inefficiency of the collimator shown in Fig. 1. Only those neutrons that leave the moderator headed almost directly for the spectrometer exit the collimator. Coaxing neutrons headed in the "wrong" directions into moving toward a spectrometer is not generally possible, but there is a trick we can play—we can install a neutron guide tube rather than a collimator. A guide tube is usually a rectangular-section tube of thick glass. The inner walls of the tube are highly polished and coated with a thin, smooth layer of nickel or, even better, isotopically pure ^{58}Ni. Neutrons that make a sufficiently small angle with a wall of the guide tube are totally reflected from the wall and therefore stay within the guide, headed toward a neutron spectrometer. The "sufficiently small" angle, in degrees, is about one-tenth of the neutron wavelength in angstroms. For example, the critical angle for thermal neutrons with a typical wavelength of 2 angstroms is 0.2 degrees. The increased neutron intensity
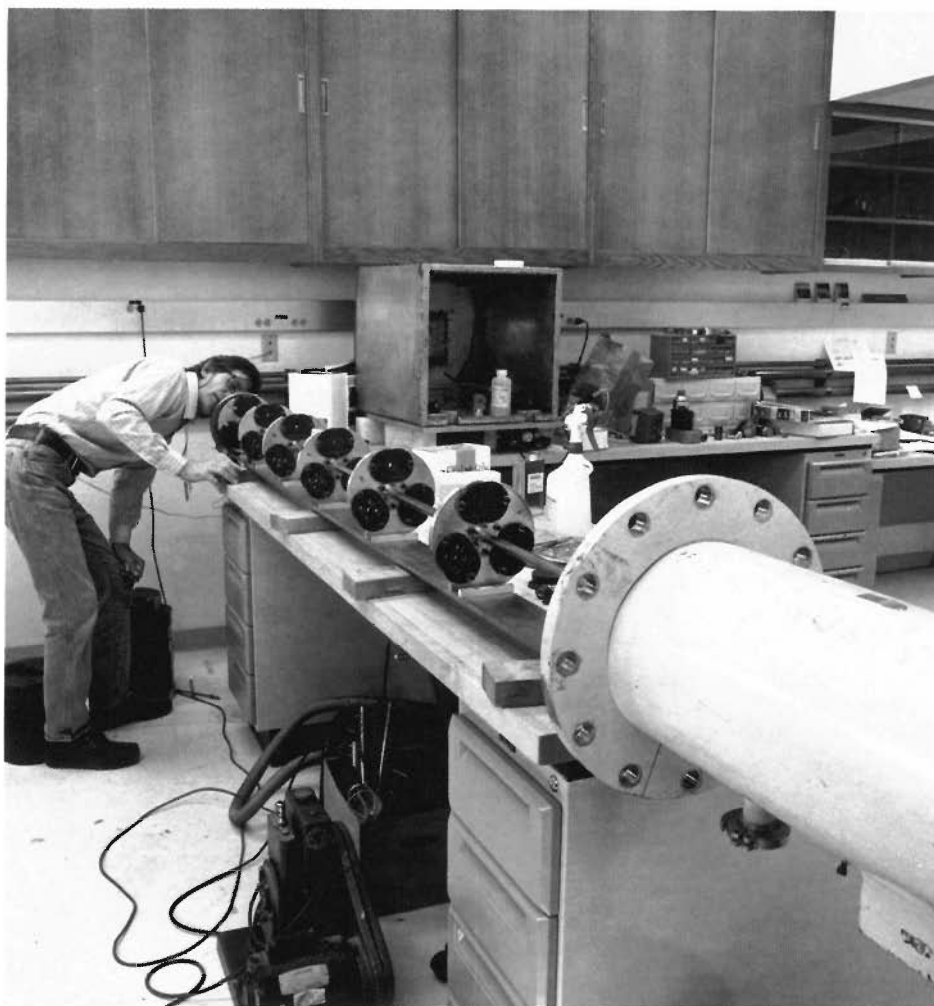
Rodney Hardee, instrument technician at LANSCE, peers along the beam path defined by the set of single holes in the top series of beam scrapers for the collimator of the Low-Q Diffractometer. (The series of beam scrapers at the 4 o'clock position also have single holes.) The series of beam scrapers at the 8 o'clock position illuminates the sample with a set of five beams rather than a single beam.

provided by a guide tube is worth its high cost, several thousand dollars per meter. The intensity gain is especially high for long-wavelength neutrons because their critical angles are greater.

Guide tubes are 10 to 100 meters long. They are evacuated to prevent neutrons from being scattered out of the tube by air. Guides have been used at reactors for some time—the first were in place at the Institut Laue

Langevin in France in the early 1970s. The National Institute of Standards and Technology is currently installing guide tubes at its Washington, D.C. reactor, and a 25-meter guide will deliver ten times more 2-meV neutrons to the new LANSCE backscattering spectrometer than would a conventional collimator.

T-zero Choppers. Little can be done to change the fact that each pulse of
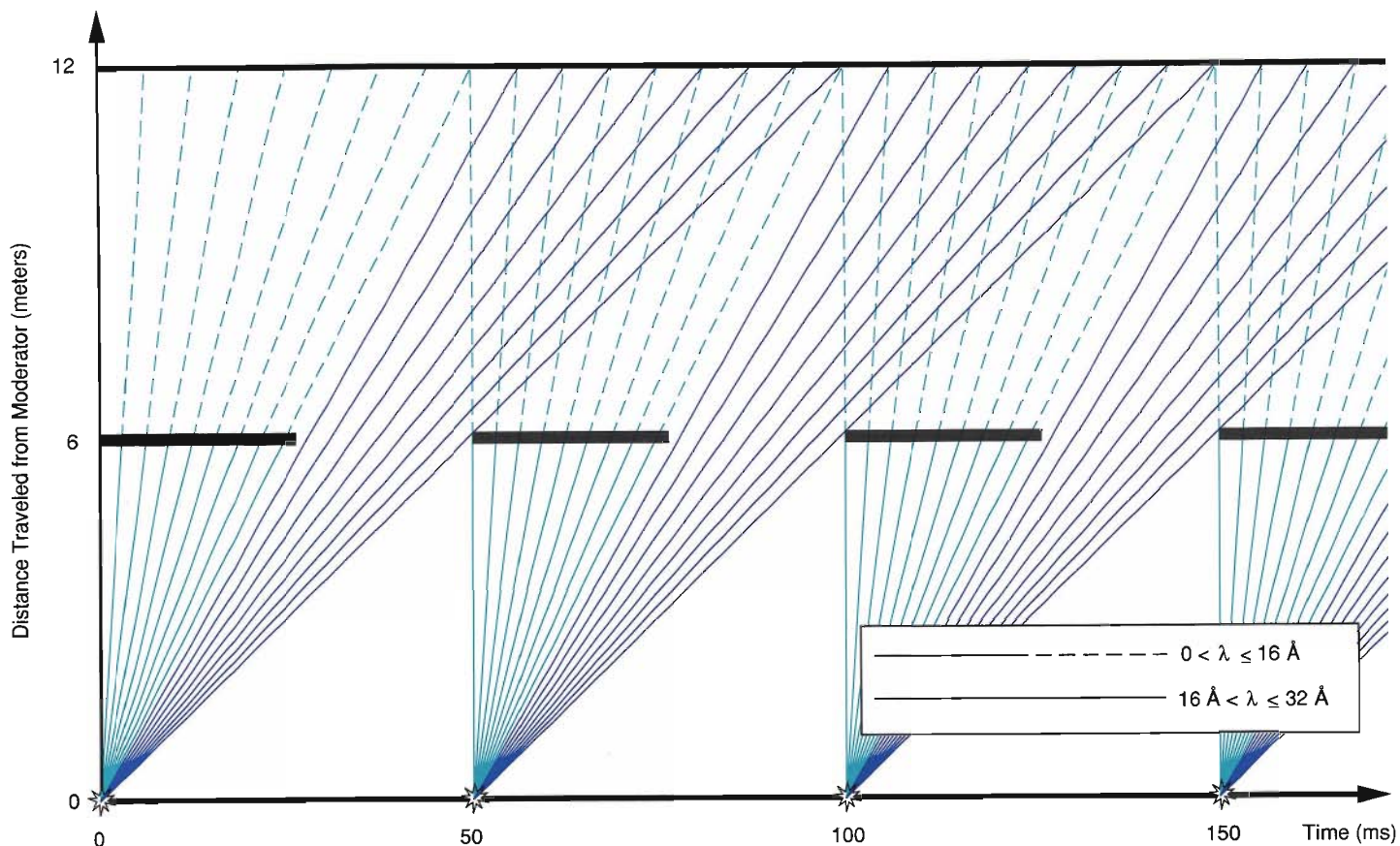
FRAME OVERLAP

Fig. 3. The neutrons in each moderated pulse begin their flight to a detector at essentially the same time (within a small fraction of a millisecond). However, because they all have different energies, as time passes they spread out along the course, reaching the same distance from the moderator at different times after departure. That situation is depicted graphically here in plots of distance traveled versus time for neutrons with various energies. Each "explosion" on the time axis represents a pulse of neutrons emerging from a moderator. Note that fast neutrons from, say, the second pulse can reach a detector 12 meters distant at the same time as slow neutrons from the first pulse. To prevent such "frame overlap," which fouls up the time-of-flight measurements, a spectrometer can be equipped with a frame-overlap chopper (See Fig. 4). When the chopper is rotated at 10 times per second (half the repetition rate of the proton pulses) and at a certain phase relative to the proton pulses, it blocks neutrons with energies corresponding to wavelengths less than 16 angstroms. That mode of chopper operation is represented here by a thick horizontal line at a distance of 6.25 meters. Changing the phase of the chopper by 90 degrees (moving the thick horizontal line to the left or right a distance equal to its length) blocks neutrons with energies corresponding to wavelengths between 16 and 32 angstroms.

useful thermal neutrons emerging from a guide tube or collimator is preceded by a bright flash of high-energy neutrons and gamma rays generated when the proton pulse strikes the spallation target. One way of reducing the background caused by that radiation is to attenuate it with a T-zero chopper, a device so named because the zero of time for each neutron pulse is defined as the moment at which the proton pulse strikes the target. The high-energy neutrons and gamma rays travel so fast that they arrive at a spectrometer essentially at the same time as the proton pulse hits the target. A T-zero chopper, a 30- to 40-centimeter-thick slug of nickel alloy, rotates into the neutron beam in synchronization with

the proton pulses and "chops" out both gamma rays and high-energy neutrons. Because the high-energy neutrons are scattered rather than absorbed, the chopper is placed inside a heavily shielded cave. The pulse of useful neutrons arrives at the chopper after the nickel slug has moved out of the beam path and passes out of the cave through a small window.

Frame-Overlap Choppers. Stripped now of background-creating high-energy neutrons, each pulse of neutrons flies on toward a sample and a detector. Initially tightly bunched, the neutrons gradually spread apart in space because their speeds differ. The spreading is depicted graphically in the distance-time diagram

of Fig. 3. The "explosions" on the time axis represent neutron pulses leaving the moderator. Examination of Fig. 3 reveals that fast neutrons in one pulse will catch up with slower neutrons in the preceding pulse before the latter have reached a neutron detector, provided the flight distance is sufficiently great. This "frame overlap" clearly gets worse the longer the flight distance. The hare always catches the most motivated tortoise if the course they run is long enough! Because a neutron detector is sensitive only to the arrival of a neutron and not to its energy, frame overlap leads to assignment of erroneous times of flight (and hence energies) to neutrons with energies sufficiently different from the average. The frame-overlap

T-ZERO AND FRAME-OVERLAP CHOPPERS

Fig. 4. The LANSCE reflectometer has both a T-zero chopper and a frame-overlap chopper. The former scatters the burst of high-energy neutrons and gamma rays (red through yellow) that precedes the pulse of moderated neutrons; the latter absorbs lower-energy neutrons that would cause frame overlap as they spread out along the flight path because of their different energies (see Fig. 3). By $t = 10$ μs the T-zero chopper, which rotates $360°$ every 50 ms, has scattered neutrons with energies down to 1 MeV. By $t = 15$ ms the T-zero chopper has scattered neutrons down to 300 meV and rotated out of the beam. Energetic thermal neutrons (green and aqua) have reached and are being absorbed by the frame-overlap chopper, which rotates $180°$ every 50 ms. The remaining neutrons in the pulse have spread out along the flight path. By $t = 40$ ms the frame-overlap chopper has absorbed neutrons down to 0.3 meV (aqua) and rotated out of the beam, allowing very slow neutrons (blue) to proceed on to the sample. By changing the phase of the frame-overlap chopper, faster neutrons, rather than slower neutrons, can be allowed to reach the sample. Here pale color represents low beam intensity.

problem is reduced significantly if the repetition rate of the neutron pulses is low, as can be seen by removing every other "explosion" on the time axis of Fig. 3. LANSCE thus has an advantage over other pulsed neutron sources because its repetition rate is lower (20 per second compared with 30 to 50 per second). However, even at a pulse repetition rate of 20 per second, a 12.5-meter flight distance leads to frame overlap if the bandwidth of the neutron wavelengths is greater than 16 angstroms.
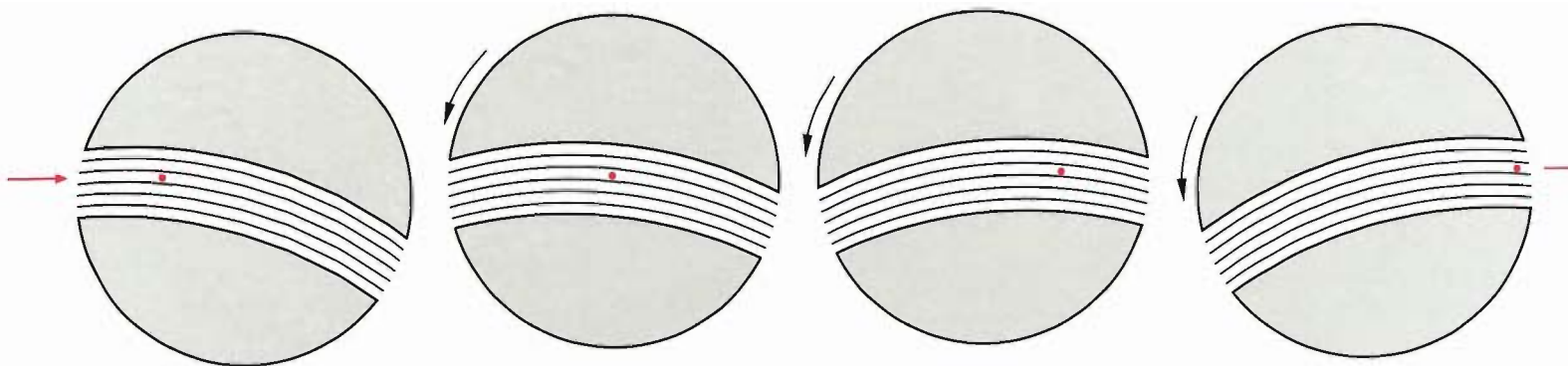
One way of avoiding frame overlap is to install a frame-overlap chopper, a device that closes for a prescribed amount of time during each pulse to prevent neutrons in a particular wavelength band from reaching the detector. The frame-overlap chopper on the reflectometer at LANSCE has a shape resembling that of a bow tie and is fabricated of a neutron-absorbing material. When the chopper is rotated at half the pulse repetition rate (10 times per second), its phase can be adjusted so that neutrons with wavelengths less than 16 angstroms are transmitted and neutrons with wavelengths between 16 and 32 angstroms are absorbed. Changing the phase of the chopper appropriately reverses the situation: Neutrons with wavelengths less than 16 angstroms are blocked, and neutrons with wavelengths between 16 and 32 angstroms are transmitted. Both wavelength bands can be used profitably on the reflectometer. The observant reader may have noticed that the chopper does not prevent frame overlap between fast neutrons in the $(n+2)$th pulse and very slow neutrons—those with wavelengths beyond 32 angstroms—in the nth pulse. Therefore, the very slow neutrons are reflected out of the beam by silicon wafers coated with nickel. Action of a T-zero chopper and a frame-overlap chopper is illustrated in Fig. 4.

Energy Selectors. Although the neutron pulses transmitted through a frame-

ENERGY-DEFINING CHOPPER

Fig. 5. A neutron (red point) traveling in the direction indicated by the red arrow can, if it has the proper speed, pass through the center of a curved slit in a rotating disk. Shown here are four snapshots of the neutron making its passage through one of several slits in an energy-defining chopper. (To an observer moving with the neutron, the slit appears to open up to allow passage.) Neutrons that do not have the proper speed strike the walls of the slit and are absorbed. A chopper of this type can be used to select a narrow band of neutron energies from either the incident or the scattered beam. Different neutron energies can be selected by changing the speed of rotation of the chopper.

overlap chopper have a reduced energy bandwidth, that bandwidth is still very broad. Such pulses can be used without further ado in elastic-scattering experiments because an elastically scattered neutron does not change energy during scattering and therefore its measured time of flight provides all the energy information required. In contrast, inelastic scattering experiments require knowledge of the energies of a detected neutron both before and after scattering. If the initial energy of an inelastically scattered neutron is known, say because only neutrons whose energies lie within a narrow band have been allowed (selected) to reach the sample, then its final energy can be calculated from its measured time of flight. Alternatively, if the final energy of an inelastically scattered neutron is known because only neutrons whose energies lie within a narrow band have been allowed to reach the detector, then its initial energy can be calculated from its measured time of flight.

Several methods are available for selecting, from either the incident or the scattered beam, neutrons whose energies lie within a narrow band. One method uses an energy-defining chopper, which can be thought of as a window

that opens briefly to allow passage of neutrons with a particular speed. Such a "window" is usually achieved by rotating an assembly of curved slits in the beam line. The curvature of the slits and the speed at which the assembly rotates determine the energy of the neutrons that pass through the slits. Other neutrons strike the walls of the slits and are absorbed. The duration of the neutron pulse transmitted by the chopper varies with the width of the slits and the speed of rotation. Figure 5 illustrates the operating principle of such an energy selector, one of which is installed, close to the sample, in the incident beam line of Pharos, a new inelastic scattering instruments at LANSCE. (The instrument is named after the ancient lighthouse at Alexandria.) Magnetic bearings allow high-speed rotation (1200 times per second) of the chopper, and the chopper slit package is made from a strong composite material containing neutron-absorbing boron fibers.

One of the most common techniques for selecting a monochromatic beam of neutrons—at both accelerator and reactor sources—makes use of Bragg diffraction from a single crystal. The orientation of the crystal determines the

energy of the neutrons diffracted toward a sample or a detector. When mounted in the scattered beam line, an energy selector of this type is called an analyzer. Each of the analyzers available for use with the LANSCE Constant-Q Spectrometer consists of an array of single crystals rather than one single crystal. All of the crystals have the same orientation, and their identical orientation is preserved when the crystals are rotated to allow diffraction from a different set of atomic planes (and hence detection of scattered neutrons with a different energy). (Realizing this feat with a single analyzer crystal is not possible because it would require a crystal of unattainable size.) Such an array can "analyze" the neutrons scattered by the sample through a wide range of angles. To increase the scattered-neutron intensity at the detector, dislocations have been introduced into each crystal by squeezing it in a special press. The squeezed crystal diffracts as if it were composed of many perfect crystal blocks ("mosaic blocks") that are slightly misoriented relative to one another. That is, it diffracts neutrons whose energies lie within a finitely (rather than infinitely) narrow

band. Thus a greater number of neutrons are recorded by the detector. The increase in measured intensity is accompanied, however, by a decrease in energy resolution because the misoriented mosaic blocks diffract slightly different energies. Such compromises are not uncommon: Neutrons are scattered so weakly by matter that obtaining a measurable signal often requires being satisfied with decreased energy resolution.

Another method for selecting neutron whose energies lie within a narrow range makes use of polycrystalline "filters" that diffract neutrons with certain energies. The most commonly used filters are beryllium and beryllium oxide, cooled by liquid nitrogen, and pyrolytic graphite. The beryllium and beryllium oxide filters, used widely in connection with cold neutron beams, are almost transparent to neutrons with energies less than 5.2 meV and 3.7 meV, respectively, and strongly diffract neutrons with higher energies. Alternate detector elements of the LANSCE Filter-Difference Spectrometer are preceded by beryllium and beryllium oxide filters. Subtracting the signals in the two sets of detector elements yields the scattering law of the sample for neutrons whose final energies lie between 3.7 meV and 5.2 meV.

Sample-Environment Equipment. Most early neutron-scattering experiments were carried out on samples under ambient conditions. Because the questions being addressed by neutron scattering have become increasingly sophisticated, equipment for providing a wide range of sample environments has become increasingly prevalent. Temperatures down to 1.4 kelvins or up to a few hundred degrees Celsius are now considered routine and are available on almost all neutron spectrometers. Also readily available are cells in which single crystals or powders can be subjected to pressures up to a few tens of
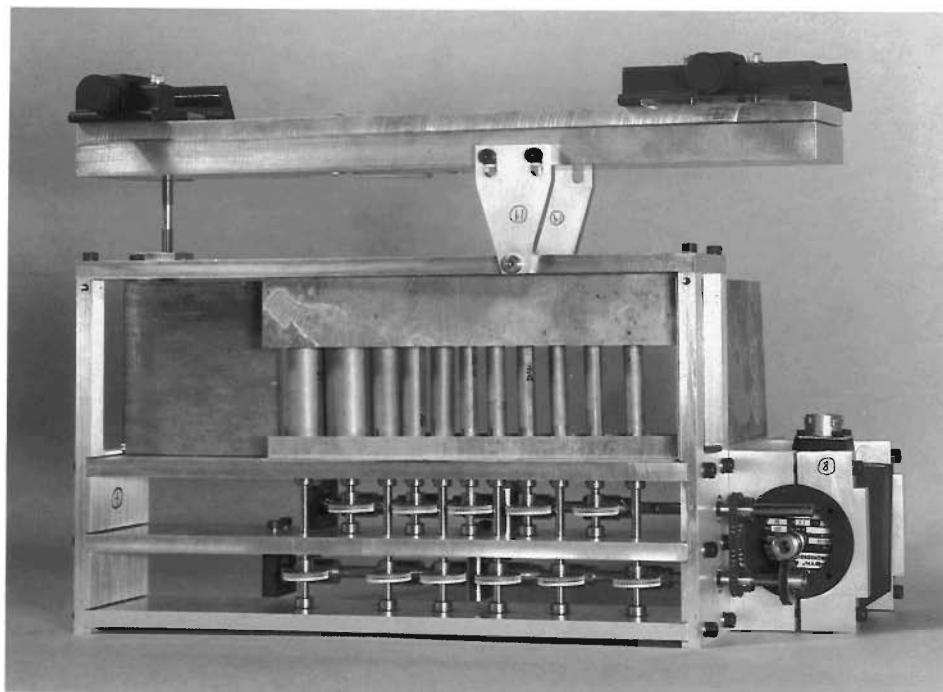
Each cylinder in this analyzer for the Constant-Q Spectrometer is a single crystal of germanium. All the crystals have the same orientation, and therefore the analyzer diffracts, toward an array of detector elements, neutrons that the sample has scattered through different angles but to the same energy. The diameters of the crystals vary because different thicknesses are required for optimum diffraction through different angles: Thicker crystals are required for smaller scattering angles. Note the mechanism for rotating the crystals in unison to orient other atomic planes perpendicularly to the plane of the analyzer. A similar analyzer containing single crystals of copper is available. It allows selection of a different set of energies and energy resolutions.

thousands of atmospheres. The pressure cells can often be mounted in cryostats and cooled to liquid-helium temperature. In addition, at some neutron-scattering centers high magnetic fields can be applied to samples.

Extreme sample environments are often easier to implement at an accelerator source than at a reactor source. Use of a white incident neutron beam implies that complete information can frequently be obtained by examining only those neutrons scattered through a narrow range of angles. For example, obtaining data that cover lattice spacings between 0.3 and 5 angstroms with the LANSCE Neutron Powder Diffractometer requires access to scattering angles

between 78 and 102 degrees in the 90-degree detector bank, whereas obtaining data that cover lattice spacings between 1 and 5 angstroms with a reactor instrument fed 2-angstrom neutrons requires access to scattering angles between 23 and 160 degrees. Thus the paraphernalia required to subject a sample to extreme temperature or pressure need not, at a spallation source, be equipped with large neutron-transparent windows. Such windows are required at a reactor source and are not always easy to implement because few structural materials are transparent to neutrons (aluminum is one of the best).

Some types of sample-environment equipment are specific to particular

types of spectrometers. For example, an active vibration-isolation system is installed on the LANSCE reflectometer. It serves to eliminate ripples on the surface of a liquid from which neutrons are being reflected. The liquid is usually contained in a Langmuir trough, which is rather like a Teflon-coated baking pan. The pan is equipped with a lid to maintain constant vapor pressure above the liquid and a boom to allow compression of monolayer films on the liquid surface.

A special apparatus that may be found on instruments designed to provide small-angle-scattering data is a Couette, or shear, cell—a device in which a fluid can be subjected to shear by being confined within the gap between two coaxial cylinders rotating at different speeds. Analysis of small-angle-scattering data for the fluid yields information about shear-induced structural changes. The shear cell that will soon be installed on the LANSCE Low-Q Diffractometer is made of fused silica, a material that adds little background to the small-angle-scattering signal. At a maximum torque of 5 newton-meters, a shearing rate of 10,000 per second will be achieved when the gap between the cylinders is 0.3 millimeter.

Detectors. Almost all neutron detectors are so-called gas detectors containing ^{3}He, one of the few isotopes that have a high cross section for absorbing neutrons. When a ^{3}He nucleus absorbs a neutron, a proton and a triton are formed. Both have considerable energy and are thus capable of knocking electrons out of nearby ^{3}He atoms. The electrical pulse that occurs as the electrons migrate to an anode maintained at a high voltage is detected by an electronic circuit. The neutron is then said to have been counted (or recorded or detected). The ^{3}He gas is generally contained within a thin-walled, small-diameter tube of aluminum or stainless steel.
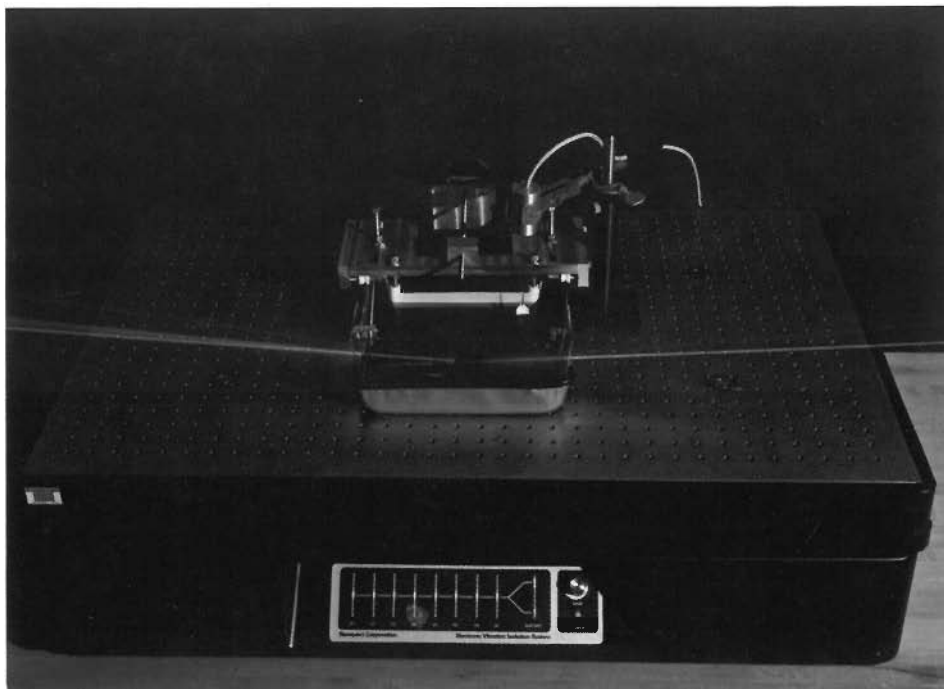
A Langmuir trough is used to contain a liquid whose surface is being studied by neutron reflectometry. The vibration-isolation system on which the trough sits eliminates ripples on the surface. Laser beams are used to simulate the neutron beams during alignment of the sample.

A wire strung along the tube axis serves as the anode; the tube wall serves as the cathode. The efficiency of a neutron detector (the fraction of the neutrons entering the detector that are counted) increases with the number of ^{3}He atoms in the tube, that is, with the product of gas pressure and tube volume. In addition, because the cross section of ^{3}He for absorbing neutrons increases with decreasing neutron energy, so also does the detector efficiency. Typical efficiencies are not far from unity.

Each time a proton pulse is fired at the LANSCE spallation target, an electronic clock is started. The clock is read when the electrical pulse created by absorption of a neutron is detected and that time is recorded by the data-acquisition system as the measured time of flight of the neutron. Because the electrical pulse occurs at some time after the neutron arrives at the detector,

the measured time of flight is longer than the true time of flight. However, the delay between the neutron's arrival and detection of the electrical pulse is about a microsecond, whereas the true time of flight of the neutron is on the order of several milliseconds. Therefore, the delay introduces a negligible error in the measured time of flight.

The simplest neutron detectors do not provide information about where a neutron is absorbed. More sophisticated detectors, called linear-position-sensitive detectors, do provide such information in one dimension, namely the axial dimension. Therefore a single linear-position-sensitive detector can be used to count neutrons scattered through different angles. Position sensitivity can be extended to two dimensions by using a hollow disk to contain the ^{3}He instead of a tube. A grid of anode wires strung across a circular cross section of

the disk provides the two-dimensional position information.

Many of the neutron detectors used at LANSCE are of the simplest type. Their size varies between 5 mm in diameter by 150-mm long to 25 mm in diameter by 1 meter long. Linear-position-sensitive detectors varying in length from 200 mm to almost 1 meter are installed on several LANSCE spectrometers. The accuracy with which those detectors can determine the position at which the neutron was detected varies between 1.8 mm and about 5 mm. Two LANSCE diffractometers are equipped with a two-dimensional position-sensitive detector. The one on the Single-Crystal Diffractometer covers a square area 25 centimeters on a side, whereas the one on the Low-Q Diffractometer covers a circular area more than four times as large—and is considerably more expensive!

Each instrument has a special gas detector called a beam monitor, which contains a low density of ^{10}B nuclei (as BF_3) and detects only about one neutron out of every hundred thousand. The beam monitor provides the energy spectrum of the neutrons incident on a sample. That information is used when the neutron scattering law for the sample is calculated from the data provided by the detectors in the scattered beam.

Data-Acquisition and -Analysis Systems. Electronic signals from the neutron detectors on each LANSCE spectrometer are fed into a data-acquisition system solely dedicated to that spectrometer. Because most of the spectrometers are equipped with a detector system containing more than one detector element, the number of neutrons recorded per second can be very large—up to 100,000 on the High-Intensity Powder Diffractometer, for example. The LANSCE data-acquisition systems are designed to handle such a barrage of information, storing both the time of
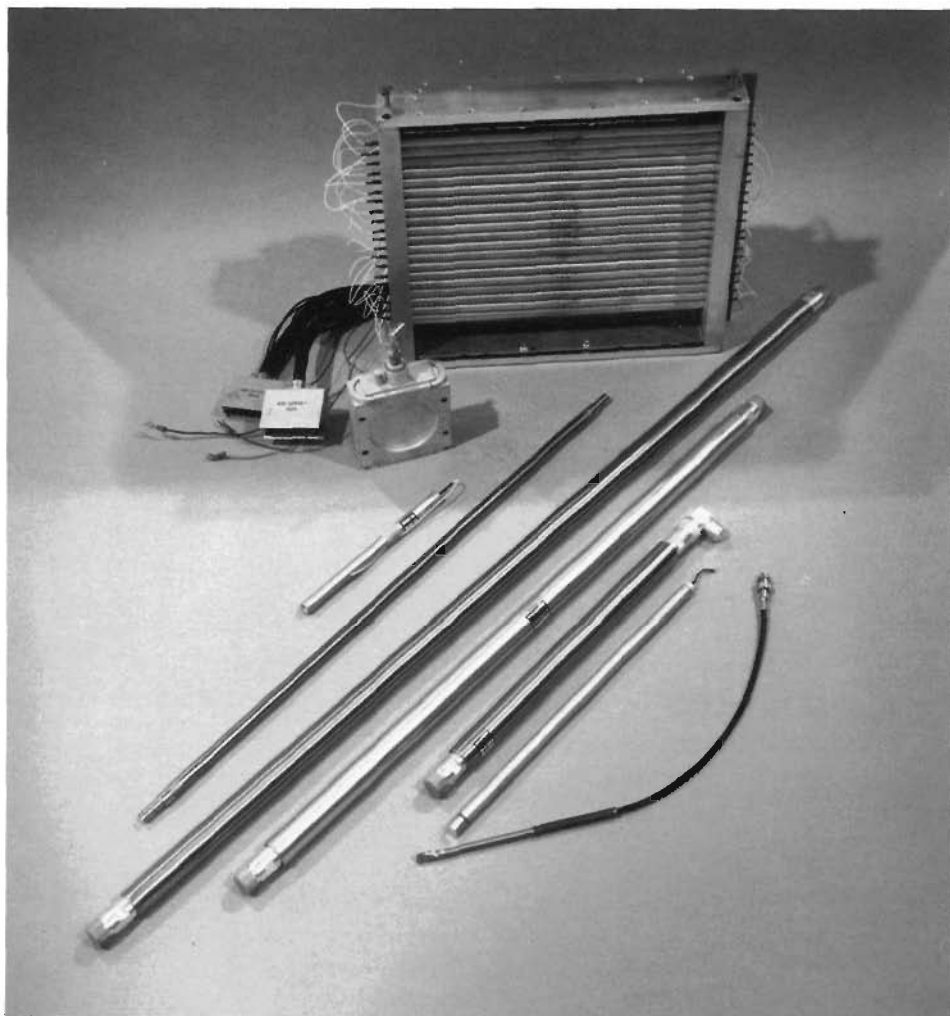
This display of gas detectors used at LANSCE includes, from bottom right, a 6-mm by 100-mm detector from the 5-degree bank of the High-Intensity Powder Diffractometer; a 12-mm by 300-mm detector from the Neutron Powder Diffractometer; a 25-mm by 300-mm detector; a 25-mm by 600-mm detector; a 25-mm by 1-meter detector from Pharos; a 12-mm by 600-mm linear-position-sensitive detector; a 12-mm by 150-mm detector from the Constant-Q Spectrometer; a beam monitor; and the array of 12-mm by 300-mm linear-position-sensitive detectors from the Surface Profile Analysis Reflectometer.

flight and detector element (scattering angle) corresponding to each detected neutron. Each data-acquisition system is based on an electronic transmission line called a FASTBUS. Along this computer "highway" thirty-two "lanes" of signals are directed according to internationally standard protocols. A FASTBUS is capable of handling 2,000,000 events

per second.

Data from each LANSCE instrument are delivered to a VAX-station computer with 9 megabytes of memory; data files are written on a local magnetic 300-megabyte disk. Data are then archived on a 150-gigabyte optical disk. (Such a disk could store the entire contents of *Encyclopaedia Britannica* 600 times).

Thirty data-analysis systems are available, each based on a VAX 3100 workstation and providing color graphics and dial-in capabilities.

The veritable deluge of data that pours out of a spectrometer at a spallation source can be an advantage for survey experiments. On the other hand, knowing what measurements to make next is very difficult when the full import of data already obtained has not yet been comprehended. Therefore, high-speed computers are needed to convert raw data into neutron scattering laws, as are advanced computer graphics and image-processing software to provide insightful views of data and results of calculations.

The LANSCE User Program

Because neutrons are much in demand and available only at a few specialized facilities, the experimenter who needs neutrons is forever chasing after them. He or she often applies well in advance to more than one facility to ensure getting some precious beam time on an instrument. Beam time is of course available at LANSCE only when LAMPF is up and running, and its running period is limited by funding constraints to about five or six months a year. LANSCE issues a call for proposals before each scheduled running period. In response, scientists from universities, industry, and other research facilities around the world submit their proposals, which are examined by an external program advisory committee LANSCE shares with the Intense Pulsed Neutron Source at Argonne National Laboratory. An internal program advisory committee also exists; it considers beam-time allocation for work of programmatic interest to the Laboratory.

Oversubscription for instrument beam time by a factor of 2—even more on the Low-Q Diffractometer and the High-Intensity Powder Diffractometer—has
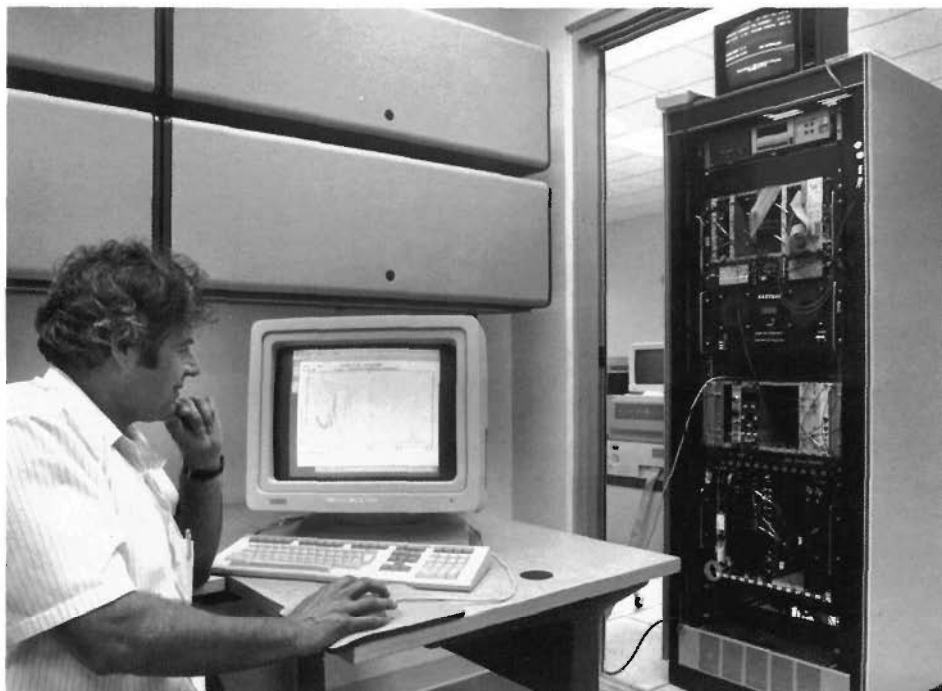
Walter Kalceff, a user from the Department of Applied Physics, University of Technology, Sydney, Australia, uses the data-acquisition system for the High-Intensity Powder Diffractometer to check on the progress of an experiment.

been the norm during the three years the LANSCE user program has been in effect. There is no charge to researchers for nonproprietary experiments, but DOE cost-recovery rules apply to experiments not publishable in the open literature.

LANSCE operates around the clock during the run cycles. Giving users directions to the facility is quite easy. One just says, "... and from the entrance gate of TA-53, just continue on down La Mesita Road for about a mile until you see a pink building on your right."

"'Pink?" questions the user.

"Yes, you won't be able to miss it—even at night."

And when they arrive, they note that the desks inside aren't made of gray metal and the kitchen has hot-pink counters. More important, they note that the new 18,000 square-foot

experiment hall and the well-equipped data rooms are part of a facility that gives them a higher peak neutron flux on their sample than any other spallation source in the world.

LANSCE has seven working neutron spectrometers; two more are under construction. Because an experiment team often includes two or three people, a visitor count between ten and twenty at any one time during the run cycles is not unusual. Most experiments take between two and ten days to complete; powder-diffraction and small-angle-scattering experiments are relatively quick, whereas inelastic-scattering experiments may take a few weeks. Each user is assigned a local contact—one of the LANSCE scientific staff—who assists in setting up the user's experiment and explains the subtleties of the data-acquisition system. Arrangements can be made for the user who wishes to stay

on at LANSCE for a few days after the experiment to complete an analysis, or at least to get the data into a form more convenient for analysis back home. In most cases, the working relationship between visitors and LANSCE scientists quickly develops into a true collaboration, one in which ideas and knowledge are freely exchanged.

LANSCE provides the target assembly, the data-acquisition and -analysis systems, the neutron-scattering instruments, and the required support services. The Medium Energy Physics Division operates the PSR that delivers proton pulses to LANSCE and the WNR facility. Support for LANSCE and PSR is given by the Office of Basic Energy Sciences of the DOE and by the Laboratory. Operation of the LAMPF accelerator itself is handled by the Medium Energy Physics Division with support from the DOE-OER Office of High Energy and Nuclear Physics.

In July 1989 the Laboratory's neutron-scattering center was officially dedicated in honor of former New Mexico Congressman Manuel Lujan, Jr. When now Secretary of Interior Lujan arrived at LANSCE for the dedication ceremonies, he noticed pictures of the early days of Los Alamos hanging on the wall. Secretary Lujan went straight to the picture of the post exchange and said he worked there making sodas during the war. It seems appropriate that a research center at Los Alamos should be named for a native son of San Ildefonso and long-term member of the House Committee on Science, Space, and Technology.

Conclusion

LANSCE has demonstrated the unique capabilities of a high-intensity pulsed neutron source and has furthered the development and refinement of neutron-scattering instrumentation. It is truly an outstanding tool for

research in many areas of condensed-matter science. In addition, LANSCE plays an important role in the technological advancements of our society and provides a unique educational opportunity for graduate students in a wide range of disciplines. ∎

Acknowledgments

Further Reading

R. N. Silver. 1986. The Los Alamos Neutron Scattering Center. *Physica* 137B: 359–372.

Dianne K. Hyer, editor. 1990. *Condensed Matter Research at LANSCE*. Los Alamos National Laboratory report LALP-90-7.

Dianne K. Hyer, editor. *from the tip of the LANSCE*. Los Alamos, New Mexico: Los Alamos National Laboratory.

Physics Today, January 1985. A special issue on neutron scattering.

C. G. Windsor. 1981. *Pulsed Neutron Scattering*. London: Taylor and Francis.

W. Gavin Williams. 1988. *Polarized Neutrons*. Oxford: Clarendon Press.

Ewald Balcar and Stephen W. Lovesey. 1989. *Theory of Magnetic Neutron and Photon Scattering*. Oxford: Clarendon Press.

Malcolm F. Collins. 1989. *Magnetic Critical Scattering*. Oxford: Oxford University Press.

D. K. Hyer, editor. 1989. *Advanced Neutron Sources 1988: Proceedings of the 10th Meeting of the International Collaboration on Advanced Neutron Sources (ICANS X) held at Los Alamos, 3–7 October 1988*. Institute of Physics Conference Series Number 97. Bristol: Institute of Physics.

Varley F. Sears. 1989. *Neutron Optics: An Introduction to the Theory of Neutron Optical Phenomena and Their Applications*. Oxford: Oxford University Press.

Dianne K. Hyer received her Bachelor's degree from Louisiana State University and Master's degree from the University of New Mexico. She holds a lifetime teaching credential from the state of California and professional teaching credentials in both elementary and secondary education from the state of New Mexico. She was the first Computer Coordinator for the Los Alamos School District and has taught in both the mathematics and the computer science departments at the University of New Mexico, Los Alamos. Dianne joined the Laboratory in 1985 as a staff member in the Physics Division's inertial confinement fusion program. She moved to LANSCE in 1987 and, as science coordinator, organized its user program. Dianne is currently co-author and Project Director of the Los Alamos SWOOPE Project—an innovative environmental science program in which scientists, teachers, and students collaborate in gathering environmental data of national concern. Dianne enjoys music, philosophy, contemplating the universe, her children, being in the middle of ideas that grow, producing musicals, and catamaran racing.

BIOLOGY
on the scale of

COO⁻

^+H_3N — C — H

CH_2

amino acid

10

Calmodulin

100

nucleosome
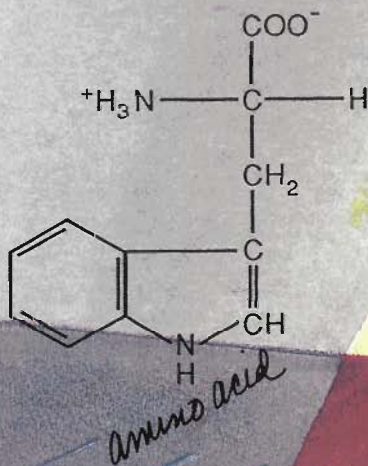
The technique of small-angle neutron scattering illuminates length scales from tens to thousands of angstroms. It has been used to locate the different proteins (red and yellow) in the ribosome, to orient the DNA (red) and histone (blue) components of chromatin, and to determine the shape of calmodulin in solution.

The ultimate challenge of the life scientist is to understand how the molecular constituents of living systems carry out their biological functions. What molecular interactions bring about and control such complicated functions as metabolism, reproduction, defense against invading organisms, and response to external stimuli?

Although none of those functions are fully understood, much of the progress so far has come from uncovering the *structures* of the biomolecules involved. In biology structure and function are intimately related. A most striking illustration is DNA, molecules of which contain all the genetic information for life. The discovery by Watson and Crick of the double-helix structure of DNA led immediately to a molecular picture of how the genetic instructions for making proteins are stored within an organism as well as inherited

NEGLECTED DIMENSIONS

by Jill Trewhella

Rhibosome

1000

Tobacco Mosaic Virus

10000

White Blood Cell
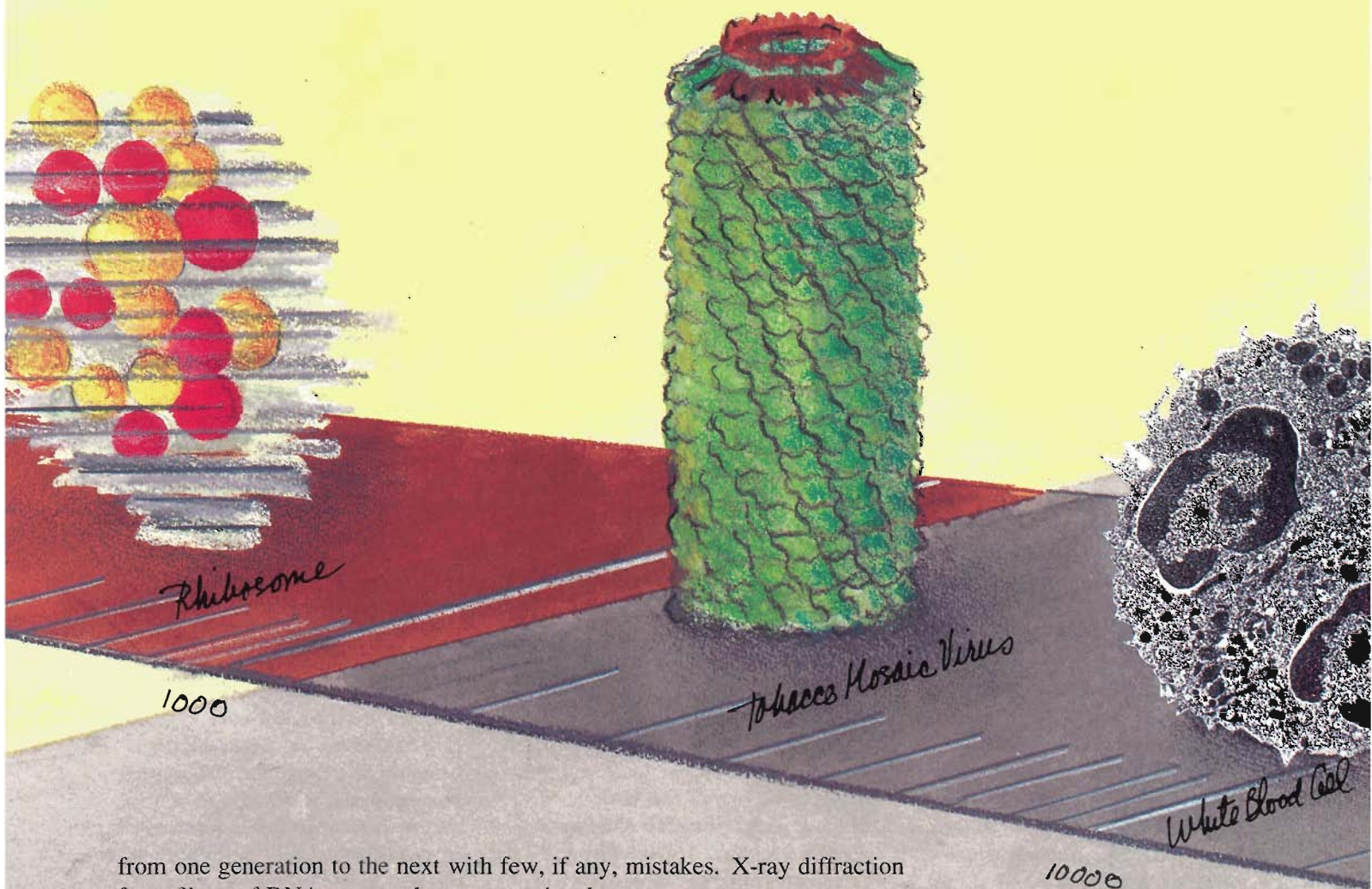
from one generation to the next with few, if any, mistakes. X-ray diffraction from fibers of DNA was used to uncover its elegant structure.

Another major breakthrough in understanding structure-function relationships at the molecular level was the elucidation, through x-ray crystallography, of the structures of various proteins. Proteins are the working agents in all life processes. They catalyze biochemical reactions, they transport molecules and ions from one cellular compartment to another, they convert light to chemical

energy, they generate and transmit nerve impulses, they control cell differentiation and growth, they provide mechanical support, they even form tiny motors that propel microorganisms through their world much as an outboard motor propels a boat.

An organism manufactures a protein by joining together in a linear polypeptide chain a certain combination of some twenty different amino acids. The number of amino acids in the chain and their sequence, or order, along the chain are specified by the organism's DNA. It is the sequence of the amino acids that determines the function of a protein— by determining the way the polypeptide chain folds up into its functional structure, or conformation. The unique conformation of each protein species creates a unique chemical environment, which, in turn, may facilitate some specialized molecular interaction. Hence, each protein is able to carry out its function, or sometimes a set of functions, faithfully.

What tools are available for determining the structures of biological molecules and the functional units they compose? The task is difficult because the molecular weights of the molecules are quite large, 1000 to 1,000,000 daltons. (Such large molecules are often referred to as macromolecules.) Nonetheless, techniques such as x-ray crystallography and nuclear-magnetic-resonance spectroscopy have given us an exquisitely detailed picture of the locations of individual chemical groups within some macromolecules and, in special cases, within macromolecular assemblies such as small viruses. Ideally, one would like to have that level of structural detail for all biological molecules in all functional states. However, after decades of effort, high-resolution structural data have been obtained for only a few hundred of the thousands of proteins whose amino-acid sequences are known, and those data often pertain to only one functional state.

At the level of the cell, light mi-

croscopy has provided us with a view of the functioning of large assemblies of biological molecules. But between the level of individual chemical groups and the cellular level is another realm, referred to by Wolfgang Oswald in the 1920s as "the world of neglected dimensions." Populating that realm, which is highlighted in the opening figure, are the biological units responsible for controlling replication and transcription of genetic information, manufacturing proteins, transmitting neural impulses, converting light to neural impulses, contracting muscle fibers, and so on. Even today that structural realm remains relatively mysterious. The functional units are so large, so complex, and so difficult to crystallize that no single technique is adequate to tackle the job of determining their structures. Progress has been made only through the interplay of many different techniques. Among them is neutron scattering.

The Contributions of Neutron Scattering

Studies ranging from high-resolution neutron scattering from crystals to low-resolution neutron scattering from particles dispersed in a solvent have made important contributions to the structural database in biology. High-resolution crystallography has provided information, for example, about the locations of hydrogen atoms, and hence of conformation-stabilizing hydrogen bonds, in biological molecules and about the structure of water molecules in the immediate vicinity of the surfaces of biological molecules. The latter information is important because physical or chemical combination of water, the fluid of life, with biological molecules influences their conformations and often facilitates their functions.

At the other extreme, the structure of the basic subunit in the chromosome, namely the nucleosome (consisting of a

NEUTRON SCATTERING LENGTHS

The neutron scattering length of an atom is a measure of the strength of the scattering interaction between neutrons and the atomic nucleus. Listed here are scattering lengths of atoms commonly found in biological molecules. (Deuterium, ^2_1H, is included for reasons given in the text.) Also listed are values for the analogous parameter in x-ray scattering. Note that the scattering lengths vary unpredictably from one atom to another and that the range of variation is not great. In contrast, the x-ray scattering factors increase monotonically with atomic number.

Atom	Neutron Scattering Length (10^{-12} cm)	X-Ray Scattering Factor (electrons)
^1_1H	−0.38	1.0
^2_1H	0.67	1.0
$^{12}_6C$	0.66	6.0
$^{14}_7N$	0.94	7.0
$^{16}_8O$	0.58	8.0
$^{31}_{15}P$	0.51	15.0

length of DNA wound around a protein core), was first determined by scattering neutrons from nucleosomes in solution. Low-resolution neutron scattering is also being used to locate the many different protein species within the ribosome, the assembly responsible for translating RNA into proteins. That difficult work has been going on for fifteen years and is not yet completed.

Neutron scattering has also contributed structural information that will help explain how the transmembrane protein bacteriorhodopsin serves as a light-driven pump of protons, or hydrogen

ions. Bacteriorhodopsin is found in certain primitive bacteria and is responsible for their ability to use light to generate metabolic energy when oxygen is scarce. Absorption of light by bacteriorhodopsin creates an electrochemical gradient across the bacterial membrane, which serves as the source of metabolic energy. Neutron diffraction revealed in part how the protein chain folds back and forth across the membrane, presumably bringing together the chemical groups responsible for the proton translocations that create the gradient, but more work is needed to unravel the exact mechanism.

Our studies of the protein calmodulin, presented below, are an example of the use of neutron scattering to study, under physiological conditions, the conformational changes that occur as proteins carry out their functions.

In general, neutron scattering is useful for studying biological units ranging in size from small proteins to animal viruses. Moreover, it is a particularly powerful tool for studying composite entities containing more than one molecular species. That special power arises from a large difference between the scattering of neutrons by hydrogen (a major constituent of all biological molecules) and its neighboring isotope deuterium.

The Neutron as a Probe of Biomolecules

In "Neutron Scattering—A Primer," hereafter referred to as the primer, we learn that unlike x rays, which scatter from the electron cloud surrounding an atomic nucleus, neutrons scatter from the atomic nucleus itself. If the wavelength of the neutron is much larger than the size of the nucleus (as it is in all low-energy neutron-scattering experiments), then the nucleus is considered to be located at a point $\mathbf{r}'$, and the interaction between the neutron and the

---

Table 2

CONTRAST MATCHING

When low-energy (long-wavelength) neutrons scatter at small angles from biological molecules in solution, the parameter that governs the observed intensity of scattered neutrons is the "contrast" between the biological molecules and the solvent, that is, the difference between their scattering-length densities. Therefore, by matching the scattering-length density of the solvent to the scattering-length density of one component of a two-component biological macromolecule, the contribution of that component to the observed neutron scattering can be eliminated. Listed here are the mean scattering-length densities of a nucleotide (DNA), a protein (histone), and lipids and the mean scattering-length densities of water (the usual solvent) containing various percentages of D_2O.

	Mean Scattering-Length Density $(10^{10} \text{ cm}^{-2})$
DNA	2.01
Histone	3.81
Lipids	between −0.2 and 0.8
0% D_2O	−0.56
100% D_2O	6.38
x% D_2O	$0.01[6.38x - 0.56(100 - x)]$
37% D_2O	2.01
63% D_2O	3.81

---

nucleus can therefore be described by a delta-function pseudopotential, $b\delta(\mathbf{r}-\mathbf{r}')$. The parameter b, the so-called scattering length, is related to the total scattering cross section of the nucleus, $4\pi b^2$. It turns out that the scattering lengths of all elements commonly found in biological macromolecules, except hydrogen, are positive and similar in magnitude. In contrast, the scattering length

of hydrogen is *negative* (because of a nuclear resonance) and equal in magnitude to about half the values of the scattering lengths of the other elements in biomolecules (Table 1). The negative scattering length of hydrogen means that the neutrons scattered from hydrogen are 180° out of phase with the neutrons scattered from elements with positive scattering lengths. Both the sign and the relatively large magnitude of hydrogen's scattering length make that element far easier to "see" with neutrons than with x rays.

In this article we will focus on small-angle neutron scattering from macromolecules in solution. That technique is useful for determining overall shapes and sizes, molecular weights, and internal composition variations of the large molecules, rather than the locations of individual atoms. Consequently, a scattering-length density (the sum of the scattering lengths of the nuclei within a volume element dV divided by the volume element) is relevant rather than individual scattering lengths. We will denote the scattering-length density at a particular point $\mathbf{r}$ within a particle by $b_p(\mathbf{r})$ and the mean scattering-length density of the particle by $\bar{b}_p$. Because various types of biological molecules, such as proteins, polynucleotides (DNA and RNA), and lipids (the basic constituents of membranes), contain different proportions of hydrogen relative to other elements and because the mean scattering-length density varies rapidly with the proportion of hydrogen, the mean scattering-length density of those types of molecules differ significantly. Serendipitously, their mean scattering-length densities lie between those for pure H_2O and pure D_2O (Table 2). That fact is crucial for the powerful technique called contrast matching, a technique that allows us to see, for example, only the protein component in a biological assembly containing proteins and DNA or proteins and lipids.

Contrast Matching

At the right is a whimsical example of contrast matching in which one figure blends into the background pattern and the other figure stands out. Another example of contrast matching, more analogous to the situation in neutron scattering, is the viewing of black and white kittens playing on black or white carpets. On a white carpet the black kittens stand out sharply, whereas the white kittens blend into the carpet. In other words, the white kittens are "contrast matched" to the background. The reverse effect is achieved by placing the kittens on a black carpet. The black kittens are then matched to the background, and only the white kittens are easily visible. In neutron scattering from macromolecules in solution, the solvent plays the part of the carpet, the molecules play the part of the kittens, and the contrast is determined by the *difference* between the mean scattering-length densities of the solvent and the molecules.

If the molecules and the solvent have the same scattering-length density, the waves scattered from any point have the same amplitude and phase. When integrated over the entire sample volume, the scattering at any angle (other than zero) cancels because arriving waves have all phases with equal probability. In other words, no small-angle scattering occurs. On the other hand, suppose the dispersed molecules and the solvent have different scattering-length densities, b_p and b_s. Each wave scattered from some point within a molecule may be thought of as a sum of two waves, one of amplitude b_s and one of amplitude $b_p - b_s$. The waves of amplitude b_s cancel with the waves of equal amplitude scattered from the solvent, whereas the waves of amplitude $b_p - b_s$ interfere constructively or destructively with other waves scattered from other points within the molecule. As described below, con-

©1988 Universal Press Syndicate

When the monster came, Lola, like the peppered moth and the arctic hare, remained motionless and undetected. Harold, of course, was immediately devoured.

structive interference of those waves occurs mostly at small angles about the direction of the incident neutrons. Thus the name small-angle scattering is applied to the phenomenon. The observed intensity of the scattered neutrons is equal to the square of the amplitude of the scattered waves and is thus proportional to $(b_p - b_s)^2$.

Now consider a particle composed of two subunits, each with a different mean scattering-length density. By varying the D_2O-to-H_2O ratio in the solvent, the scattering-length density of the solvent can be matched to that of one of the subunits. Then most of the observed scattering is due to the other subunit. Contrast matching was the technique used in the 1970s to elucidate the structure of the nucleosome, which contains

DNA and proteins. At that time it was not clear whether the protein component of the nucleosome was outside or inside the DNA component. The DNA component has the same scattering-length density as a solvent containing approximately 63 percent D_2O, whereas the protein component has the same scattering-length density as a solvent containing approximately 37 percent D_2O. Neutron-scattering data obtained at each of those match points, as well as at other solvent scattering-length densities, gave information on the relative orientation of the two components and led to the model shown in Fig. 1.

Many biochemical units contain protein species whose mean scattering-length densities are so similar that the technique of contrast matching is not directly applicable. The ribosome in the opening figure is an example. However, the range of variation among protein species can be increased by replacing some of them with "deuterated" versions. A deuterated version of a protein, as you might guess, contains deuterium rather than hydrogen. It can be prepared in various ways. For example, if the protein of interest is a bacterial protein, the bacterium is simply cultured on a medium containing D_2O rather than H_2O. As the bacterium multiplies, it produces deuterated versions of various proteins, including the protein of interest. The locations of the different protein species in a bacterial ribosome were determined by small-angle neutron-scattering studies on a set of ribosomes, each of which had been reconstituted with deuterated versions of two protein species prepared in the above manner. (Deuterated versions of nonbacterial proteins can be obtained by using recombinant DNA technology. Clones of the gene for the protein of interest are first combined with the DNA of a bacterium. The bacterium is then cultured as above, and the protein of interest harvested.)
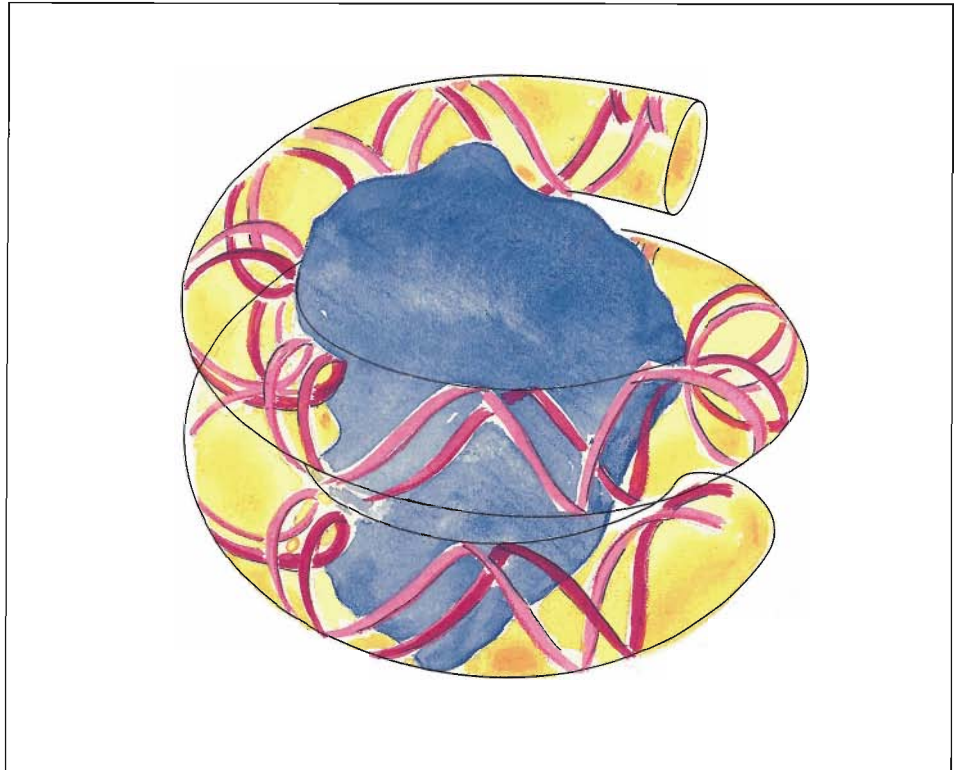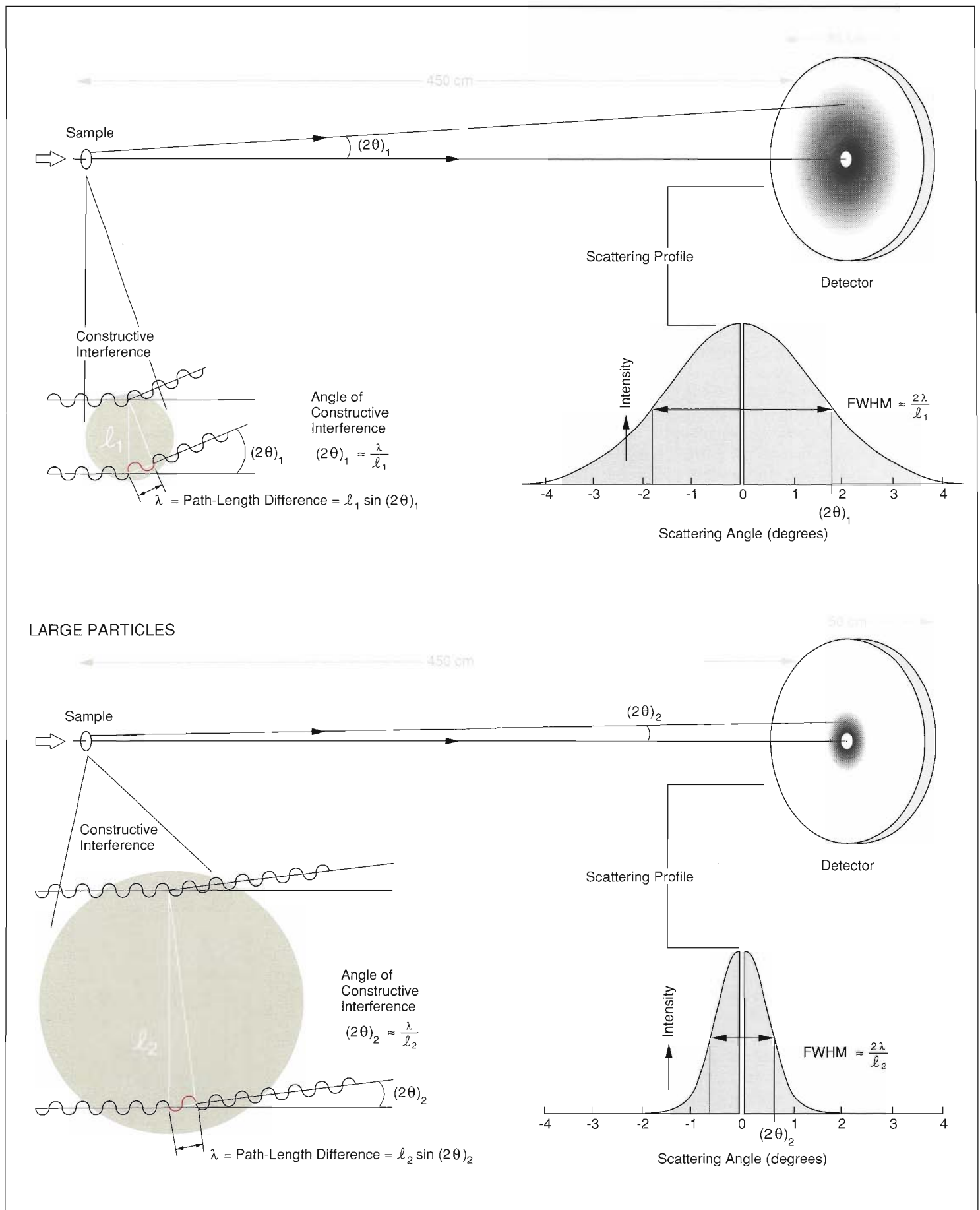

THE NUCLEOSOME: A TWO-COMPONENT SYSTEM

Fig. 1. The nucleosome, the basic subunit of chromosomes, contains DNA and proteins known as histones. That the DNA was wrapped around the proteins, rather than vice versa, was first deduced by using contrast matching to measure separately the small-angle neutron-scattering profiles of the DNA and the protein components.

Small-Angle Neutron Scattering

Figure 2 shows the experimental setup for small-angle neutron-scattering experiments. A sample of particles dispersed in a small volume (of the order of 100 cubic millimeters) of solution is placed in the neutron beam, and a detector located several meters away measures the intensity of the neutrons scattered through various angles 2θ relative to the direction of the incident beam. To protect the detector, a small beam stop (a few square millimeters) absorbs the unscattered neutrons—as well as those scattered through zero degrees and through angles very close to zero. The intensity of the scattered neutrons (represented by shades of gray in Fig. 2) varies with the scattering angle and, if the dispersed particles are randomly oriented, is isotropic about the beam axis. Figure 2 shows a peak in intensity centered at zero degrees and extending over a small range of scattering angles. (Note that the beam stop prevents direct observation of the intensity at scattering angles equal and very nearly equal to zero. Instead $I(0)$ is determined by extrapolation of the observed scattering profile.) Such a peak occurs if the wavelength of the incident neutrons, λ, is small compared with the particle size and large compared with internuclear distances. The peak is produced by constructive interference of waves scat-

Sample

$(2\theta)_1$

Scattering Profile

Detector

Constructive
Interference

Intensity

ℓ_1

Angle of
Constructive
Interference

$(2\theta)_1$

$(2\theta)_1 \approx \dfrac{\lambda}{\ell_1}$

FWHM $\approx \dfrac{2\lambda}{\ell_1}$

λ = Path-Length Difference = $\ell_1 \sin (2\theta)_1$

$(2\theta)_1$

Scattering Angle (degrees)

LARGE PARTICLES

Sample

$(2\theta)_2$

Detector

Scattering Profile

Constructive
Interference

Intensity

ℓ_2

Angle of
Constructive
Interference

$(2\theta)_2 \approx \dfrac{\lambda}{\ell_2}$

$(2\theta)_2$

FWHM $\approx \dfrac{2\lambda}{\ell_2}$

λ = Path-Length Difference = $\ell_2 \sin (2\theta)_2$

$(2\theta)_2$

Scattering Angle (degrees)

SMALL-ANGLE NEUTRON SCATTERING

Fig. 2. In a small-angle neutron-scattering experiment a detector located relatively far from the sample records the intensity *I* of the scattered neutrons as a function of scattering angle 2θ. *I*(0) cannot be measured directly because the neutrons scattered through zero degrees are coincident with the unscattered neutrons. Both are absorbed by a small beam stop coaxial with the incident beam. The observed scattering profile (graph of *I* versus 2θ) at very small scattering angles contains information about the size and shape of the particles in the sample. For example, the width of the scattering profile varies inversely with the size of the scattering particles. In the case of spherical particles, the maximum separation between scattering points for which interference may occur is the particle diameter (ℓ). Integration over the particle volume yields a scattering profile that is approximately Gaussian (at small angles) with a full width at half maximum (FWHM) of (1.14) (2λ/ℓ). The figure shows the condition for constructive interference between the scattered waves from a pair of nuclei separated by a distance ℓ: the difference between their path lengths, ℓ sin 2θ, must equal λ, the wavelength of the incident neutrons. For small angles sin 2θ ≈ 2θ, so the condition for constructive interference becomes 2θ ≈ λ/ℓ.

tered from pairs of nuclei within each particle, which occurs when the difference in path length of the two scattered waves is equal and nearly equal to an integral multiple of λ. Now the waves scattered at most angles from the many, many pairs of nuclei within the particle will tend to cancel out and produce a very small signal. Only the waves scattered at small angles (for which the path-length differences are also small)

will tend to add up and produce a large peak. The shape of the peak contains information about the size of the particles. For example, as shown in Fig. 2, the width of the peak (at half maximum) varies approximately inversely with the diameter of the particles.

To explain how we extract more exact quantitative information from the small-angle scattering profile, we build on the discussion in the primer. Consider the scattering in vacuum of neutrons from a particle whose scattering-length density is described by $b_p(\mathbf{r})$. Equation 4 in the primer states that the scattering profile of the neutrons is given by

$$I(Q) = \left| \int_{V_p} b_p(\mathbf{r}) e^{i\mathbf{Q}\cdot\mathbf{r}} d^3\mathbf{r} \right|^2, \quad (1)$$

where $\mathbf{Q} \equiv \mathbf{k}_{\text{initial}} - \mathbf{k}_{\text{final}}$ is the scattering vector, $Q \equiv |\mathbf{Q}| = 2\pi \sin\theta/\lambda$, $d^3\mathbf{r}$ is a volume element, and V_p is the volume of the particle. Suppose now that the neutrons are scattered from a particle in a matrix of uniform scattering-length density b_s. Then Eq. 1 becomes

$$I(Q) = \left| \int_{V_p} (b_p(\mathbf{r}) - b_s) e^{i\mathbf{Q}\cdot\mathbf{r}} d^3\mathbf{r} \right|^2. \quad (2)$$

(The quantity $b_p(\mathbf{r}) - b_s$ in Eq. 2 is called the contrast factor.) If, as is generally the case, the neutrons are being scattered by more than one particle, the integral in Eq. 2 must be multiplied by the number of particles N_p (provided the particles are identical). For simplicity, that multiplicative constant is not explicitly given in the equations that follow.

By using the fact that, for any complex number z, $|z|^2 = zz^*$ (where z^* is the complex conjugate of z), Eq. 2 becomes

$$I(Q) = \int_{V_p} \int_{V_p} (b_p(\mathbf{r}) - b_s)(b_p(\mathbf{r}') - b_s) \\ \times e^{i\mathbf{Q}\cdot(\mathbf{r}-\mathbf{r}')} d^3\mathbf{r}\, d^3\mathbf{r}'. \quad (3)$$

The scattering of neutrons from biological molecules in solution is generally spherically averaged because the molecules are generally randomly oriented. The spherical average of Eq. 3 can be expressed in terms of a so-called pair-distribution function $P(R)$, where $R \equiv |\mathbf{r} - \mathbf{r}'|$:

$$I(Q) = 4\pi \int_0^{d_{\max}} P(R) \frac{\sin QR}{QR} dR, \quad (4)$$

where $d_{\max}$ is the maximum dimension of the particles. (The transition from Eq. 3 to Eq. 4 is more easily comprehended by recalling that the spherical average of e^{ix} is $\sin x/x$.) The pair-distribution function $P(R)$ in Eq. 4 is a weighted frequency distribution of the distances $|\mathbf{r} - \mathbf{r}'|$ between all pairs of scattering centers (that is, nuclei); each frequency is weighted by the factor $(b_p(\mathbf{r}) - b_s)(b_p(\mathbf{r}') - b_s)$.

Implicit in the derivation of Eq. 4 are several important assumptions: (1) all the particles in the solution are identical; (2) the particles are randomly oriented; (3) the concentration of the particles is sufficiently low that neutrons scattered by different particles do not interfere with each other; and (4) thermodynamically, the solution is a two-component system consisting of particles dissolved in a homogeneous solvent.

Structural information about the particles is deduced from an analysis of $P(R)$, which is determined by measuring $I(Q)$ and calculating the inverse Fourier transform of Eq. 4:

$$P(R) = \frac{1}{2\pi^2} \int_0^\infty I(Q)QR \sin(QR)dQ. \quad (5)$$

(Although Eq. 5 indicates that the integration should extend from zero to infinity, $I(Q)$ can be measured only over a finite Q range. That experimental limitation can introduce truncation errors into the calculated $P(R)$. Also note that

$I(Q)$ is measurable only if N_p is relatively large.) One piece of information obtained from $P(R)$ is the overall shape of the scattering particle. For a spherical particle of uniform scattering-length density, $P(R)$ is symmetric (Fig. 3a). It has a maximum near a value of R equal to the radius of the sphere and falls to zero at $R = 0$ and at an R value equal to the sphere's diameter. As the particle becomes more asymmetric, so does $P(R)$. Figure 3b shows pair-distribution functions for several two-lobed particles with uniform scattering-length density. Note that those functions are quite sensitive to the relative dispositions of the two lobes. In all cases $P(R) = 0$ for all R values greater than the maximum distance between scattering centers within the particle, and hence one can estimate the maximum dimension of the molecule, d_{max}.

Although $P(R)$ is a complicated function of R, its integral over R is simply $\left[(\bar{b}_p - b_s)V_p\right]^2$, the square of the mean contrast factor times the particle volume. Moreover, according to Eq. 3 the intensity of the neutrons scattered through zero degrees, $I(0)$, is also equal to $\left[(\bar{b}_p - b_s)V_p\right]^2$:

$$I(0) = \int_{V_p}\int_{V_p}(b_p(\mathbf{r}) - b_s)(b_p(\mathbf{r}') - b_s)d^3\mathbf{r}\,d^3\mathbf{r}'$$
$$= \left[(\bar{b}_p - b_s)V_p\right]^2$$
$$= \int_0^{R_{max}} P(R)dR.$$

(6)

$I(0)$ is one of the most commonly cited small-angle neutron-scattering parameters. Since the neutrons scattered through zero degrees are coincident with the unscattered beam, $I(0)$ cannot be measured directly. Instead, $I(0)$ is determined by extrapolating the measured scattering profile to $2\theta = 0$ (see below). By determining $I(0)$ at different values of b_s, we can determine $\bar{b}_p$ and V_p. In particular, according to Eq. 6 a plot of

(a) Single-Lobed Objects
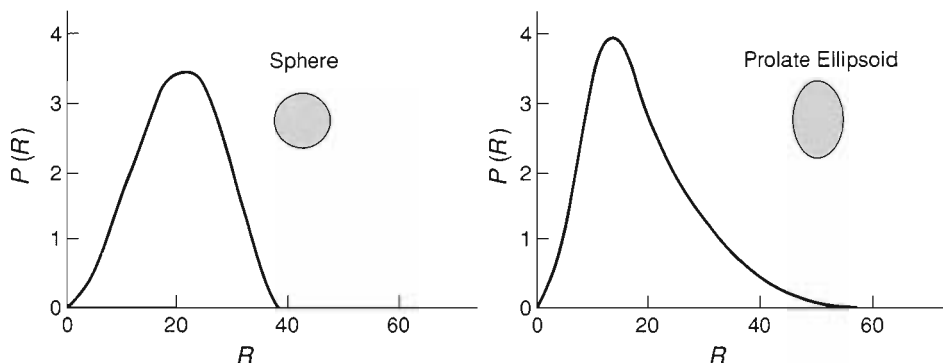
(b) Two-Lobed Objects
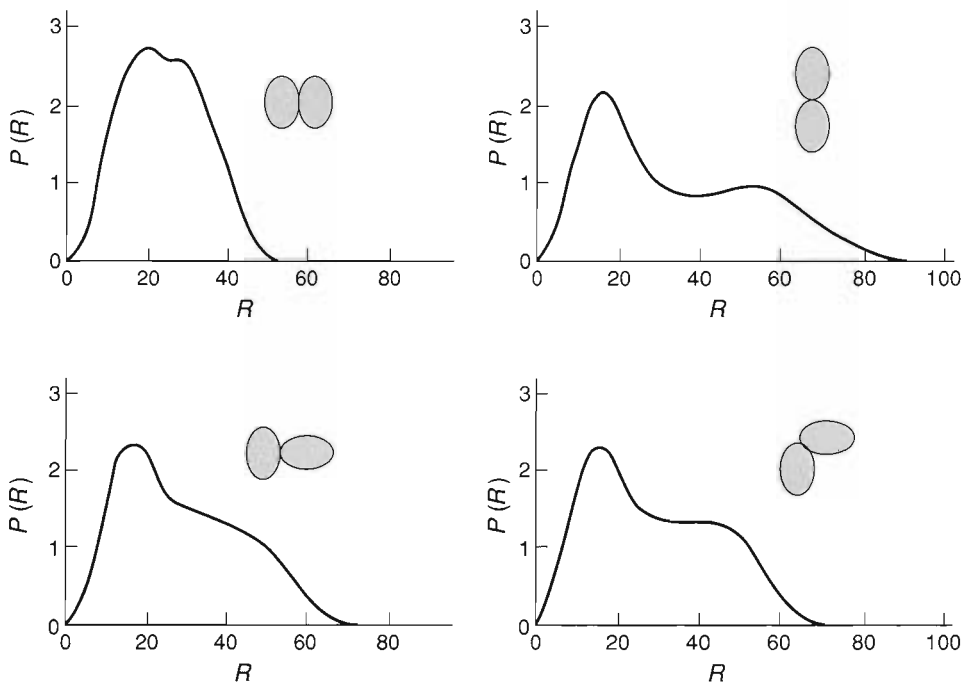
PAIR-DISTRIBUTION FUNCTIONS FOR OBJECTS OF DIFFERENT SHAPE

Fig. 3. General structural features of an object are reflected in its pair-distribution function $P(R)$, which is determined from its small-angle neutron-scattering profile. If the object has a uniform scattering-length density, $P(R)$ is proportional to the frequency distribution of the distances R between pairs of scattering centers (nuclei). Shown here are pair-distribution functions for objects of uniform scattering-length density and various shapes. (a) The pair-distribution functions for a sphere and a prolate ellipsoid show that the asymmetry of $P(R)$ increases with the asymmetry of the object's shape. (b) The pair-distribution functions for various two-lobed objects demonstrate the sensitivity of $P(R)$ to the relative orientation of the two lobes.

$\sqrt{I(0)}$ versus b_s is a straight line with a slope of $-V_p$ and an intercept with the b_s axis of $\bar{b}_p$. Also, because the molecular weights of proteins are approximately proportional to their volumes, $I(0)$ serves as a check that proteins in a sample are dispersed as single particles rather than as aggregates.

Another commonly cited parameter is R_g, the radius of gyration, which is defined by the equation

$$R_g^2 = \frac{\int_{V_p}(b_p(\mathbf{r}) - b_s)|\mathbf{r}|^2 d^3\mathbf{r}}{\int_{V_p}(b_p(\mathbf{r}) - b_s)d^3\mathbf{r}}. \qquad (7)$$

The radius of gyration of neutron scattering is analogous to the radius of gyration of classical mechanics, which is defined by an equation exactly like Eq. 7 except that a mass density replaces the contrast factor. (The numerator and denominator in Eq. 7 are analogous, respectively, to the moment of inertia and the mass of classical mechanics.) The radius of gyration can also be determined as the second moment of the pair-distribution function:

$$R_g^2 = \frac{\int_0^{R_{\max}} P(R)R^2 dR}{\int_0^{R_{\max}} P(R)dR}. \qquad (8)$$

Using $P(R)$ to calculate R_g makes use of scattering data over the entire measurable Q range.

An alternative method for calculating R_g was developed by Guinier. He showed that the innermost portion of the scattering profile can be approximated well by a Gaussian curve:

$$I(Q) = I(0)e^{-\frac{1}{3}R_g^2 Q^2}. \qquad (9)$$

The value of $R_g Q$ below which the Guinier approximation is valid depends on the particle shape; typically, however, it is valid for $R_g Q < 1$. Equation 9 implies that a plot of $\ln I(Q)$ versus Q^2 is a straight line with a slope of $-\frac{1}{3}R_g^2$ and an extrapolated intercept

with the $\ln I(Q)$ axis of $\ln I(0)$. Guinier analysis is a common method for determining $I(0)$ as well as R_g. Later, when we present neutron-scattering data for a two-protein system at various contrast factors, we will see how an analysis of the variation of R_g with contrast factor allows us to deduce the relative positions of the two components, provided they have different mean scattering-length densities.

The theory for small-angle x-ray scattering is essentially the same as the theory presented here for small-angle neutron scattering except that x-ray scattering factors are substituted for neutron scattering lengths. X-ray scattering factors increase monotonically with the number of electrons in the electron cloud surrounding the atomic nucleus (that is, with the atomic number of the atom), and, because the dimensions of the electron cloud are of the same order as the wavelength of the x rays, they decrease with increasing scattering angle. Isotopes of the same element scatter x rays identically (because the number of electrons remains the same), and there are no negative x-ray scattering factors. Hence, although the scattering factor of a solvent can be changed by, for example, adding glucose or sucrose, the mean scattering factor of one component of a two-component biological assembly cannot be changed by isotopic substitution. However, x rays are easier and cheaper to produce than neutrons, and, when feasible, x-ray scattering is used to complement neutron scattering.

One final point. We mentioned above that neutron scattering from particles randomly oriented in solution yields only spherically averaged information about particle structure. One way to increase the information content of the scattering profile is to orient the particles along some direction in space. The sidebar "Ferrofluids—A New Alignment Technique" discusses a new method for orienting biological assemblies in so-

lution, which promises to yield more detailed structural information.

All the neutron-scattering experiments to be described below were performed at the Laboratory's pulsed neutron source (see "LANSCE—A Facility for Users") with the Low-Q Diffractometer (LQD), an instrument optimized for biological applications. Because the detector must be placed far from the sample to measure the neutron-scattering profile with high angular resolution and because biological molecules are inherently weak neutron scatterers, the flux of neutrons incident on the sample must be as large as possible to achieve a measurable intensity of scattered neutrons at the detector. The LQD is designed to maximize the incident flux and yet permit intensity measurements at very small scattering angles. The smallest angle at which useful intensity data can be obtained determines in part the largest particles that can be studied with the instrument.

The Q range measured with the LQD (0.002–0.5 angstrom^{-1}) covers small-angle scattering from objects with dimensions from tens to thousands of angstroms. The wide Q range also allows use of the LQD for low-resolution diffraction measurements. (Diffraction from plant viruses is discussed in "Ferrofluids—A New Alignment Technique.") Moreover, unlike the case at steady-state neutron sources, all of those measurements can be made without reconfiguring the instrument.

## Calmodulin and Biochemical Regulation

In the remainder of this article, we will focus on structural studies of the protein calmodulin, which mediates the regulatory effect of calcium ions on many biochemical processes. Through a combination of small-angle neutron and x-ray scattering, we have begun to study the conformational changes that

CALCIUM REGULATION MEDIATED BY CALMODULIN

Fig. 4. Calcium ions, Ca^{2+}, regulate many intracellular processes, and many of the regulatory functions of Ca^{2+} are mediated by the protein calmodulin. That is, a local increase in Ca^{2+} concentration causes Ca^{2+} to bind to calmodulin, and the Ca^{2+}-calmodulin complex binds to a target enzyme and thereby activates a catalytic site on the enzyme. The activated target enzyme then catalyzes, or speeds up, a biochemical reaction, such as the addition of phosphate groups to some protein. Regulation by Ca^{2+} through calmodulin is involved in such processes as muscle contraction, transmission of nerve impulses, the degradation of glycogen to glucose, and cell division.

1. Calcium-ion concentration increases

Calcium Ions

Calmodulin

2. Calcium ions bind to calmodulin, forming Ca^{2+}-calmodulin complex

Ca^{2+}-Calmodulin

3. Ca^{2+}-calmodulin binds to and activates target enzyme

Inactive Target Enzyme

Ca^{2+}-Calmodulin

Activated Target Enzyme

4. Activated target enzyme catalyzes biochemical reaction, such as phosphorylation

CATALYSIS

Phosphate Group

Protein

Phosphorylated Protein

calmodulin undergoes as it carries out its function as mediator. But before we get ahead of the story, let's focus on the process of regulation.

Living systems must regulate a myriad of interdependent biochemical processes. In other words, processes must be turned on and off or slowed down and speeded up as needed. Regulation generally involves the interaction of small messenger molecules or ions with proteins. The messengers fall naturally into three major classes defined by the time scale of the processes they control: hormones regulate intercellular processes that occur on time scales of days to hours; small organic molecules known as cyclic nucleomonophosphates (cyclic AMP, for example) regulate intracellular processes that occur on time scales of minutes to seconds; and the divalent (doubly charged) calcium ion, Ca^{2+}, regulates a large number of diverse intracellular processes that occur on time scales of seconds to milliseconds. Among the Ca^{2+}-regulated processes are mitosis (cell division), the degradation of glycogen (the storage
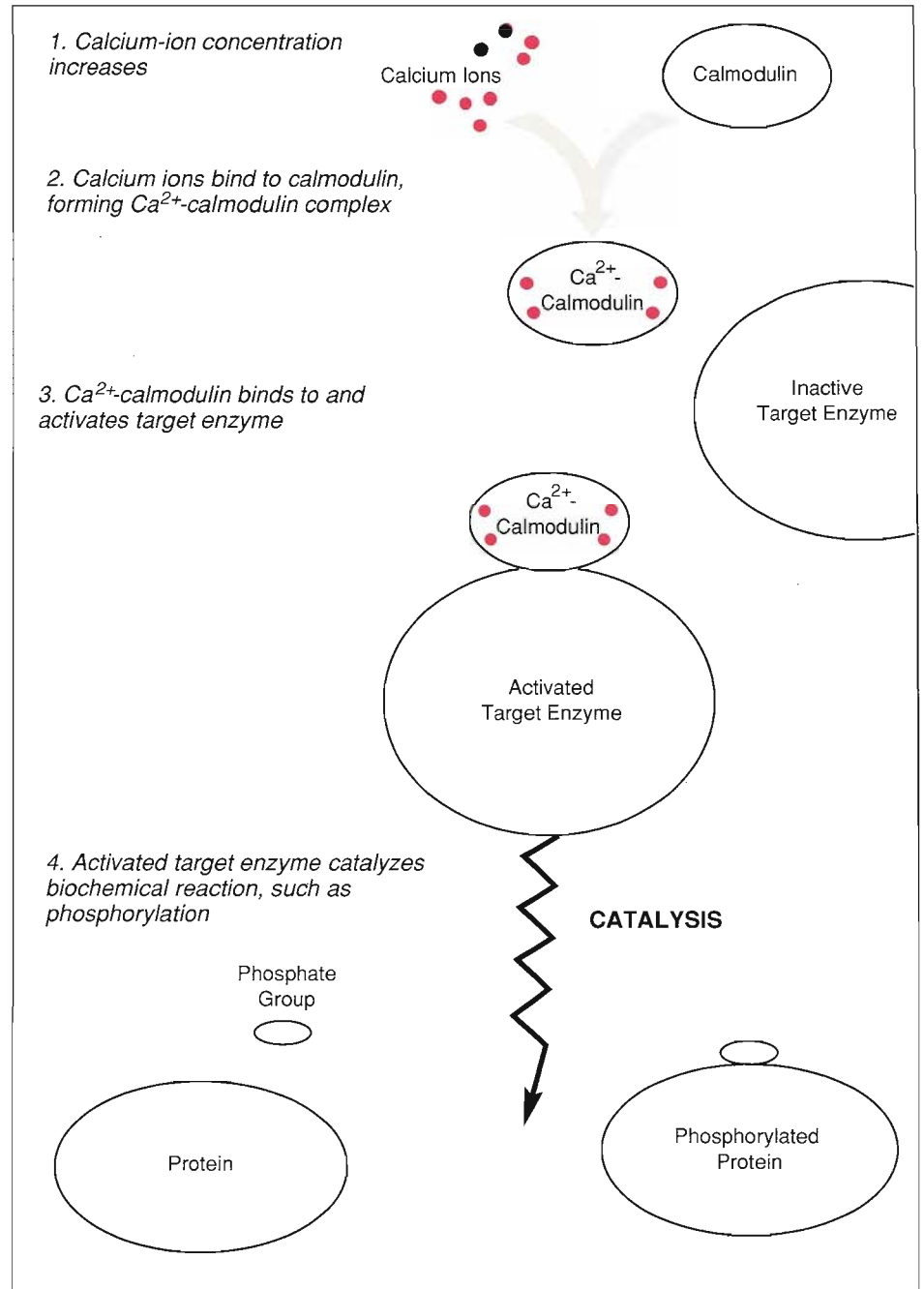
form of glucose) to glucose, muscle contraction, and transmission of nerve impulses. All those processes are regulated by localized changes in the concentration of calcium ions.

The question of how the simple calcium ion can regulate so many types of processes has attracted an enormous amount of attention. Many of the regulatory effects of Ca^{2+} are mediated by the protein calmodulin. Binding of Ca^{2+} to calmodulin is thought to induce some kind of conformational change that enables Ca^{2+}-calmodulin to bind to and thereby activate a "target" en-
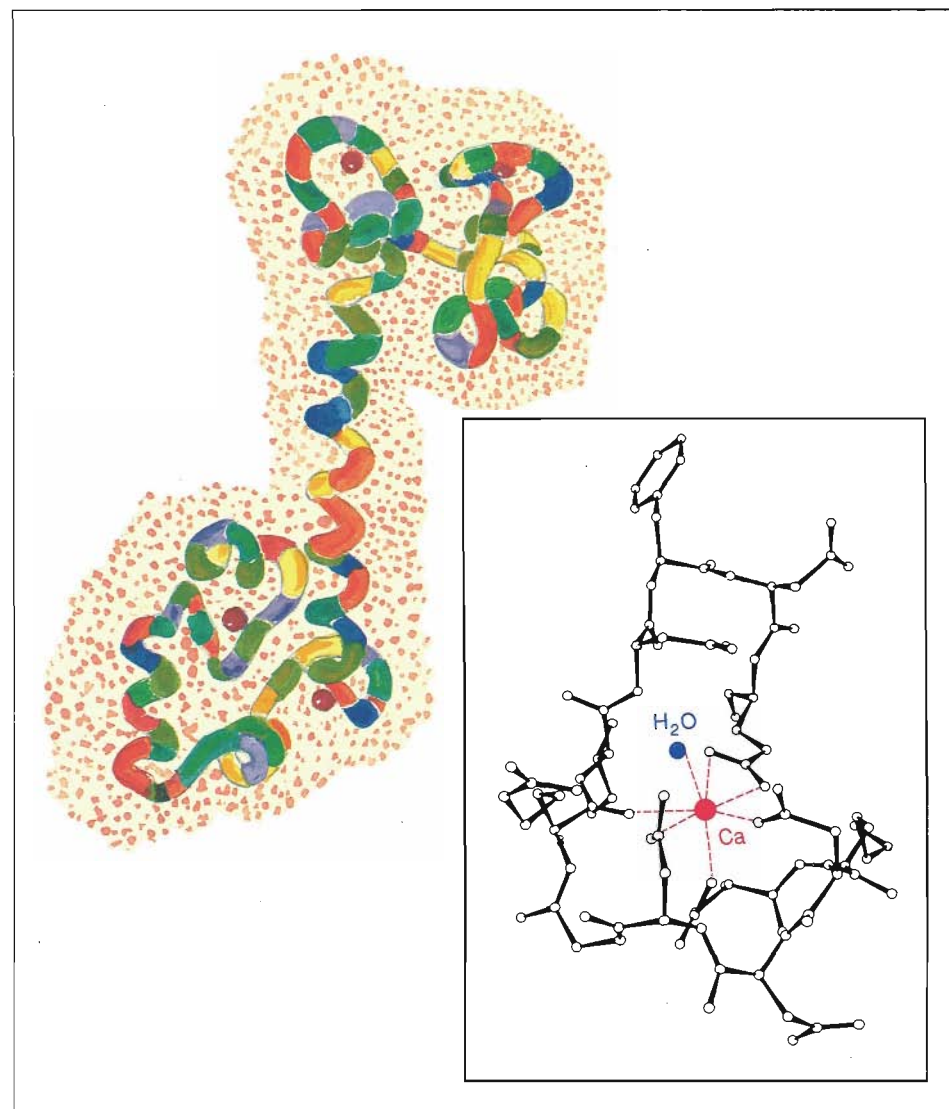
zyme. The activated target enzyme then catalyzes a biochemical reaction. The activated catalytic site on the target enzyme is sometimes far from the Ca^{2+}-calmodulin-binding site, and the mechanism of such "activation at a distance" is unknown (Fig. 4). We do know that Ca^{2+}-calmodulin can activate many different target enzymes (one at a time). Moreover, the constancy of the amino-acid sequence of calmodulin throughout evolution suggests that sequence changes almost anywhere in calmodulin are fatal to the organism and must therefore destroy key functions. Thus, the

CRYSTAL STRUCTURE OF Ca²⁺-CALMODULIN

Fig. 5. The structure of crystalline Ca²⁺-calmodulin was deduced from x-ray diffraction data for calmodulin crystallized at low pH and in the presence of Ca²⁺. The structure is depicted here in a model of calmodulin's backbone superimposed on a space-filling model (left). The structure consists of two globular lobes connected by an extended α helix of about eight turns. Located on each globular lobe are two Ca²⁺-binding sites (red). The basic structure of all four Ca²⁺-binding sites is the same (right).

sites for binding to different target enzymes may be distributed over the entire calmodulin molecule, and the activation mechanism may vary from one target enzyme to another.

What features of calmodulin's structure allow it to support such diverse activities? Calmodulin is a water-soluble protein, and most such proteins fold into globular structures that minimize the number of hydrophobic amino-acid residues exposed to water. But x-ray-diffraction data for Ca^{2+}-calmodulin (calmodulin without Ca^{2+} has not yet been crystallized) show that the polypeptide chain folds into two globular lobes connected by an extended α helix of about eight turns (Fig. 5). Two Ca^{2+}-binding sites are found in each globular lobe, and all four Ca^{2+}-binding sites have a high degree of structural similarity with each other and with the Ca^{2+}-binding domains of other Ca^{2+}-activated regulatory proteins, such as parvalbumin and troponin C.

The α helix connecting the two globular domains of calmodulin is presumed to be important in its interactions with target enzymes. Consequently, it has been the subject of considerable interest. Helical structures in more compact proteins are stabilized by intramolecular or sometimes intermolecular interactions. However, the crystal packing of Ca^{2+}-calmodulin shows no evidence of any such stabilizing interactions. So what happens to the α helix under physiological conditions? Is it stable in the extended conformation of the crystal form, or is it flexible and able to twist or bend?

The two-lobed structure of Ca^{2+}-calmodulin is an ideal subject for study by both x-ray and neutron scattering because the scattering data are quite sensitive to the relationship between the two globular domains. Moreover, when coupled with the techniques of selective deuteration and contrast matching, neutron scattering allows us to see the

shape of Ca^{2+}-calmodulin when it is bound to a target enzyme.

X-Ray Studies—
The Solution Structure

My colleagues and I have gathered small-angle x ray-scattering data that reveal differences between the crystal and solution structures of Ca^{2+}-calmodulin and between the solution structures of calmodulin and Ca^{2+}-calmodulin.

First we gathered small-angle x-ray-scattering data for calmodulin dispersed in a solution containing enough calcium
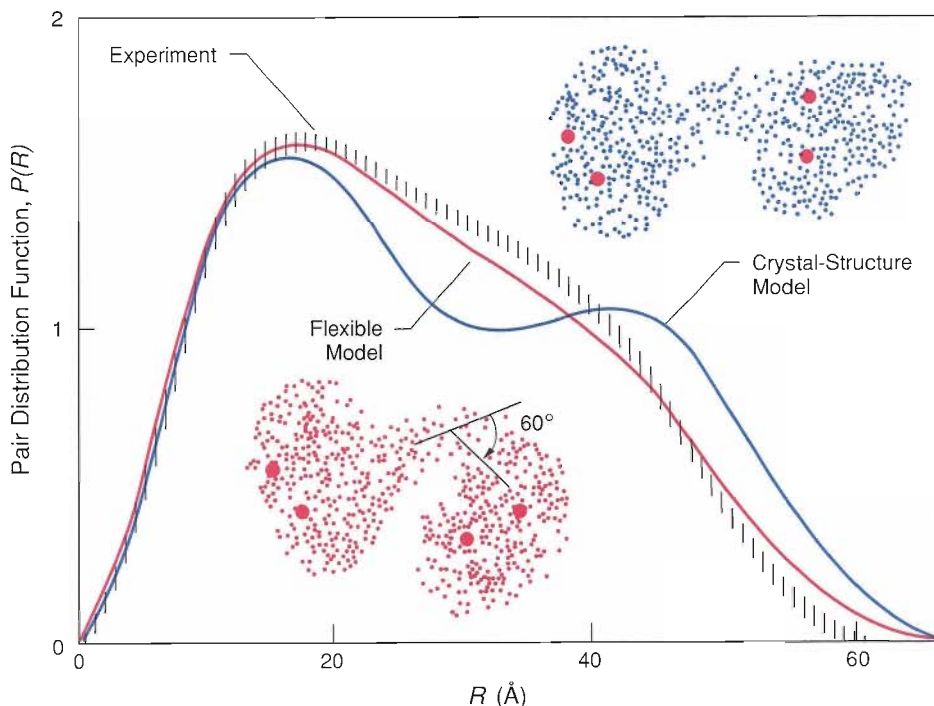
ions to saturate its Ca^{2+}-binding sites. Figure 6 compares the pair-distribution function derived from those data with two theoretical pair-distribution functions. Both theoretical pair-distribution functions were based on the same pattern of scattering centers, namely the pattern deduced from the reported crystal structure of Ca^{2+}-calmodulin. However, in one case the portion of the pattern representing the α helix was allowed to be flexible so that on average the two globular lobes are closer together than in the crystal form. It should be pointed out that a uniform

density of scattering centers within calmodulin would yield similar theoretical pair-distribution functions. Thus calmodulin is fairly well approximated as a molecule of uniform scattering-length density.

As expected, all the pair-distribution functions in Fig. 6 bear some resemblance to the pair-distribution functions for double-lobed, uniform-density structures shown in Fig. 3b. Note that the experimental $P(R)$ falls to zero at a smaller R value than does the $P(R)$ based on the (inflexible) crystal structure. Therefore the solution structure of Ca^{2+}-calmodulin has a smaller maximum dimension than does the crystal structure. More important, the experimental $P(R)$ lies below the $P(R)$ for crystalline Ca^{2+}-calmodulin at large R values and has no saddle at intermediate R values. In other words Ca^{2+}-calmodulin in solution has fewer widely separated scattering centers and more intermediately separated scattering centers than does crystalline Ca^{2+}-calmodulin. On the other hand, the $P(R)$ based on a flexible α helix agrees well with the experimental $P(R)$. Thus the x-ray-scattering studies indicate that in solution the interconnecting helix in Ca^{2+}-calmodulin is flexible and the globular lobes are, on average, about 10 angstroms closer together.

We repeated the x-ray studies, this time eliminating Ca^{2+} from the solution. Using both the pair-distribution function and Guinier analysis (Eqs. 8 and 9), we found that the radius of gyration of calmodulin is 5 percent smaller than that of Ca^{2+}-calmodulin. From complementary chemical studies we know that the increase in R_g due to binding of Ca^{2+} results in the exposure of hydrophobic regions in each globular lobe. Those nonpolar regions are thought to be important in the interactions of Ca^{2+}-calmodulin with target enzymes.

The issue of the flexibility of calmodulin's α helix remains controversial,
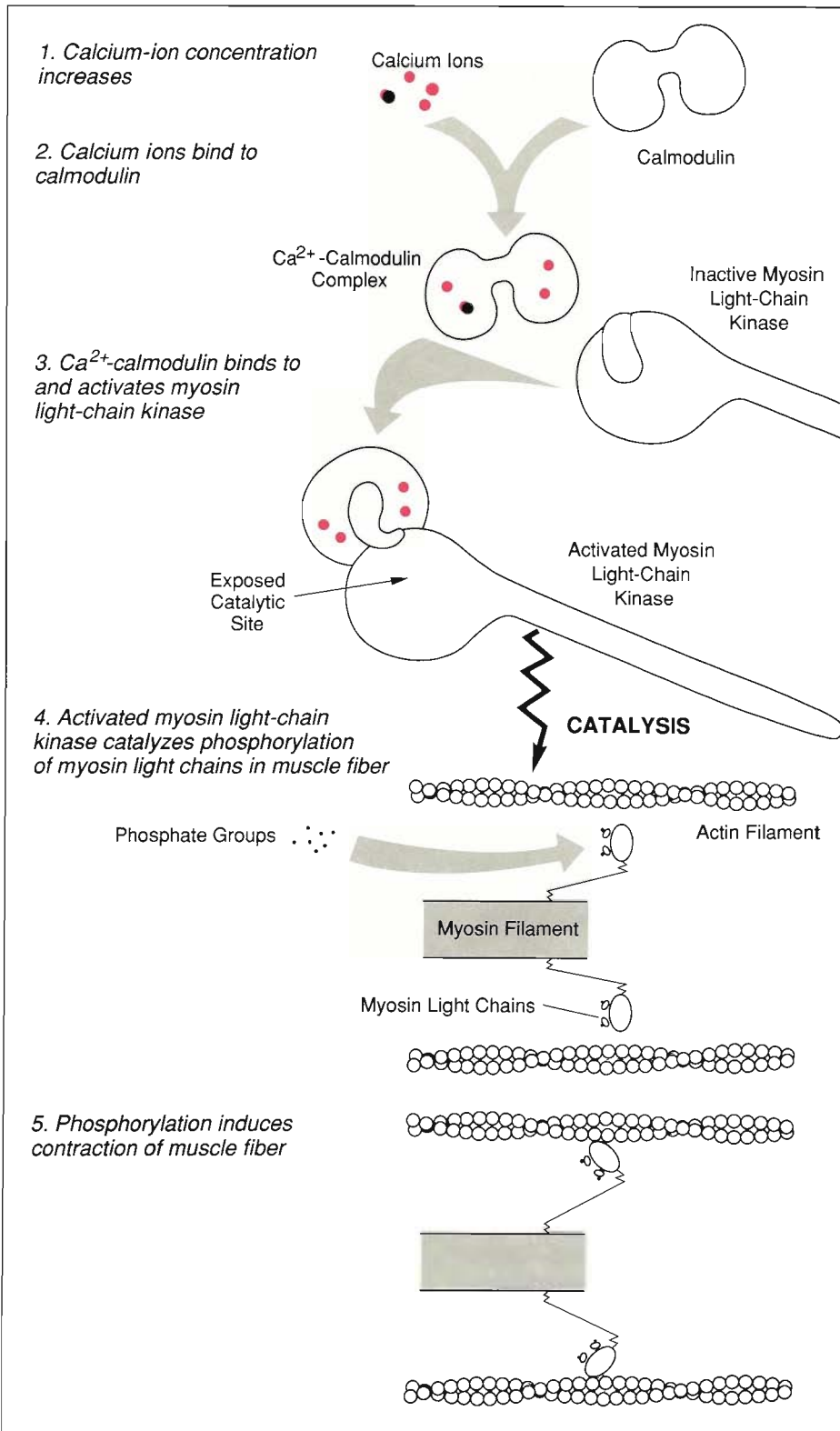

X-RAY RESULTS FOR Ca^{2+}-CALMODULIN IN SOLUTION

Fig. 6. Shown here is the pair-distribution function calculated from small-angle x-ray-scattering data for calmodulin in a solution containing sufficient Ca^{2+} to saturate calmodulin's Ca^{2+}-binding sites. Also shown are two theoretical pair-distribution functions and the pattern of scattering centers on which each is based. The crystal-structure scattering-center pattern was deduced from the crystal structure of Ca^{2+}-calmodulin; the flexible-model scattering-center pattern is identical except that the portion corresponding to calmodulin's interconnecting helix is allowed to be flexible rather than rigid. The fact that the pair-distribution function based on the flexible model more closely matches the experimental pair-distribution function suggests that in solution calmodulin's interconnecting helix is flexible.

in part because more than one structure may be consistent with the pair-distribution function obtained from the spherically averaged scattering data. To pin down the solution structure more precisely, we have attempted a novel experiment in which $^{240}Pu^{3+}$ is bound to the Ca^{2+}-binding sites in calmodulin. The triply charged ion of plutonium-240 has the same ionic radius as Ca^{2+}, and we have shown that it can substitute for Ca^{2+} in studies of biochemical processes dependent on Ca^{2+}-calmodulin. Plutonium-240 was chosen as a substitute for Ca^{2+} because 0.28-angstrom

neutrons form a short-lived bound state with plutonium-240 and thus have a much greater probability of scattering from plutonium-240 than from the other nuclei in calmodulin. That "resonant" scattering gave rise to a measurable interference pattern that directly reflects the distances between the Ca^{2+}-binding sites. With some further refinement the experiment will provide crucial information about the relationship between the two lobes. In addition, the small-angle scattering from Pu^{3+}-calmodulin showed that it has the same structure as Ca^{2+}-calmodulin.

1. Calcium-ion concentration increases

Calcium Ions

Calmodulin

2. Calcium ions bind to calmodulin

Ca^{2+}-Calmodulin Complex

Inactive Myosin Light-Chain Kinase

3. Ca^{2+}-calmodulin binds to and activates myosin light-chain kinase

Exposed Catalytic Site

Activated Myosin Light-Chain Kinase

CATALYSIS

4. Activated myosin light-chain kinase catalyzes phosphorylation of myosin light chains in muscle fiber

Phosphate Groups

Actin Filament

Myosin Filament

Myosin Light Chains

5. Phosphorylation induces contraction of muscle fiber

REGULATION OF MUSCLE CONTRACTION BY Ca^{2+}, CALMODULIN, AND MLCK

Fig. 7. Muscle contracts (and relaxes) as filaments of myosin and actin slide past each other. One molecular mechanism that enables the sliding motion involves calmodulin and the enzyme myosin light-chain kinase. A local increase in Ca^{2+} concentration causes binding of Ca^{2+} to calmodulin, and the resulting Ca^{2+}-calmodulin binds to MLCK and thereby activates a catalytic site on the enzyme. The activated MLCK then catalyzes the phosphorylation of a subunit of myosin, which enables the relative motion of myosin and actin.

Neutron Studies—The Interaction with Target Enzymes

Although x-ray and chemical studies by us and others have yielded knowledge about the structural changes in calmodulin that result from binding of calcium ions, much less is known about the molecular interactions of Ca^{2+}-calmodulin with its various target enzymes. The difficulties associated with isolating large quantities of purified target protein and the very high molecular weights of most target enzymes are major obstacles to structural studies.

Myosin light-chain kinase (MLCK) from rabbit skeletal muscle is one of the best characterized of the enzymes activated by Ca^{2+}-calmodulin. When Ca^{2+}-calmodulin binds to MLCK, the enzyme catalyzes the binding of phosphate groups to a subunit of the muscle protein myosin. The "phosphorylation" of myosin is thought to induce a twitch response in muscle cells (Fig. 7). The domain of MLCK to which Ca^{2+}-calmodulin binds was the first such domain to be identified and sequenced. It consists of twenty-seven amino-acid residues located toward one end of the polypeptide chain. A synthetic polypeptide with an amino-acid sequence iden-

Contrast Variation
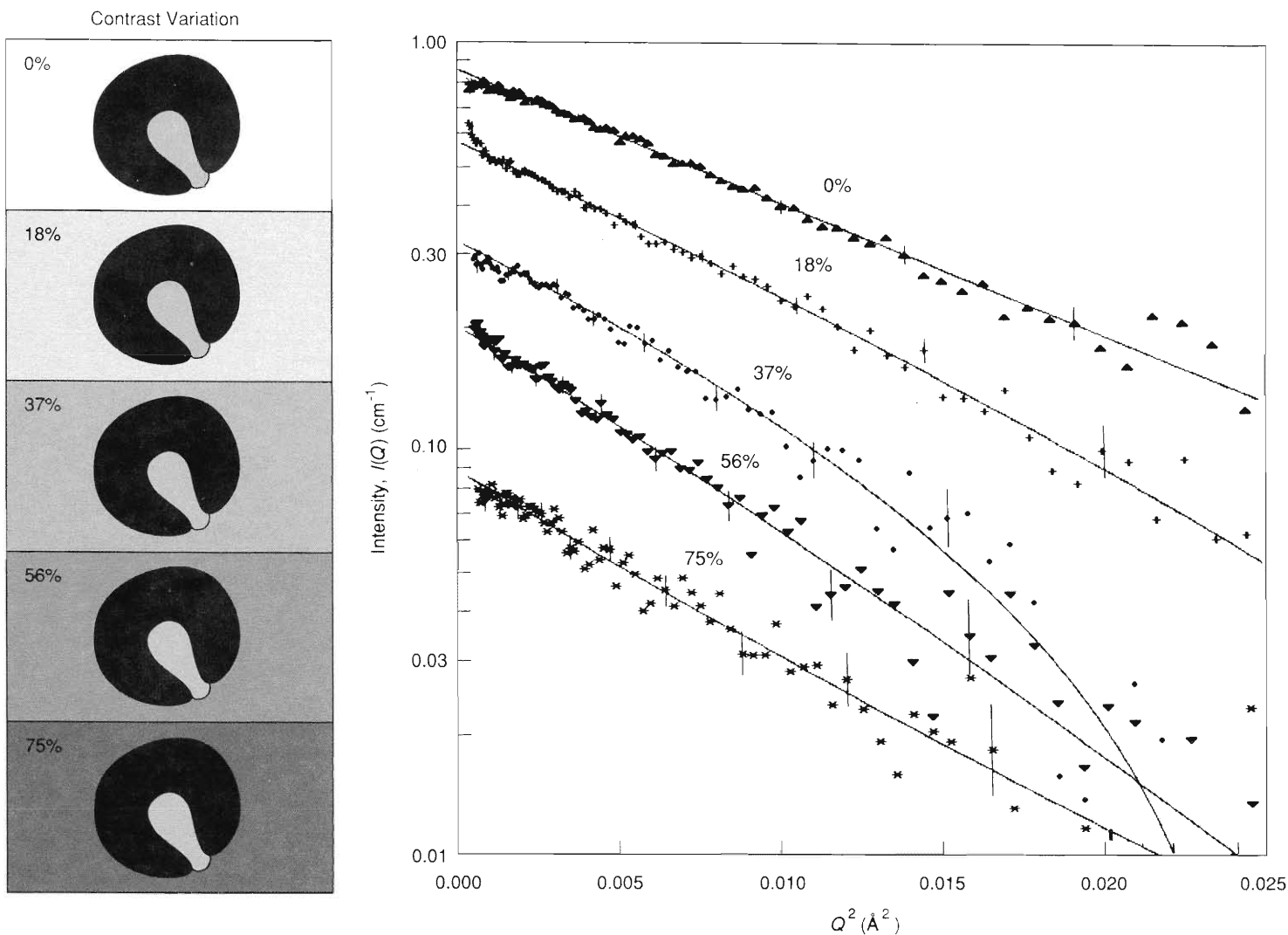
NEUTRON CONTRAST-VARIATION DATA FOR Ca^{2+}-CALMODULIN-MLCK-I

Fig. 8. Small-angle neutron-scattering data for Ca^{2+}-calmodulin complexed with the synthetic peptide MLCK-I were obtained at different contrast factors by varying the percentage of D$_2$O in the solvent. Since the calmodulin was deuterated, the contrast factor was greatest when the solvent was pure H$_2$O and progressively decreased as the percentage of D$_2$O was increased. Consequently, the maximum intensity decreases as the D$_2$0 content of the solvent increases. MLCK-I is contrast matched to a solvent containg 37 percent D$_2$O; therefore the observed scattering at that D$_2$O content is dominated by scattering from calmodulin. The figures alongside the data represent by shades of gray the scattering contrast among solvent, Ca^{2+}-calmodulin, and MLCK-I.

tical to that of the MLCK binding domain has been used in several instances as a model for studying the interaction between MLCK and Ca^{2+}-calmodulin. The synthetic polypeptide, referred to as MLCK-I, has been shown to bind one-to-one and with high affinity to Ca^{2+}-calmodulin. Moreover, binding of the synthetic polypeptide inhibits Ca^{2+}-calmodulin from activating MLCK and other target enzymes. Thus the binding of MLCK-I appears to be a good model for the binding of the entire enzyme. We knew, from circular-dichroism (selective absorption of circularly polarized

light) and nuclear-magnetic-resonance studies of calmodulin and MLCK-I, that upon binding, MLCK-I assumes an α-helical conformation and Ca^{2+}-calmodulin undergoes global structural changes. However, those studies did not provide details of the final structures of the two components of the complex.

We decided to do some small-angle x-ray- and neutron-scattering experiments to determine the relative positions of the two components in the complex as well as the overall shape changes that occur on binding. Our initial guess was that the structure of Ca^{2+}-calmodulin in the complex might be more like its crystal structure because of stabilizing interactions between the α helix of MLCK-I and the interconnecting α helix of calmodulin.

We collected neutron-scattering data for the complex formed by MLCK-I and Ca^{2+}-calmodulin in five aqueous solvents containing various percentages of D_2O. By using deuterated calmodulin, we ensured that Ca^{2+}-calmodulin and MLCK-I had different mean scattering-length densities. Figure 8 shows plots of $I(Q)$ versus Q^2 for the five solvents and the Fourier transforms of the pair-distribution functions fitted to the data (solid lines). Because the 37-percent D_2O solvent has approximately the same scattering-length density as MLCK-I, the solid line for the 37-percent D_2O solvent is a first-order approximation to the Fourier transform of the pair-distribution function for the complexed Ca^{2+}-calmodulin alone.

We also collected neutron-scattering data for deuterated Ca^{2+}-calmodulin in pure H_2O in the absence of MLCK-I. As shown in Figure 9a, the pair-distribution function for Ca^{2+}-calmodulin deduced from the neutron-scattering data is asymmetric and indicates an elongated structure consistent with the flexible model. Note also that Figure 9a is similar the pair-distribution function for Ca^{2+}-calmodulin in solution deduced

(a) Ca^{2+}-Calmodulin

(b) Ca^{2+}-Calmodulin-MLCK-I

R (Å)
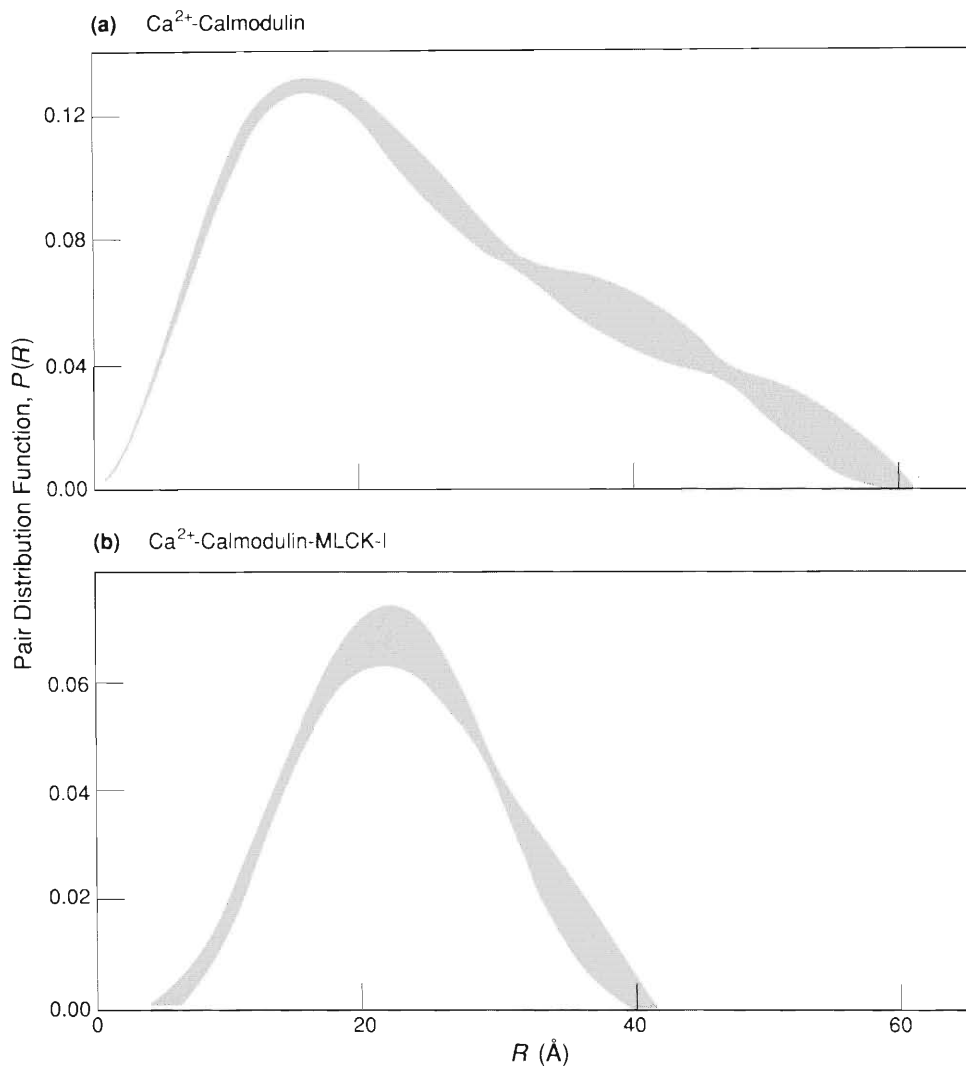
NEUTRON RESULTS FOR Ca^{2+}-CALMODULIN AND Ca^{2+}-CALMODULIN-MLCK-I

Fig. 9. The pair-distribution function for Ca^{2+}-calmodulin (a) resembles that of a double-lobed structure, whereas the pair-distribution function for Ca^{2+}-calmodulin-MLCK-I (b) resembles that of a spherical structure. Apparently when Ca^{2+}-calmodulin binds to the peptide, the lobes of calmodulin bend around the peptide. Both pair-distribution functions were deduced from small-angle neutron-scattering data. The solvent in both cases was pure H_2O.

from x-ray-scattering data (see Fig. 6). The differences that do exist can be attributed mostly to differences in the Q range measured but also to differences between the neutron scattering-length density and the electron density of Ca^{2+}-calmodulin. Both x-ray and neutron data support very similar

structures for Ca^{2+}-calmodulin in solution.

Now compare Figure 9a with Fig. 9b, the pair-distribution function for Ca^{2+}-calmodulin when it is compexed with MLCK-I in pure H_2O. The result is a surprise. Rather than becoming more like the crystal form, the structure of the

79

complexed Ca^{2+}-calmodulin becomes much more like a sphere, and its maximum dimension decreases by about 30 percent. Moreover, the radius of gyration of the Ca^{2+}-calmodulin-MLCK-I complex is much smaller than that of Ca^{2+}-calmodulin.

The data set in Fig. 8 contains much more structural information than do data obtained from a single solvent. In particular, through a method of analysis developed by Stuhrmann, the relative dispositions of the two proteins in the Ca^{2+}-calmodulin-MLCK-I complex can be deduced. We begin by determining values for R_g^2 and d_{max} from each pair-distribution function (Table 3). Note that the values of R_g^2 and d_{max} vary as the percentage of D_2O in the solvent varies. The smallest d_{max}, at 37-percent D_2O (where MLCK-I does not contribute to the scattering), characterizes the maximum dimension of Ca^{2+}-calmodulin when it is bound to MLCK-I. The variation of d_{max} among the results for the other solvents (to which both Ca^{2+}-calmodulin and MLCK-I contribute) reflects relatively large uncertainties arising from the finite range of Q for which we can measure $I(Q)$.

The variation in R_g is more interesting. As the solvent is varied, the contrast factor for each component in the complex changes, and so does the relative contribution of each component to R_g. Changes in R_g thus reflect variations in scattering-length density within the complex. Stuhrmann's method involves separating R_g^2 into a term that is independent of variations in scattering-length density, R_V^2, and terms that are not. That is,

$$R_g^2 = R_V^2 + \frac{\alpha}{b_c} - \frac{\beta}{b_c^2}, \qquad (10)$$

where $b_c \equiv \bar{b}_p - b_s$, the mean contrast factor of the whole complex. R_V can be interpreted as the radius of gyra-

---

Table 3

VARIATION OF STRUCTURAL PARAMETERS WITH CONTRAST FACTOR

Analysis of the pair-distribution function derived from small-angle neutron-scattering data provides the structural parameters $I(0)$, R_g, and d_{max}. Listed here are values of those parameters for Ca^{2+}-calmodulin dispersed in H_2O and for Ca^{2+}-deuterated calmodulin-MLCK-I dispersed in water containing various percentages of D_2O. As discussed in the text, the variation of R_g with contrast factor contains further structural information.

	Percentage D_2O	$I(0)$ (cm^{-1})	R_g (Å)	d_{max} (Å)
Ca^{2+}-calmodulin	0	0.965 ± 0.006	19.4 ± 0.2	59 ± 5
Ca^{2+}-($^2H_1$) calmodulin-MLCK-I	0	0.846 ± 0.005	15.1 ± 0.5	45 ± 5
	18	0.564 ± 0.004	16.0 ± 0.2	45 ± 5
	37	0.322 ± 0.004	16.7 ± 0.2	41 ± 5
	56	0.199 ± 0.006	18.2 ± 0.6	49 ± 5
	75	0.088 ± 0.002	18.0 ± 0.5	53 ± 10

---

tion of the complex at infinite contrast, that is, when the complex has a *uniform* scattering-length density, or no variation in scattering-length density. Thus, R_V is a measure of the overall shape of the whole complex and remains constant as the mean contrast factor is varied. The coefficient α is related to the second moment of $(b_p(\mathbf{r}) - b_s)$ about the mean and therefore measures the change of scattering-length density with distance from the center of scattering power. The coefficient β is related to the square of the first moment of $(b_p(\mathbf{r}) - b_s)$ and measures the separation between the centers of scattering power of the two scattering components.

To determine R_V, α, and β, we fit a parabola to a plot of R_g^2 versus b_c^{-1}. Obviously we must know b_c, and therefore $\bar{b}_p$, to perform the Stuhrmann analysis. We mentioned earlier that $\bar{b}_p$ can be determined from a linear fit to a plot of $\sqrt{I(0)}$ versus b_s.

We applied the Stuhrmann analysis to the values of R_g listed in Table 3. We obtained a positive value for

α, which indicates that the component with the higher scattering-length density (deuterated Ca^{2+}-calmodulin) is situated more toward the outside of the complex and the component with the lower scattering-length density (MLCK-I) is situated more toward the inside of the complex. The nonzero value derived for β indicates that the centers of scattering power of the two proteins are not coincident.

The values we derived for α, β, and R_V from the Stuhrmann analysis had large errors. The mean contrast factors of all the samples we studied were positive because $\bar{b}_p$, the mean scattering-length density of the complex containing deuterated Ca^{2+}-calmodulin and MLCK-I exceeded the scattering-length density of pure D_2O. As a result, we calculated R_V by extrapolation rather than by interpolation. The resulting uncertainty in R_V results in large errors in α and β.

Nevertheless, our qualitative results for α and β are consistent with the idea that, when MLCK-I interacts with Ca^{2+}-calmodulin, MLCK-I binds near

calmodulin's interconnecting helix and the two globular lobes of calmodulin form additional contacts with each other by closing in around at least part of the peptide. An artist's conception of the configuration is shown in Fig. 10. Thus our neutron-scattering experiments have yielded a surprising view of the interaction between Ca^{2+}-calmodulin and the binding domain of one target enzyme.

That view is not necessarily applicable to calmodulin's interaction with all of its target enzymes. In particular, our experiments on phosphorylase kinase (PhK) show rather different results. PhK, a large enzyme containing many subunits, catalyzes the transfer of a phosphate group from ATP to glycogen phosphorylase, thereby activating glycogen phosphorylase as a catalyst in the conversion of glycogen to glucose (Fig. 11). PhK is made up of four identical copies each of four different polypeptides, denoted α, β, γ, and δ. The γ subunit contains the catalytic site; the α, β, and δ subunits play regulatory roles in response to various physiological signals. The protein composing the δ subunit is calmodulin. When Ca^{2+} binds to those intrinsic calmodulin molecules, the catalytic activity of the γ subunit increases twentyfold. Calmodulin's activation of PhK is fundamentally different from its activation of MLCK in that the calmodulin remains associated with the δ subunit in the absence of Ca^{2+}.

It has recently been shown that Ca^{2+}-calmodulin binds to the γ subunit at two noncontiguous sites, each containing twenty-five amino-acid residues. The binding between Ca^{2+}-calmodulin and the γ subunit in PhK apparently encompasses a more extensive region than does the binding of Ca^{2+}-calmodulin to MLCK.

We have recently completed x-ray studies of Ca^{2+}-calmodulin complexed with one or the other or both of two synthetic polypeptides corresponding to
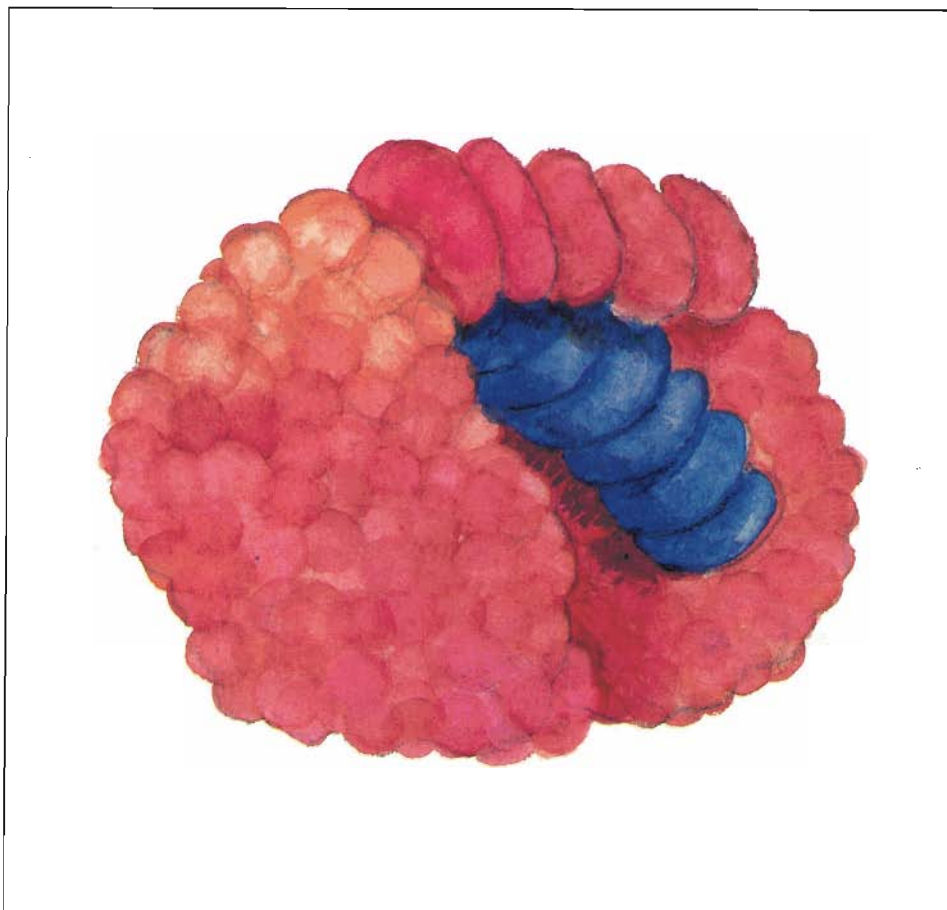
STRUCTURE OF Ca^{2+}-CALMODULIN BOUND TO MLCK-I

Fig. 10. This artist's conception of Ca^{2+}-calmodulin bound to MLCK-I is based on small-angle neutron-scattering data. Those data suggest that the α-helical peptide is in contact with calmodulin's interconnecting α helix and that the two globular lobes of calmodulin bend around the peptide to form additional contacts.

the two binding domains on the γ subunit of PhK. The two synthetic polypeptides are designated PhK5 and PhK13. In addition, we have gathered neutron-scattering data for deuterated Ca^{2+}-calmodulin complexed with both synthetic polypeptides. The results are again a surprise. The binding of PhK5 alone results in a contraction of Ca^{2+}-calmodulin quite similar to that induced by MLCK-I. The x-ray-scattering data for the complex give values of 17.1 angstroms and 50 angstroms for

R_g and d_{max}, respectively. Apparently PhK5 and MLCK-I bind similarly to calmodulin. The amino-acid sequences of PhK5 and MLCK-I are such that both peptides are predicted to have a high propensity for forming an amphipathic helix, a configuration that may facilitate interaction with the interconnecting helix of calmodulin. The sequence of PhK13 is such that the peptide is predicted to have little or no α helix but significant β structure. The x-ray data for Ca^{2+}-calmodulin com-

REGULATION OF GLYCOGEN DEGRADATION BY Ca^{2+} THROUGH CALMODULIN INTRINSIC TO PhK

Fig. 11. Another target enzyme of Ca^{2+}-calmodulin is phosphorylase kinase (PhK), a large (1.3 million daltons) enzyme consisting of four identical copies of each of four different polypeptide subunits, one of which is calmodulin. PhK is involved in the cascade of interactions, initiated by a hormone binding to a cell, that results in the degradation of glycogen to glucose. Metabolism of the glucose then provides energy for use by the cell in various processes, including muscle contraction. The activation of PhK by Ca^{2+}-calmodulin differs from that of most of Ca^{2+}-calmodulin's target enzymes in that the calmodulin is an intrinsic subunit of PhK. Binding of Ca^{2+} to those intrinsic calmodulin subunits activates the subunits containing the sites at which catalysis of the activation of glycogen phosphorylase occurs.

plexed to PhK13 show no contraction at all, but rather an overall extension of the structure ($R_g \approx 28$ angstroms and $d_{max} \approx 90$ angstroms). The complex with both PhK5 and PhK13 is also extended, as indicated by R_g and d_{max} values not significantly different from those of Ca^{2+}-calmodulin-PhK13.

Neutron-scattering data for deuterated Ca^{2+}-calmodulin complexed with both PhK5 and PhK13 add further insight into the nature of the complex. The scattering-length densities of both peptides are approximately equal to that of a solvent containing 40 percent D_2O. The pair-distribution function determined from scattering data for the complex dispersed in such a solvent shows that calmodulin itself is extended ($R_g \approx 26$ angstroms and $d_{max} \approx 90$ angstroms). That result indicates that PhK13 influences the way in which Ca^{2+}-calmodulin binds to PhK5. In particular, PhK13 apparently prevents PhK5 from inducing a conformational change in Ca^{2+}-calmodulin similar to the contraction
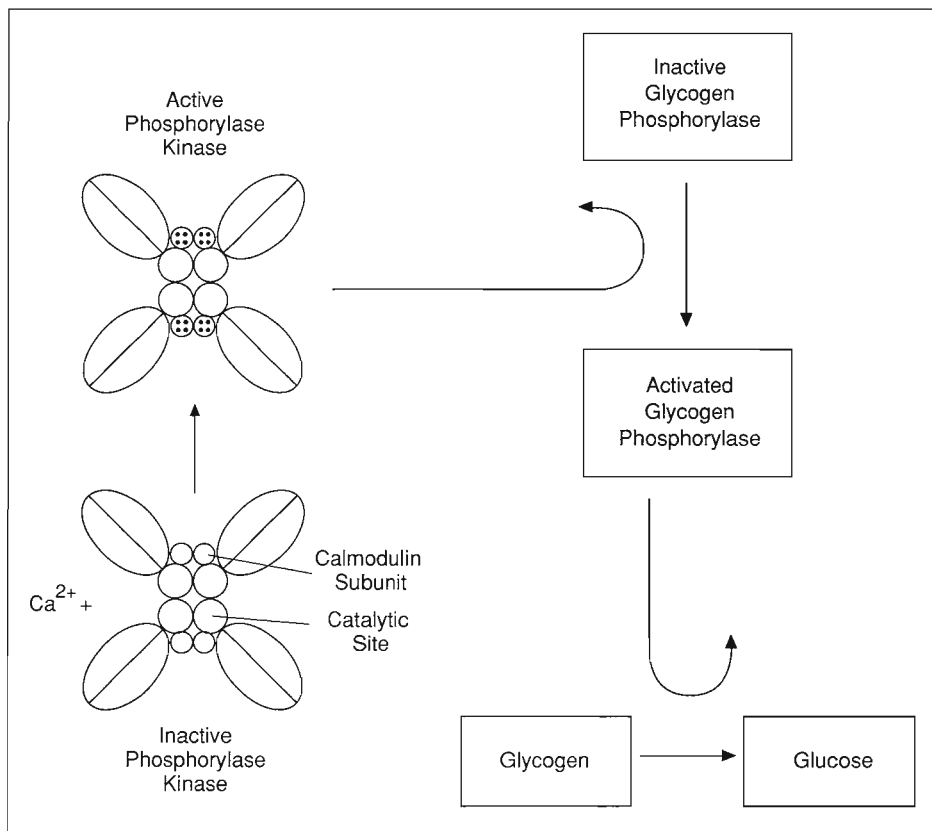
induced by MLCK-I. The differences between the conformational changes induced in Ca^{2+}-calmodulin by the two target enzymes (phosphorylase kinase and myosin light-chain kinase) are likely to be due to differences in the nature of its interactions with the two enzymes, and they provide a clue as to the purpose of the unusual interconnecting helix in calmodulin. That flexible domain clearly facilitates changes in the relationship between the globular domains of calmodulin that allow its binding to a diversity of target enzymes and its activation of a wide variety of biochemical processes. Our goal is to use neutron-scattering data, in conjunction with other structural data, to build up a complete picture of the molecular basis for Ca^{2+}-dependent activation of target enzymes by calmodulin. Future studies will focus on developing methods

for preparing complexes of deuterated calmodulin with intact target enzymes. In addition, the interactions of Ca^{2+}-calmodulin with other target enzymes will be investigated.

As is apparent from this discussion, unraveling the molecular basis of a biological function is a long and arduous task. Neutron-scattering studies, although difficult, are providing a view of biological molecules in functional states that were heretofore unobservable. Their role in studies of biological function should continue to grow and to yield important clues concerning the structural dynamics of complex biological units. ∎
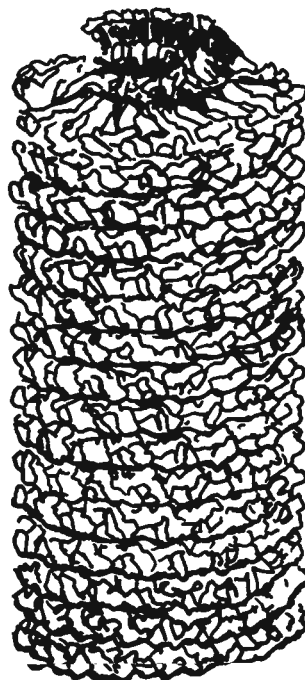
FERROFLUIDS
a new alignment technique

Typical neutron-scattering experiments on particles in solution yield only one-dimensional (spherically averaged) data because the scattering particles are randomly oriented. Such data can provide a model for the general shape and the boundaries of the scattering particles, but often more than one model is consistent with the data. However, if the particles are partially ordered by being given a definite orientation, the resulting scattering data contain more detailed structural information.

Alignment in one dimension has been achieved for the rod-shaped tobacco mosaic virus by applying shearing forces to an aqueous gel of the viruses. Two-dimensional x-ray-diffraction data for the aligned viruses led to a complete three-dimensional structure at a resolution of 3.6 angstroms (Fig. 1). Although that result is very impressive, the technique is not applicable to many biological structures.

Here we present preliminary studies of a new technique for aligning elongated biological assemblies in solution. The technique involves dispersing the assemblies in a ferrofluid (a fluid in which magnetic particles are suspended) and applying a moderate magnetic field. (Note that because the magnetic particles generally contain iron, which absorbs x rays strongly, the ferrofluid alignment technique is applicable to neutron-scattering experiments but not to x-ray-scattering experiments.) Magnetic forces cause the moments, or spins, of the magnetic particles to align along the direction of the mag-

STRUCTURE OF TMV

Fig. 1. The rod-shaped tobacco mosaic virus is about 3000 angstroms long and 180 angstroms in diameter. Its structure at a resolution of 3.6 angstroms was deduced from x-ray diffraction data for an aqueous gel of viruses. The viruses were oriented in one direction by shearing forces. The viral genetic material (RNA) is enclosed within a protein coat consisting of a helical array of many copies of a single protein species. The distance between the turns of the helix is 23 angstroms.
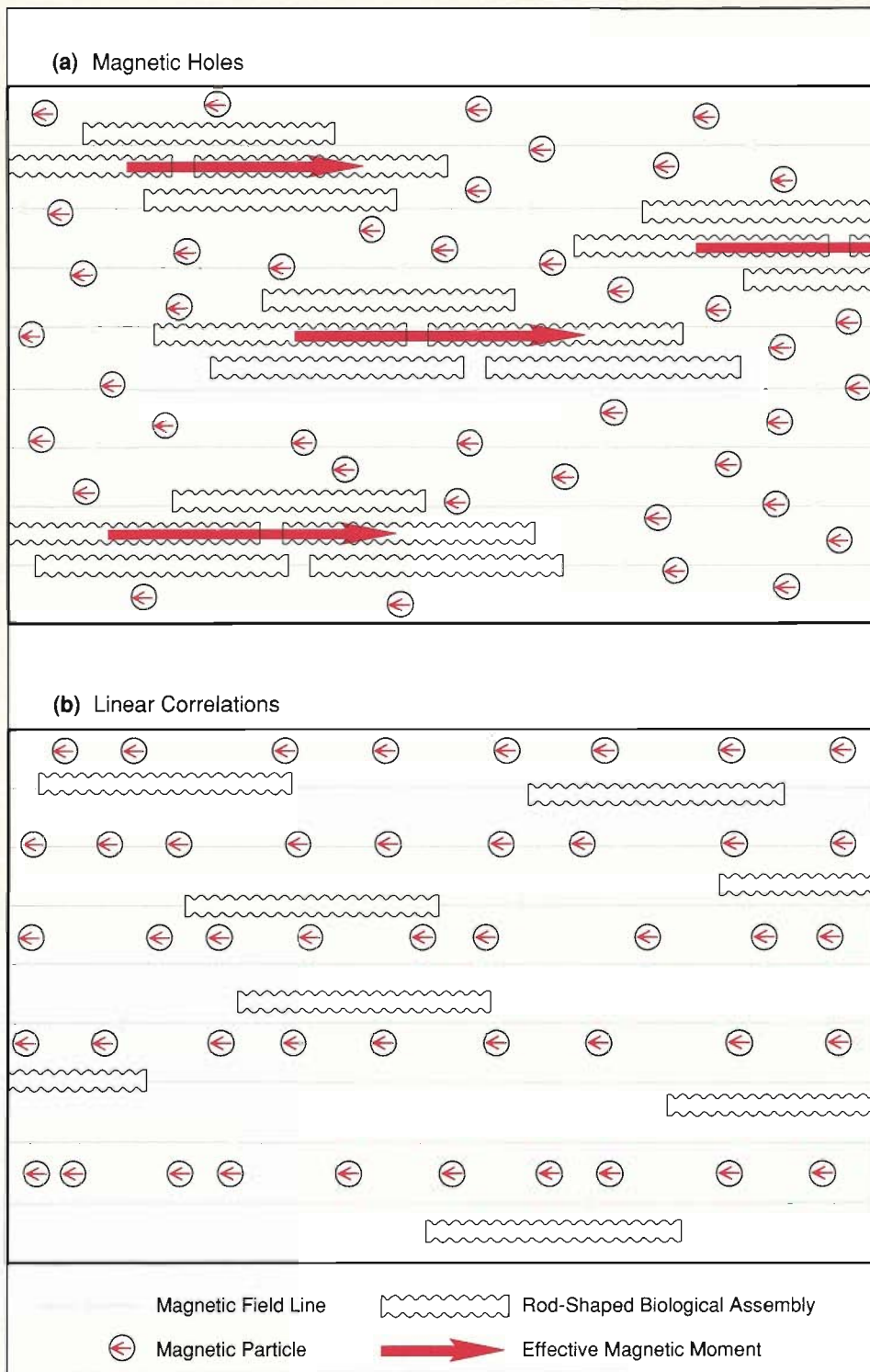
netic field, and their alignment, in turn, causes alignment of the elongated biological assemblies. Some biological assemblies have intrinsic diamagnetic moments and will therefore align along a magnetic field in the absence of mag-

netic particles. However, very strong magnetic fields are usually required. In contrast, the ferrofluid technique requires no intrinsic magnetic properties of the biological assemblies and only moderate magnetic fields.

We have tested the ferrofluid technique on two viruses, the tobacco mosaic virus (TMV) and the tobacco rattle virus (TRV), and have obtained neutron-diffraction data that testify to its success. Our tests focused not only on obtaining a high degree of alignment but also on understanding how the alignment comes about. Figure 2 shows two possible mechanisms. One possibility is that the biological assemblies act like "magnetic holes." That is, by displacing the ferrofluid, they acquire effective magnetic moments equal in magnitude and opposite in sign to the sum of the moments of the magnetic particles in the displaced ferrofluid. The effective moments then align along the applied magnetic field. In our experiments, however, the effective moments would be large enough to cause alignment of the assemblies in moderate magnetic fields only if the assemblies existed as ordered domains, or aggregates. Then each domain becomes a magnetic hole oriented along the applied field (Fig. 2a). The other possible mechanism of alignment invokes linear correlations among the magnetic particles, that is, a tendency for them to line up in rows along the applied field. Because disrupting those linear correlations would require energy, the nonmagnetic elongated assemblies also tend to line up, with their long axes along the field, in rows between the rows of magnetic particles (Fig. 2b). It seems likely at this time that both alignment mechanisms operate to different degrees depending on the relative concentrations of magnetic particles and biological assemblies and whether the solution conditions favor the formation of ordered domains of the biological assemblies.

(a) Magnetic Holes

(b) Linear Correlations

Magnetic Field Line ~~~~~~~ Rod-Shaped Biological Assembly

Magnetic Particle ⊖ Effective Magnetic Moment ➡

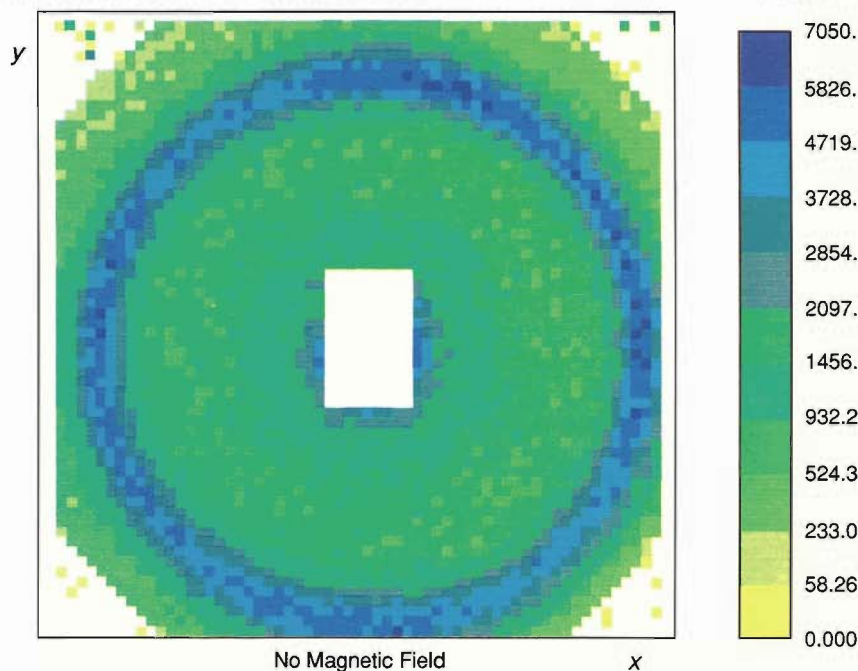## POSSIBLE MECHANISMS OF ALIGNMENT IN A FERROFLUID

Fig. 2. Elongated biological assemblies dispersed in a ferrofluid align along a moderate applied magnetic field even if they lack intrinsic magnetic properties. Shown here are two possible mechanisms of alignment. (a) Ordered domains of the biological assemblies act like magnetic holes. In other words, the domains acquire effective magnetic moments equal in magnitude and opposite in direction to the sum of the magnetic moments of the magnetic particles they displace. The effective magnetic moments cause the biological assemblies to align along the field. (b) Long-range magnetic correlations among the magnetic particles cause them to align in rows along the applied field. Lacking the energy to disrupt the linear correlations, the elongated biological assemblies also align along the field.

The magnetic particles in the ferrofluid would normally contribute to the observed scattering intensity. However, that contribution can be eliminated by matching the scattering-length density of the magnetic particles to that of the solvent. By a happy coincidence of nature, the neutron scattering-length density of the ferromagnetic material magnetite is quite close to that of D_2O. As is well known, particles of magnetite align with a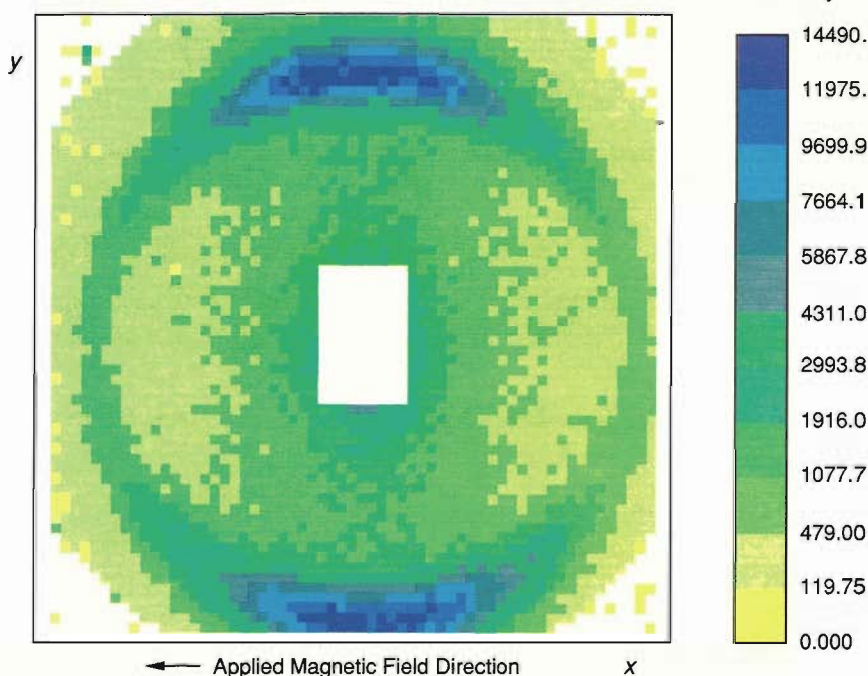n applied magnetic field. To achieve a reasonably homogeneous suspension of magnetite particles in D_2O, the particles must be coated with a detergent surfactant. The detergent must be carefully chosen, however, to avoid its degrading the biological assemblies. Through extensive electron microscopy we identified a detergent that did not degrade TMV or TRV. We then deuterated the detergent to match the scattering-length densities of magnetite and D_2O.

To find out how well our ferrofluid

(a) Isotropic Diffraction Peak at 363 angstroms

Intensity

y

■	7050.
■	5826.
■	4719.
■	3728.
■	2854.
■	2097.
■	1456.
■	932.2
■	524.3
■	233.0
■	58.26
■	0.000

No Magnetic Field   *x*

(b) Anisotropic Diffraction Peak at 363 angstroms

Intensity

y

■	14490.
■	11975.
■	9699.9
■	7664.1
■	5867.8
■	4311.0
■	2993.8
■	1916.0
■	1077.7
■	479.00
■	119.75
■	0.000

◄— Applied Magnetic Field Direction   *x*

SMALL-ANGLE NEUTRON-SCATTERING DATA FOR TMV IN A FERROFLUID

Fig. 3. Color-coded intensity data for neutrons scattered from tobacco mosaic virus dispersed in a ferrofluid. The intensity corresponding to the color of a detector pixel is related to the number of neutrons detected in the pixel. (a) In the absence of a magnetic field, the scattering is isotropic and a diffraction peak (a circle of high intensity) appears at a scattering angle corresponding to a distance of 363 angstroms. The diffraction peak reflects the packing distance between individual viruses in ordered domains. (b) When a magnetic field of 0.3 tesla is applied along the *x* direction, the diffraction peak appears only in the *y* direction, that is, perpendicular to the applied field. That result means that the packing distances are aligned perpendicular to the field and the long axes of the viruses are aligned along the field. The degree of alignment cannot be explained by intrinsic diamagnetism of the viruses and must therefore be caused by the ferrofluid.
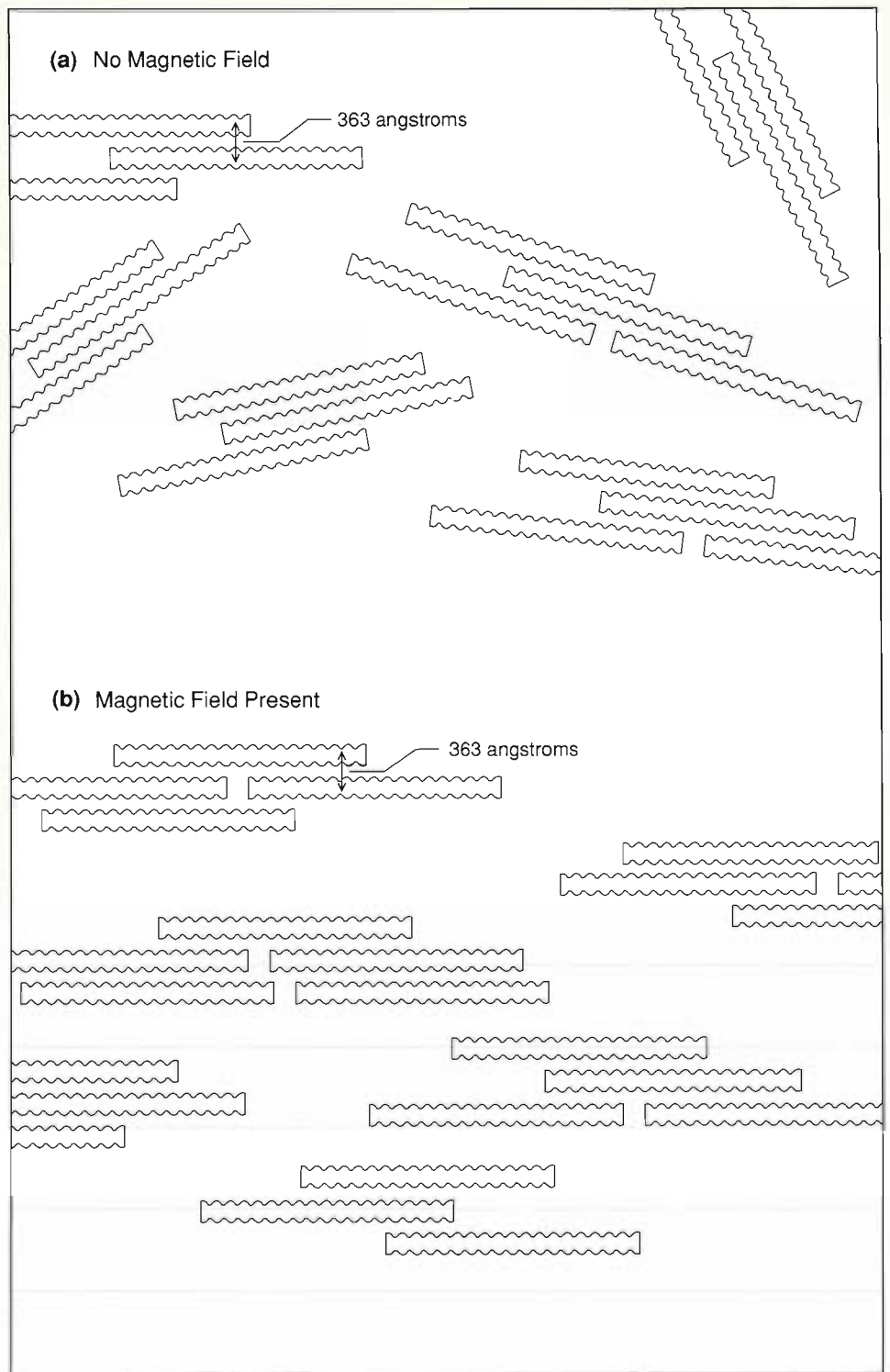
worked, we obtained neutron-scattering data for 0.06-volume-fraction dispersions of each virus in the ferrofluid. In the absence of a magnetic field, the scattering from TMV was isotropic, as expected (Fig. 3a). We also saw a diffraction peak corresponding to a distance of 363 angstroms. That diffraction peak arises because the viruses formed ordered domains in which the packing distance was 363 angstroms (Fig. 4a). (The packing distance is determined by

a repulsive electrostatic force between the viruses and varies with the ionic strength of the solution.) In a modest magnetic field of 0.3 tesla, the scattering was anisotropic, and the diffraction peak at 363 angstroms appeared only in the direction perpendicular to the field (Fig. 3b). Those results suggest that the TMV domains were aligned in such a way that the long axes of the viruses were parallel to the applied field.

If the TMV domains do so align,

ORDERED AGGREGATES OF TMV IN A FERROFLUID

Fig. 4. Tobacco mosaic viruses tend to form ordered domains in solution, as shown in the electron micrograph below. The viruses within a domain are kept at a certain distance by electrostatic forces. That distance depends on the pH and ionic strength of the solution and in our experiments was 363 angstroms. (a) In the absense of a magnetic field, the domains are randomly oriented in a ferrofluid. (b) An applied magnetic field aligns the magnetic particles in the ferrofluid, which, in turn, causes the domains to align along the field. Thus the 363-angstrom spacing between the viruses is perpendicular to the field and produces diffraction peaks only in that direction.

(a) No Magnetic Field

363 angstroms

(b) Magnetic Field Present

363 angstroms

another diffraction peak should be seen, one produced by the 23-angstrom distance between the turns of the helical viral coat protein (see Fig. 1). That peak should occur at scattering angles much greater than those included in the data of Fig. 3 and at a direction perpendicular to the direction at which the diffraction peak due to the packing distance is observed. We moved the detector closer to the sample and did indeed see a diffraction peak corresponding to a
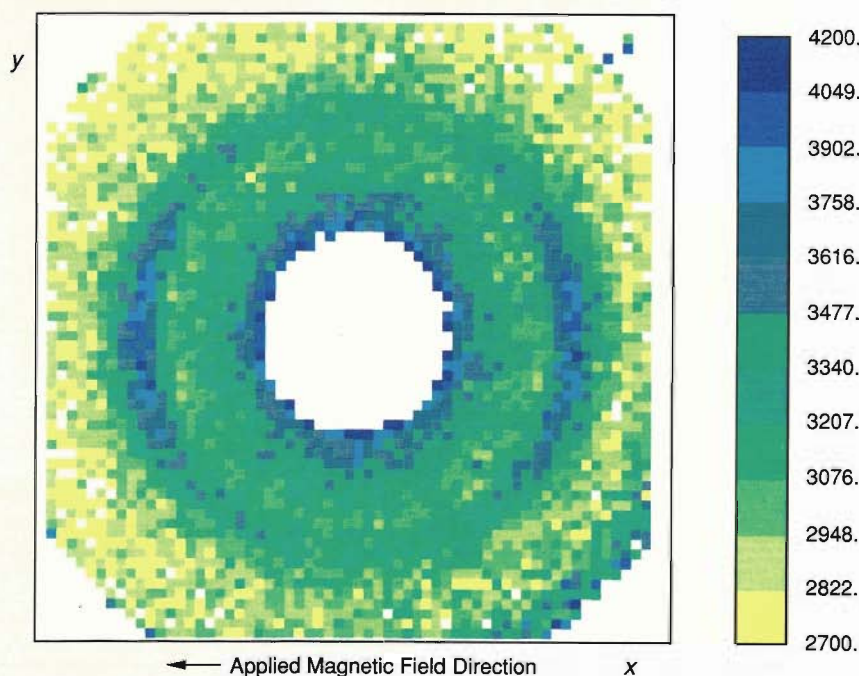
23-angstrom helical pitch (Fig. 5a). The existence of the peak confirms the structural integrity of the viruses in the ferrofluid, and its orientation parallel to the magnetic field is consistent with alignment of the long axes of the viruses parallel to the magnetic field (Fig. 4b).

Although our experiments demonstrated alignment, we wondered whether it was caused by intrinsic magnetic properties of TMV or by one of the mechanisms depicted in Fig. 2. The

(a) Diffraction Peak at 23 angstroms

Intensity

| 25000. |
| 22986. |
| 21057. |
| 19212. |
| 17453. |
| 15777. |
| 14186. |
| 12680. |
| 11258. |
| 9921.0 |
| 8668.2 |
| 7500.0 |

← Applied Magnetic Field Direction   *x*

(b) Diffraction Peak at 25 angstroms

Intensity

| 4200. |
| 4049. |
| 3902. |
| 3758. |
| 3616. |
| 3477. |
| 3340. |
| 3207. |
| 3076. |
| 2948. |
| 2822. |
| 2700. |

← Applied Magnetic Field Direction   *x*

NEUTRON DIFFRACTION PEAKS FROM TMV AND TRV

Fig. 5. Color-coded intensity data for (a) tobacco mosaic virus and (b) tobacco rattle virus dispersed in a ferrofluid in the presence of a 0.3-tesla applied magnetic field. Note the diffraction peaks at 23 and 25 angstroms, which correspond to the regular spacing between the turns of the helical protein coats of TMV and TRV, respectively. The diffraction peaks appear only in the *x* direction, that is, along the direction of the applied field. Thus, the ferrofluid has aligned the viruses with their long axes along the applied magnetic field.

trated samples of TRV did not align in a normal solution even in a high applied magnetic field (7 teslas). However, when TRV was dispersed in the ferrofluid, a field of 0.5 tesla was sufficient to produce alignment comparable to that obtained for TMV in phosphate buffer. Figure 5b shows a diffraction peak corresponding to the 25-angstrom pitch of the helical TRV protein coat, similar to the peak obtained for TMV in Fig. 5a. The position of the peak indicates that the viruses are aligned parallel to the applied magnetic field. Moreover, like TMV in phosphate buffer, TRV showed no evidence of any domains.

Our preliminary studies demonstrated that the ferrofluid technique can orient elongated biological assemblies irrespective of their intrinsic magnetic properties and without disrupting their structural integrity. Furthermore, the orientation is sufficiently great to facilitate measurement of internal structural parameters. Future studies will concentrate on increasing the degree of alignment so that higher-resolution data can be obtained. ∎

viruses could have been ordered by diamagnetic moments; however, the magnetic field strength was smaller than is generally needed to achieve the degree of alignment indicated by the data. Such alignment often requires the presence of ordered domains in the sample. So we repeated the experiments, this time adding phosphate buffer, the ions of which are known to disrupt domain structure. Thus alignment based on the intrinsic diamagnetism of the viruses is

not expected in the presence of phosphate buffer. We found that the viruses were still aligned but to a lesser degree. That result suggests not only that the ordering is due to the presence of the ferrofluid but also that the mechanism depicted in Fig. 2b may enhance the alignment of TMV.

The tobacco rattle virus is genetically unrelated but morphologically quite similar to TMV. TRV shows only poor orientation under shear, and concen-

Acknowledgments

Further Reading

Benno P. Schoenborn, editor. 1976. *Neutron Scattering for the Analysis of Biological Structures: Report of Symposium Held June 2–6, 1975.* Brookhaven National Laboratory report 50543.

B. Jacrot. 1976. The study of biological structures by neutron scattering from solution. *Reports on Progress in Physics* 39: 911–953.

O. Glatter and O. Kratky, editors. 1982. *Small Angle X-ray Scattering.* London: Academic Press.

Jill Trewhella, J.-L. Popot, D. M. Engelman, and G. Zaccai. 1986. Neutron diffraction studies of bacteriorhodopsin. *Physics* 136B: 249.

Jill Trewhella, Jean-Luc Popot, Giuseppe Zaccaï, and Donald M. Engelman. 1986. Localization of two chymotryptic fragments in the structure of renatured bacteriorhodopsin by neutron diffraction. *The EMBO Journal* 5: 3045.

Philip Cohen and Claude B. Klee, editors. 1988. "Calmodulin." In *Molecular Aspects of Cellular Regulation,* edited by Philip Cohen. Amsterdam: Elsevier.

Douglas B. Heidorn and Jill Trewhella. 1988. Comparison of the crystal and solution structures of calmodulin and troponin C. *Biochemistry* 27: 909.
D. B. Heidorn, P. A. Seeger, S. E. Rokop, D. K. Blumenthal, A. R. Means, H. Crespi, and J. Trewhella. 1989. Changes in the structure of calmodulin induced by a peptide based on the calmodulin-binding domain of myosin light chain kinase. *Biochemistry* 28: 6757.

R. A. Kowluru, D. B. Heidorn, S. P. Edmondson, M. W. Bitensky, A. Kowluru, N. W. Downer, T. W. Whaley, and J. Trewhella. 1989. Glycation of calmodulin: Chemistry and structural and functional consequences. *Biochemistry* 28: 2220.

D. B. Heidorn and J. Trewhella. Low-resolution structural studies of proteins in solution: Calmodulin. Accepted for publication in *Comments on Molecular and Cellular Biophysics: A Journal of Critical Discussion of the Current Literature (Comments on Modern Biology A).*

J. Trewhella, D. B. Heidorn, and P. A. Seeger. Solution structures of calcium-binding proteins: A small-angle-scattering study. *Journal of Molecular Crystals and Liquid Crystals* 180A: 45-53.

J. Trewhella, D. K. Blumenthal, S. Rokop, and P. A. Seeger. The solution structure of calmodulin complexed with two peptides based on the regulatory domain of the catalytic subunit of phosphorylase kinase. Submitted to *Biochemistry.*

Keiichi Namba and Gerald Stubbs. 1986. Structure of tobacco mosaic virus at 3.6 Å resolution: Implications for assembly. *Science* 231: 1401.

John B. Hayter, Roger Pynn, Stuart Charles, Arne T. Skjeltorp, Jill Trewhella, Gerald Stubbs, and Peter Timmins. 1989. Ordered macromolecular structures in ferrofluid mixtures. *Physical Review Letters* 62: 1667.

Jill Trewhella is an Australian-born biophysicist. She received her B.Sc. and M.Sc. degrees in physics at the University of New South Wales. Her studies there in molecular structure included an x-ray-crystallographic analysis of an anti-cancer agent that interacts with DNA to inhibit replication. That work directed her interests toward structural biology. Her Ph.D. research at the University of Sydney utilized physical techniques, in particular NMR spectroscopy, to study the structure and function of heme proteins. She came to the United States in the fall of 1980 as a postdoctoral fellow at Yale University, where she used neutron diffraction to study the structure of the membrane protein bacteriorhodopsin. She came to the Laboratory's Life Sciences Division in 1984 to develop a structural biology program in association with the then newly developing neutron-scattering center. Since that time her research program in several areas of structural biology has utilized neutron and x-ray scattering, as well as other biophysical techniques. One of her major interests is in the molecular mechanisms of calcium regulation. She is also actively involved in structural studies of chromatin and molecules of the immune system.
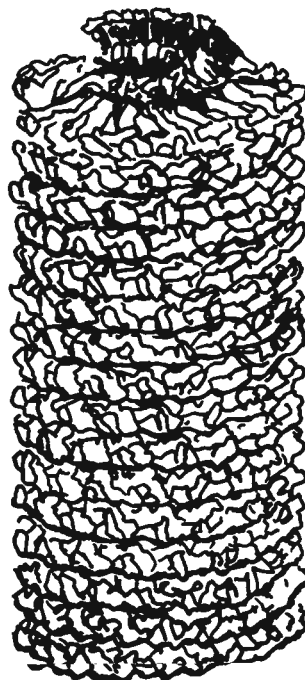
FERROFLUIDS
a new alignment technique

Typical neutron-scattering experiments on particles in solution yield only one-dimensional (spherically averaged) data because the scattering particles are randomly oriented. Such data can provide a model for the general shape and the boundaries of the scattering particles, but often more than one model is consistent with the data. However, if the particles are partially ordered by being given a definite orientation, the resulting scattering data contain more detailed structural information.

Alignment in one dimension has been achieved for the rod-shaped tobacco mosaic virus by applying shearing forces to an aqueous gel of the viruses. Two-dimensional x-ray-diffraction data for the aligned viruses led to a complete three-dimensional structure at a resolution of 3.6 angstroms (Fig. 1). Although that result is very impressive, the technique is not applicable to many biological structures.

Here we present preliminary studies of a new technique for aligning elongated biological assemblies in solution. The technique involves dispersing the assemblies in a ferrofluid (a fluid in which magnetic particles are suspended) and applying a moderate magnetic field. (Note that because the magnetic particles generally contain iron, which absorbs x rays strongly, the ferrofluid alignment technique is applicable to neutron-scattering experiments but not to x-ray-scattering experiments.) Magnetic forces cause the moments, or spins, of the magnetic particles to align along the direction of the mag-

STRUCTURE OF TMV

Fig. 1. The rod-shaped tobacco mosaic virus is about 3000 angstroms long and 180 angstroms in diameter. Its structure at a resolution of 3.6 angstroms was deduced from x-ray diffraction data for an aqueous gel of viruses. The viruses were oriented in one direction by shearing forces. The viral genetic material (RNA) is enclosed within a protein coat consisting of a helical array of many copies of a single protein species. The distance between the turns of the helix is 23 angstroms.
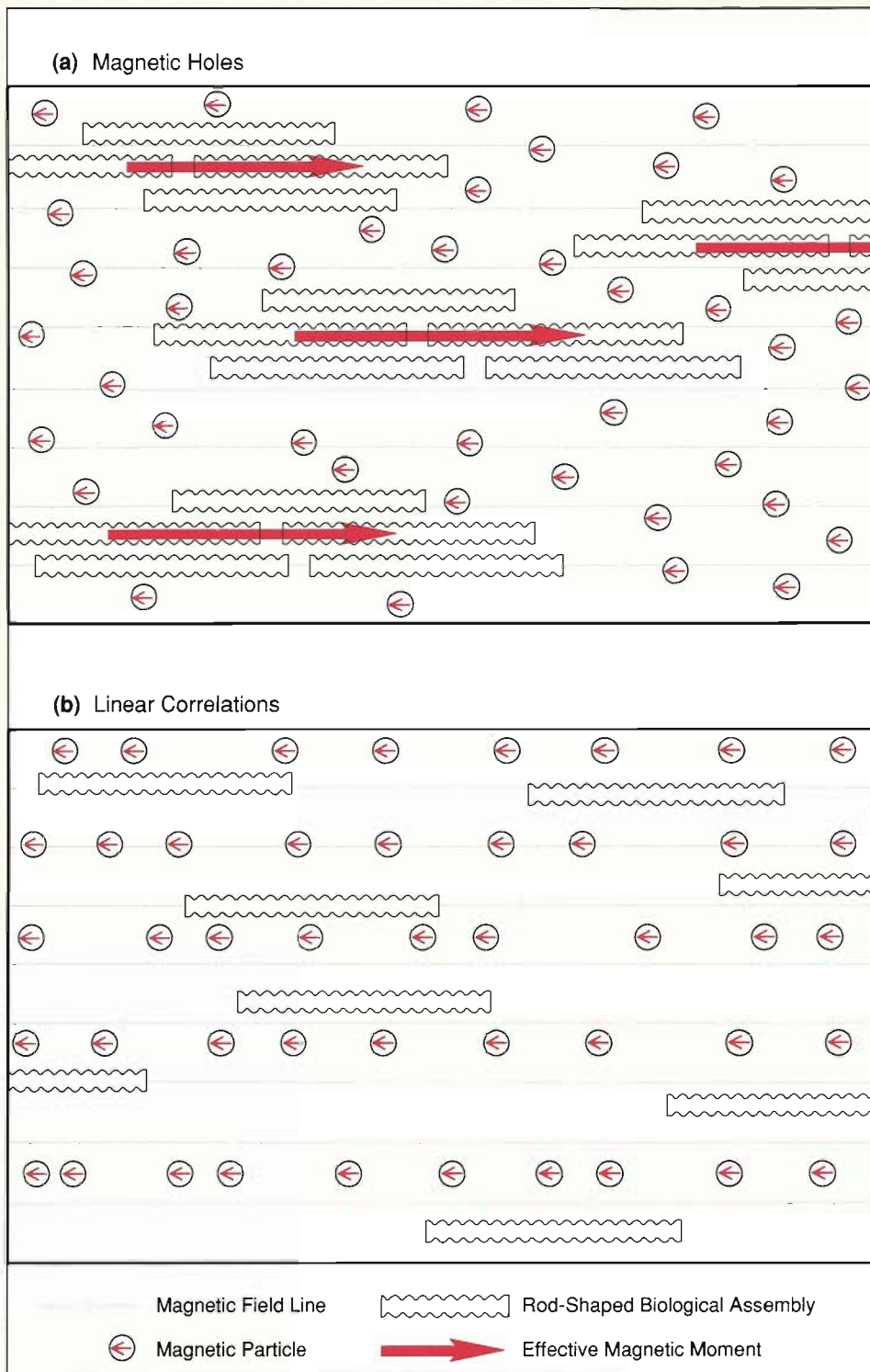
netic field, and their alignment, in turn, causes alignment of the elongated biological assemblies. Some biological assemblies have intrinsic diamagnetic moments and will therefore align along a magnetic field in the absence of mag-

netic particles. However, very strong magnetic fields are usually required. In contrast, the ferrofluid technique requires no intrinsic magnetic properties of the biological assemblies and only moderate magnetic fields.

We have tested the ferrofluid technique on two viruses, the tobacco mosaic virus (TMV) and the tobacco rattle virus (TRV), and have obtained neutron-diffraction data that testify to its success. Our tests focused not only on obtaining a high degree of alignment but also on understanding how the alignment comes about. Figure 2 shows two possible mechanisms. One possibility is that the biological assemblies act like "magnetic holes." That is, by displacing the ferrofluid, they acquire effective magnetic moments equal in magnitude and opposite in sign to the sum of the moments of the magnetic particles in the displaced ferrofluid. The effective moments then align along the applied magnetic field. In our experiments, however, the effective moments would be large enough to cause alignment of the assemblies in moderate magnetic fields only if the assemblies existed as ordered domains, or aggregates. Then each domain becomes a magnetic hole oriented along the applied field (Fig. 2a). The other possible mechanism of alignment invokes linear correlations among the magnetic particles, that is, a tendency for them to line up in rows along the applied field. Because disrupting those linear correlations would require energy, the nonmagnetic elongated assemblies also tend to line up, with their long axes along the field, in rows between the rows of magnetic particles (Fig. 2b). It seems likely at this time that both alignment mechanisms operate to different degrees depending on the relative concentrations of magnetic particles and biological assemblies and whether the solution conditions favor the formation of ordered domains of the biological assemblies.
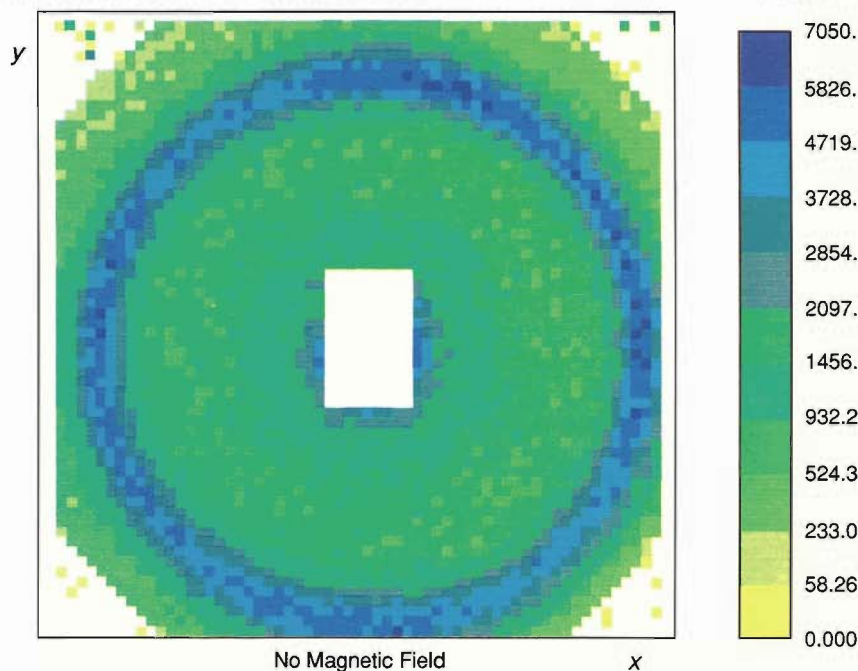
POSSIBLE MECHANISMS OF ALIGNMENT IN A FERROFLUID

Fig. 2. Elongated biological assemblies dispersed in a ferrofluid align along a moderate applied magnetic field even if they lack intrinsic magnetic properties. Shown here are two possible mechanisms of alignment. (a) Ordered domains of the biological assemblies act like magnetic holes. In other words, the domains acquire effective magnetic moments equal in magnitude and opposite in direction to the sum of the magnetic moments of the magnetic particles they displace. The effective magnetic moments cause the biological assemblies to align along the field. (b) Long-range magnetic correlations among the magnetic particles cause them to align in rows along the applied field. Lacking the energy to disrupt the linear correlations, the elongated biological assemblies also align along the field.

(a) Magnetic Holes

(b) Linear Correlations

Magnetic Field Line ~~~~~ Rod-Shaped Biological Assembly

⟵ Magnetic Particle ➡ Effective Magnetic Moment

The magnetic particles in the ferrofluid would normally contribute to the observed scattering intensity. However, that contribution can be eliminated by matching the scattering-length density of the magnetic particles to that of the solvent. By a happy coincidence of nature, the neutron scattering-length density of the ferromagnetic material magnetite is quite close to that of D_2O. As is well known, particles of magnetite align with an applied magnetic field. To achieve a
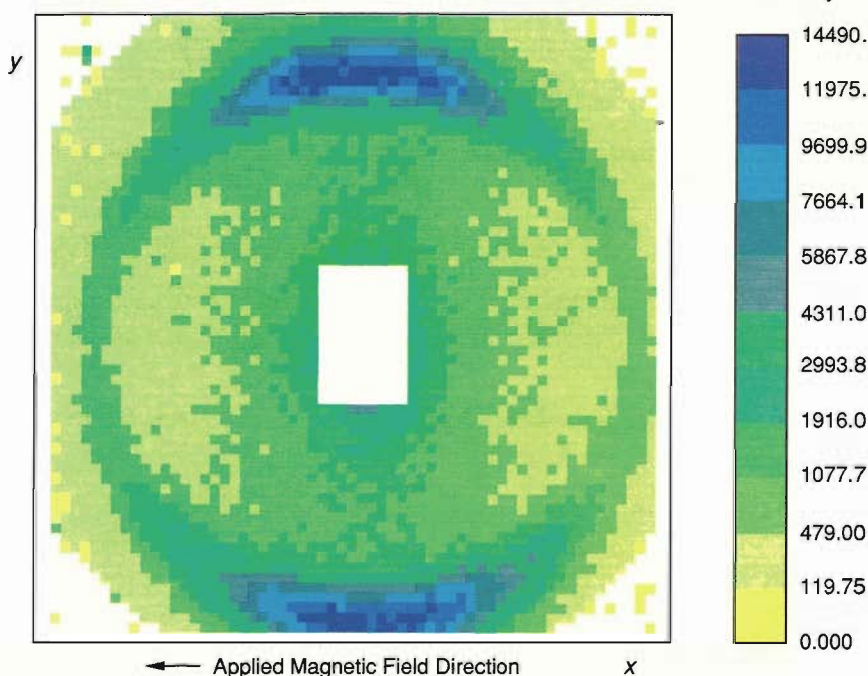
reasonably homogeneous suspension of magnetite particles in D_2O, the particles must be coated with a detergent surfactant. The detergent must be carefully chosen, however, to avoid its degrading the biological assemblies. Through extensive electron microscopy we identified a detergent that did not degrade TMV or TRV. We then deuterated the detergent to match the scattering-length densities of magnetite and D_2O.

To find out how well our ferrofluid

(a) Isotropic Diffraction Peak at 363 angstroms

No Magnetic Field

(b) Anisotropic Diffraction Peak at 363 angstroms

◄─── Applied Magnetic Field Direction   *x*

SMALL-ANGLE NEUTRON-SCATTERING DATA FOR TMV IN A FERROFLUID

Fig. 3. Color-coded intensity data for neutrons scattered from tobacco mosaic virus dispersed in a ferrofluid. The intensity corresponding to the color of a detector pixel is related to the number of neutrons detected in the pixel. (a) In the absence of a magnetic field, the scattering is isotropic and a diffraction peak (a circle of high intensity) appears at a scattering angle corresponding to a distance of 363 angstroms. The diffraction peak reflects the packing distance between individual viruses in ordered domains. (b) When a magnetic field of 0.3 tesla is applied along the *x* direction, the diffraction peak appears only in the *y* direction, that is, perpendicular to the applied field. That result means that the packing distances are aligned perpendicular to the field and the long axes of the viruses are aligned along the field. The degree of alignment cannot be explained by intrinsic diamagnetism of the viruses and must therefore be caused by the ferrofluid.
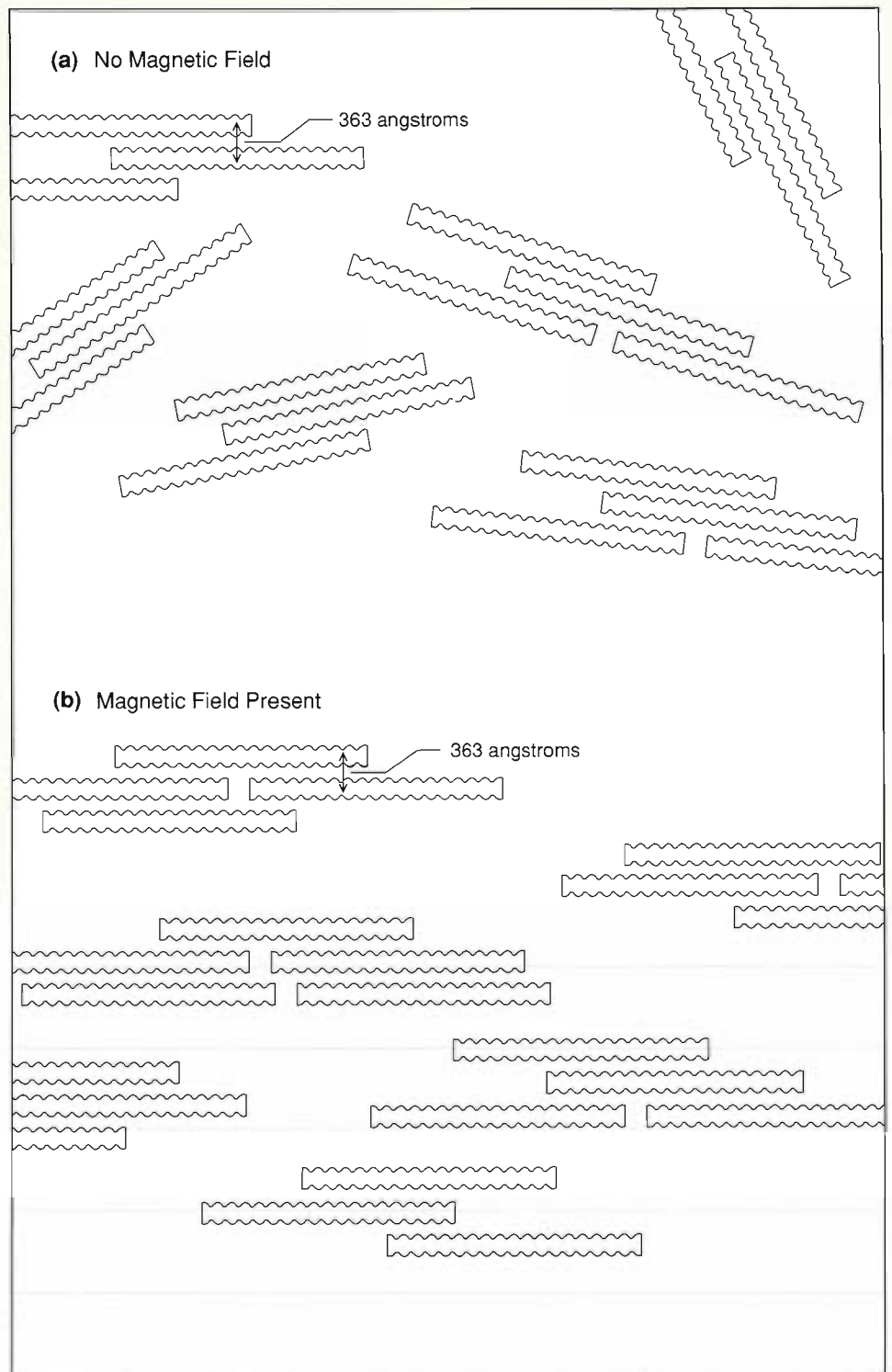
worked, we obtained neutron-scattering data for 0.06-volume-fraction dispersions of each virus in the ferrofluid. In the absence of a magnetic field, the scattering from TMV was isotropic, as expected (Fig. 3a). We also saw a diffraction peak corresponding to a distance of 363 angstroms. That diffraction peak arises because the viruses formed ordered domains in which the packing distance was 363 angstroms (Fig. 4a). (The packing distance is determined by a repulsive electrostatic force between the viruses and varies with the ionic strength of the solution.) In a modest magnetic field of 0.3 tesla, the scattering was anisotropic, and the diffraction peak at 363 angstroms appeared only in the direction perpendicular to the field (Fig. 3b). Those results suggest that the TMV domains were aligned in such a way that the long axes of the viruses were parallel to the applied field.

If the TMV domains do so align,

ORDERED AGGREGATES OF TMV IN A FERROFLUID

Fig. 4. Tobacco mosaic viruses tend to form ordered domains in solution, as shown in the electron micrograph below. The viruses within a domain are kept at a certain distance by electrostatic forces. That distance depends on the pH and ionic strength of the solution and in our experiments was 363 angstroms. (a) In the absense of a magnetic field, the domains are randomly oriented in a ferrofluid. (b) An applied magnetic field aligns the magnetic particles in the ferrofluid, which, in turn, causes the domains to align along the field. Thus the 363-angstrom spacing between the viruses is perpendicular to the field and produces diffraction peaks only in that direction.

(a) No Magnetic Field

363 angstroms

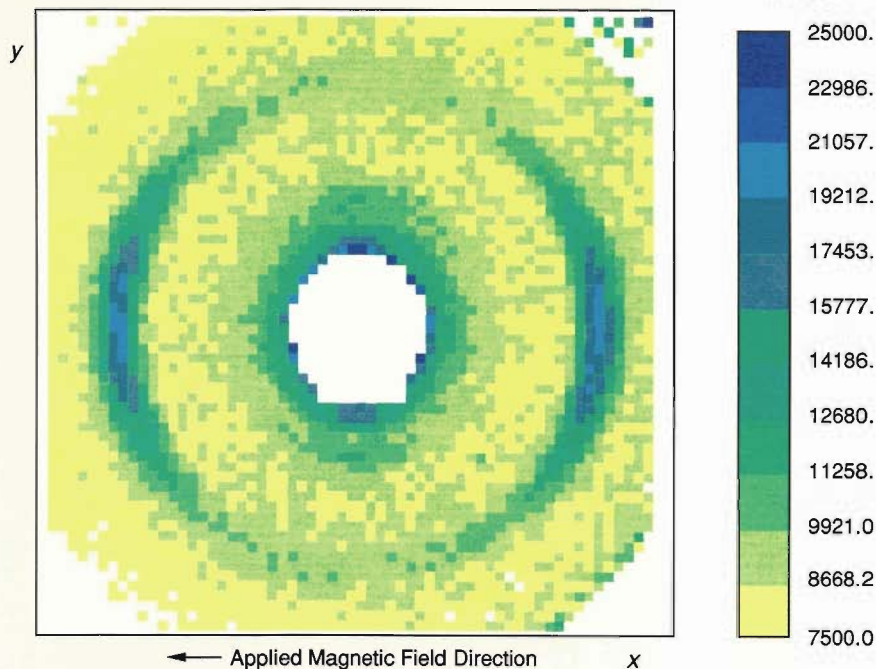**(b)** Magnetic Field Present

363 angstroms

another diffraction peak should be seen, one produced by the 23-angstrom distance between the turns of the helical viral coat protein (see Fig. 1). That peak should occur at scattering angles much greater than those included in the data of Fig. 3 and at a direction perpendicular to the direction at which the diffraction peak due to the packing distance is observed. We moved the detector closer to the sample and did indeed see a diffraction peak corresponding to a
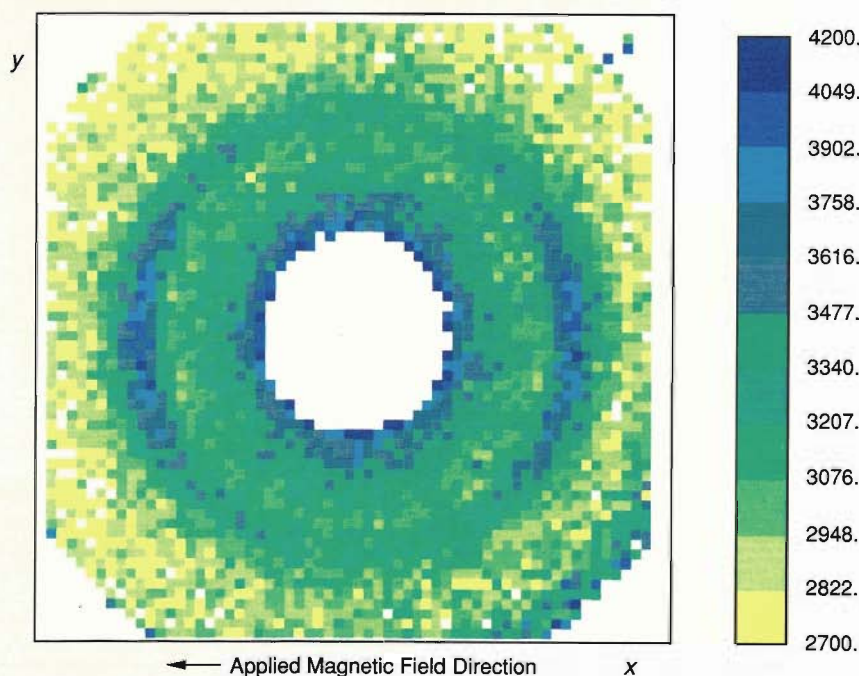
23-angstrom helical pitch (Fig. 5a). The existence of the peak confirms the structural integrity of the viruses in the ferrofluid, and its orientation parallel to the magnetic field is consistent with alignment of the long axes of the viruses parallel to the magnetic field (Fig. 4b).

Although our experiments demonstrated alignment, we wondered whether it was caused by intrinsic magnetic properties of TMV or by one of the mechanisms depicted in Fig. 2. The

(a) Diffraction Peak at 23 angstroms

Intensity

25000.
22986.
21057.
19212.
17453.
15777.
14186.
12680.
11258.
9921.0
8668.2
7500.0

◄— Applied Magnetic Field Direction   x

(b) Diffraction Peak at 25 angstroms

Intensity

4200.
4049.
3902.
3758.
3616.
3477.
3340.
3207.
3076.
2948.
2822.
2700.

◄— Applied Magnetic Field Direction   x

NEUTRON DIFFRACTION PEAKS FROM TMV AND TRV

Fig. 5. Color-coded intensity data for (a) tobacco mosaic virus and (b) tobacco rattle virus dispersed in a ferrofluid in the presence of a 0.3-tesla applied magnetic field. Note the diffraction peaks at 23 and 25 angstroms, which correspond to the regular spacing between the turns of the helical protein coats of TMV and TRV, respectively. The diffraction peaks appear only in the x direction, that is, along the direction of the applied field. Thus, the ferrofluid has aligned the viruses with their long axes along the applied magnetic field.

trated samples of TRV did not align in a normal solution even in a high applied magnetic field (7 teslas). However, when TRV was dispersed in the ferrofluid, a field of 0.5 tesla was sufficient to produce alignment comparable to that obtained for TMV in phosphate buffer. Figure 5b shows a diffraction peak corresponding to the 25-angstrom pitch of the helical TRV protein coat, similar to the peak obtained for TMV in Fig. 5a. The position of the peak indicates that the viruses are aligned parallel to the applied magnetic field. Moreover, like TMV in phosphate buffer, TRV showed no evidence of any domains.

Our preliminary studies demonstrated that the ferrofluid technique can orient elongated biological assemblies irrespective of their intrinsic magnetic properties and without disrupting their structural integrity. Furthermore, the orientation is sufficiently great to facilitate measurement of internal structural parameters. Future studies will concentrate on increasing the degree of alignment so that higher-resolution data can be obtained. ■

viruses could have been ordered by diamagnetic moments; however, the magnetic field strength was smaller than is generally needed to achieve the degree of alignment indicated by the data. Such alignment often requires the presence of ordered domains in the sample. So we repeated the experiments, this time adding phosphate buffer, the ions of which are known to disrupt domain structure. Thus alignment based on the intrinsic diamagnetism of the viruses is

not expected in the presence of phosphate buffer. We found that the viruses were still aligned but to a lesser degree. That result suggests not only that the ordering is due to the presence of the ferrofluid but also that the mechanism depicted in Fig. 2b may enhance the alignment of TMV.

The tobacco rattle virus is genetically unrelated but morphologically quite similar to TMV. TRV shows only poor orientation under shear, and concen-

NEUTRONS, SLUDGE PHYSICS *and the* LIBERTY BELL *by Roger Pynn*
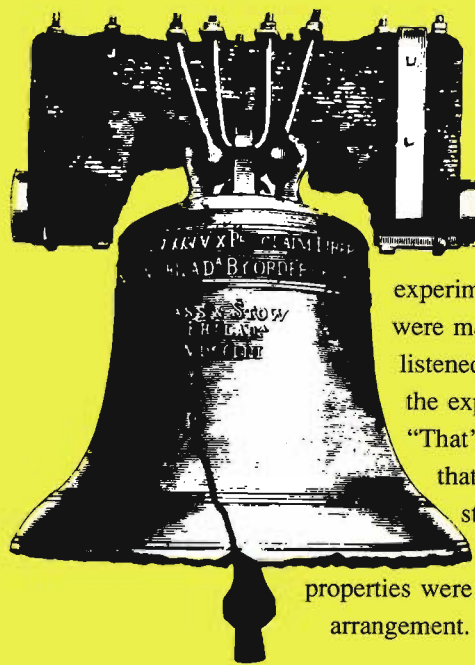
A friend of mine, a well-respected theoretical physicist at a major American university, worked for many years on polymers and colloids at one of the better-known oil companies. He told me that he once discussed his work with an equally well-respected experimentalist whose main preoccupations were magnetism and phase transitions. Having listened to my theorist friend for a few minutes, the experimentalist rendered his verdict: "That's sludge science!" He meant, of course, that the polymers and colloids my friend studied did not have a convenient monocrystalline form—their physical properties were not simplified by a symmetrical atomic arrangement. Traditionally, scientists have preferred

Why would one want to study sludge with neutrons? Our opener shows both high-tech and mundane products made of disordered materials washing in on a blue wave of sludge. In the background, fumed silica particles (red) represent the fractal aggregates elucidated by neutron scattering. The Liberty Bell, a famous example of the effects of residual stress, rounds out this gallery of neutron scattering's practical applications.

to investigate the physical properties of ordered systems, or at least of materials that are free of impurities and dirt. Unfortunately, many of the materials on which our society depends are "dirty" and quite devoid of the translational symmetry of a crystal—they are made of "sludge."

Examples of sludge abound: the oil we put in our automobile engines with all of the detergent additives that advertisements insist are so important; the various polymers we use to package food, weave cloth, make compact disks, and produce a variety of parts for dishwashers, televisions, and other appliances; the paints and cosmetics we use to color and preserve our houses and our skins; detergent; soap; toothpaste; even some things that pass for food. In today's age of advancing technology, we are constantly seeking new materials that can better serve our needs, that can carry out a new or an old function better than existing materials.

But why would one want to study sludge with neutrons? To design new materials we have to understand the relationships between material properties, material structure, material synthesis, and material performance (Fig. 1). As an example, we may decide that the steel armor of yore is impractical for the modern policeman. Instead, we need a flack-jacket that is strong enough to stop bullets, flexible enough to let the policeman move his arms, and light enough for him to go to battle without a horse. Our dream material needs to combine all these properties. Polymers—a common type of sludge—are light and flexible, but those we encounter daily as zip-lock bags or the outer casings of televisions are not strong enough. To design a polymer with the strength to stop bullets requires an understanding of the relationship between molecular structure and strength. Once we know the structure that provides the strength we want, we have to invent a way of

synthesizing a polymer with that structure. Neutrons enter the picture because they often provide an ideal method of determining the structures of materials. We may not think of sludge as having structure, and none may be visible to the naked eye, but the molecules are always organized in some way, and that organization often determines properties.

Structure, properties, synthesis: all investigations build to the performance apex of the materials research pyramid. Once we make a product from our dream material, we must find a way to ensure that it will perform adequately. For a unique, expensive product we must make this guarantee without simulating future use—which brings us to the Liberty Bell. Some years after its manufacture, the Liberty Bell cracked, probably as a result of stresses introduced when it was cast. Without realizing that their bell's fame would hang from its failure, the manufacturers might have prevented the rupture if they had understood the distribution of stresses related to the casting process. However, to achieve such an understanding without destroying the Bell or any other stressed piece of metal—a railway line, an oil pipeline, or a turbine blade—is difficult. We have to measure the strain in the material, that is, the amount by which it has been stretched or compressed during manufacture or use, without disturbing an atom. Unfortunately, most of the traditional methods employed to obtain information about residual strains involve destruction or modification of the component—hole drilling in conjunction with strain-gauge measurements and progressive removal of layers of material, for example. Today, neutron diffraction provides an alternative, nondestructive method of measuring changes in the interatomic spacings, and thus of measuring the residual strains, at different positions in the sample.

In this article I will describe a few

of neutron scattering's contributions to our understanding of the structures of several different types of sludge and to studies of the stresses in metals. Experiments of this sort are relatively new to neutron scattering. The profound contributions that neutrons have made to polymer science—such as determining the conformations of polymer molecules—started at the beginning of the seventies, and neutron measurements of residual strains in engineered components began only in the eighties. To a small extent this evolution has depended on the development of instrumentation for neutron scattering. To a far larger extent it results from a maturing of the field. In the early days, neutron-scattering researchers measured single crystals and tried to understand the properties described in graduate texts on physics: atomic and magnetic structures of inorganic solids, phonons and magnons in crystals, and the structures of simple liquids, for example. Later, they used neutrons to study the structural aspects of much more esoteric phenomena—structural phase transitions, manifestations of nonlinearity such as solitons, and the coexistence of magnetism and superconductivity. Each new phenomenon left its signature on neutron-scattering experiments by giving rise to a qualitatively different neutron scattering law. Each was a stepping stone to the study of a more complex system. Now the technique is mature enough to apply to sludge, to materials we really care about.

Of course, as a practicing scientist I strongly believe basic research should be pursued, even when its relevance to society is not instantly apparent. From lignite to lasers, most of the discoveries that have profoundly affected our lives have resulted from simple human inquisitiveness, not from a desire to make a better widget. It is important that neutron scattering continue to be used for basic research—witness the other ar-

(a) Essential Areas of Study
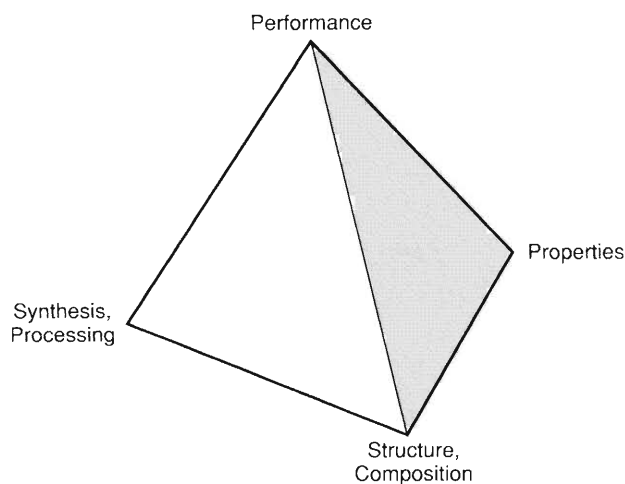
(b) An Area of Progress
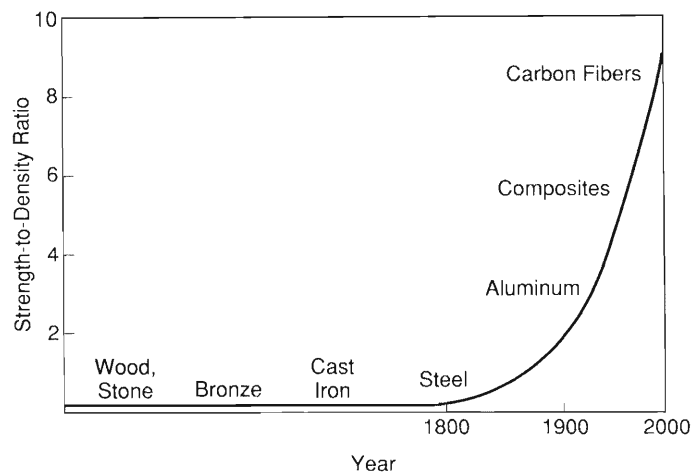
MATERIALS RESEARCH

Fig. 1. (a) This tetrahedron represents the four essential areas of materials research. Its symmetry implies that investigations of any apex must necessarily build on the other three. For example, the sharp improvement in the strength-to-density ratio of construction materials—shown on a relative scale in (b)—would not have been possible without concurrent advances in methods for synthesizing new materials, elucidating their structures, clarifying their properties, and testing their performance. Of course, the complicated demands of high-technology applications require attention to more than just strength and weight. As a simple example, lightweight composite materials that get stronger when heated are being developed for the wings and engine of hypersonic jets. (The graph is reprinted from *Materials Science and Engineering in the 1990s*, ©1989 by the National Academy of Sciences, National Academy Press, Washington, D.C.)

ticles in this issue. However, when a technique can make significant contributions to the solution of more immediate and practical problems, those problems should not be ignored. They are certainly no less challenging intellectually, and their solutions have immediate, obvious impact. Such studies contribute to incremental improvements in our technologies, while basic research often drives the revolutionary changes that become clear in hindsight.

Colloids and Amphiphilic Molecules

The first type of sludge I will discuss is colloidal sludge. A colloid is an assembly of small particles dispersed in a continuous medium of gas, liquid, or solid. Fog, smoke, milk, paint, and foam are colloidal systems encoun-

tered routinely. Because the particles are small—between 10 and 10,000 angstroms (1 angstrom = 10^{-10} meter)—they do not settle even when dispersed in gases or liquids. Rather they are kept in suspension by the kicks they receive from molecules of the carrier fluid. Many macromolecules such as polymers or proteins dissolve easily in a solvent, and in this case the colloidal particles are the molecules themselves. More fascinating and technically relevant are the so-called association colloids, one class of which consists of aggregates of relatively small *amphiphilic* molecules. Amphiphilic molecules have a polar, and hence water-soluble, head at one end and a water-insoluble tail at the other; the head is *hydrophilic*—water-loving—whereas the tail is *hydrophobic*—water-hating. In water such molecules aggregate so that their tails

are dry, assembling spontaneously into particles called *micelles* (Fig. 2). In addition, amphiphilic molecules solubilize materials that are normally completely insoluble in water—metals, ceramics, or oils, for example. The amphiphilic molecules are adsorbed onto the surfaces of "dirt" particles, and these surface layers prevent the dirt from separating out. Like the micelle, this configuration prevents contact between the tails of the amphiphilic molecules and the water environment. When small groups of amphiphilic molecules solubilize a liquid, such as oil, the resulting colloid is called a *microemulsion*.

Anyone who has tried to wash clothes or dishes in plain water knows why amphiphilic molecules are important: without soap or detergent—amphiphilic molecules by another name—dust and grease will not wash away. In addition

(a) Micelle   Water   Polar Head   **(b)** Inverted Micelle   Oil

Hydrocarbon Tail

50 Å   Amphiphilic Molecule   30 Å

(c) Oil-in-Water Microemulsion   Water

Oil

100 Å

(d) "Solubilized" Dirt Particle   Water

Dirt

100 Å

to everyday dirt, amphiphilic molecules are used to solubilize a wide variety of other materials. Detergents isolate corrosive products in motor oil by forming microemulsions; light-sensitive emulsions register images in photographic processes; bilayers of amphiphilic molecules are used to encapsulate drugs for more efficient delivery in the body; microemulsions are used for tertiary oil recovery by the petroleum industry; and fats are transported in our blood stream by micelles formed from serum lipoproteins—the high-density lipids we hear about in connection with our cholesterol measurements.

In describing the association colloids that can be made with amphiphilic molecules, I have given in Fig. 2 the conventional textbook cartoons that purport to describe the structures of various colloidal particles. However, the cartoons gloss over many important questions. For instance, are the particles really spherical? How big are they? Are all particles in a colloidal suspension the same size? How much of the particle volume is occupied by the head groups? How much by the tails? Is the tail region uniformly dense? What is the effective surface charge? How do the particles organize themselves in concentrated suspensions? And how do all of these properties depend on temperature and the concentration of amphiphilic molecules? Many of these questions can be answered by small-angle neutron-scattering (SANS) experiments, especially when the technique of contrast matching is used.

Small-angle neutron-scattering experiments are useful for measuring lengths between 10 and 10,000 angstroms, the size range in which many colloidal particles fall. As the name implies, researchers measure the intensity of neutrons scattered at very small angles, that is at very small values of the wavevector transfer, Q. For values of Q that are less than about R^{-1} (that is, for
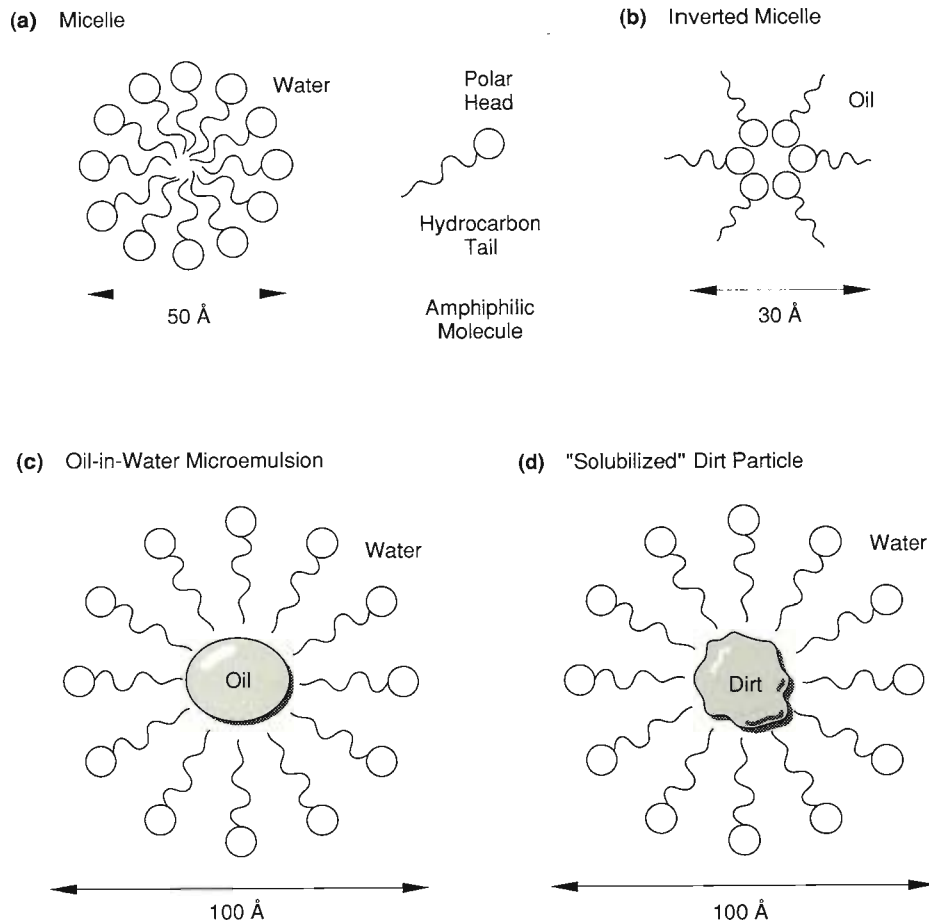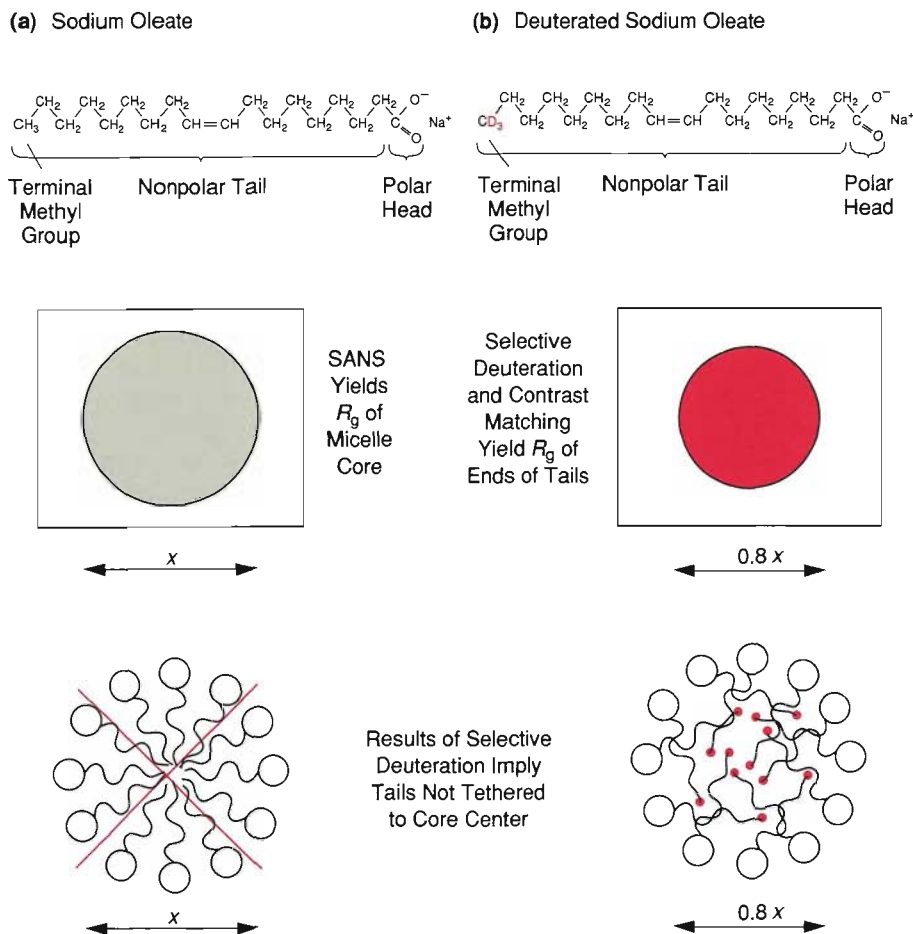
TYPES OF ASSOCIATION COLLOIDS

Fig. 2. When a sufficient number of amphiphilic molecules are dispersed in a solvent, the interaction between their hydrophilic heads and hydrophobic tails and the solvent molecules can lead to formation of various types of colloidal particles. (a) In water amphiphilic molecules can form micelles in which the hydrophobic tails group together so that only the polar heads contact the water. (b) In oil the opposite happens; the polar heads group together so that only the tails contact the oil. (c) Amphiphilic molecules can form a microemulsion of oil in water by surrounding oil droplets with their tails. (d) Detergent molecules can make dirt particles soluble in water by surrounding them. Although the effects of the hydrophobic and hydrophilic interactions suggest the general configuration of the colloidal particles, techniques such as x-ray and neutron small-angle scattering are needed to study details of size, shape, and structure.

values of QR less than about 1), where R is the approximate size of a particle, the appropriate approximation for the scattering from a dilute suspension of particles is the Guinier approximation, which in logarithmic form is

$$\ln I(Q) = -\frac{1}{3}R_g^2 Q^2 + \ln I(0). \quad (1)$$

$I(Q)$ is the scattering law, that is the intensity of scattered neutrons as a function of Q, and R_g is the so-called radius of gyration of the particle. R_g is analogous to the radius of gyration of classical mechanics (the square of which

equals a body's moment of inertia divided by its mass) and thus is a good measure of particle size. According to the Guinier approximation, we can determine R_g by plotting the logarithm of the scattered-neutron intensity versus $\frac{1}{3}Q^2$. At small values of Q, the plot should be a straight line with a slope of $-R_g^2$. In the case of a solid spherical particle whose size is determined only by one dimension—its radius—that dimension can be obtained directly from R_g; $r = \frac{1}{3}\sqrt{15}R_g$, or about $1.29R_g$. However, more dimensions are needed to define anisotropic shapes such as ellipsoids, cylinders, or pyramids. In these

(a) Sodium Oleate

(b) Deuterated Sodium Oleate

Terminal Methyl Group | Nonpolar Tail | Polar Head

SANS Yields R_g of Micelle Core

Selective Deuteration and Contrast Matching Yield R_g of Ends of Tails

Results of Selective Deuteration Imply Tails Not Tethered to Core Center

DETAILS OF MICELLAR STRUCTURE

Fig. 3. The difference between the neutron scattering lengths of hydrogen and deuterium allows researchers to see the internal structure of micelles in some detail. (a) The hydrocarbon tail of sodium oleate—a common soap—has an almost homogeneous scattering-length density along its entire span. At best, neutron scattering can show that a micelle consists of an outer shell of head groups and a core of tails with a radius of gyration x. (b) By substituting deuterium for hydrogen in the terminal methyl groups of the sodium oleate and adjusting the scattering-length density of the solvent, researchers can distinguish the ends of the tails from the tails as a whole. The 20-percent smaller radius of gyration obtained for the deuterated methyl groups (red) leads researchers to postulate a less ordered structure for the micellar core. The terminal methyl groups tend toward the center of the micelle but are not heavily concentrated there.

cases, R_g gives a good approximation of the average particle dimension.

In addition to particle size, a Guinier plot can be used to determine $I(0)$, the extrapolated value of the scattering law at $Q = 0$. This value is proportional to the average mass of the colloidal particles, that is, to the average number of amphiphilic molecules contained in a particle. Thus, simple Guinier analysis of small-angle neutron-scattering data from dilute suspensions provides immediate quantitative information on the average dimension and mass of the colloidal particles.

At larger values of Q, for exam-

ple in the *Porod region* where QR is greater than 5, the intensity of scattered neutrons yields information about the surface structures of colloidal particles. In particular, Porod's law says the scattered-neutron intensity will be proportional to Q^{-4} if the surfaces of the colloidal particles are smooth. Moreover, the constant of proportionality measures the surface area of the colloidal particles per unit volume of suspension. If the scattering law in the Porod region obeys a power law with an exponent between -3 and -4, the colloidal particles have rough surfaces, which need to be described by the frac-

tal concept described later in this article.

In addition to the size and surface area of the colloidal particles, small-angle neutron scattering also allows us to determine some details of their internal structures. For example, Fig. 2 indicates that all the hydrocarbon tails in the micellar core have the same zig-zag conformation and all point toward the center of the micelle. Is this really the case? Probably not, because such an arrangement would allow water to penetrate the micelles and come in contact with the hydrophobic tails. On the other hand, one would not expect the micellar cores to have a uniform, liquid-like density either. After all, the arrangement of the tails is constrained by the distribution of the polar heads on the particle surface. Probably the real situation lies between the two limits. The neutron-scattering technique known as *contrast matching* can help to resolve this issue.

Contrast matching is based on a simple principle: components of particles that have the same density of scattering power, or *scattering-length density*, as the solvent in which the particles are dispersed do not contribute to the observed scattered-neutron intensity. (See "Neutron Scattering—A Primer" and "Biology on the Scale of Neglected Dimensions" for a more detailed account of contrast matching.) If we could somehow match the scattering-length density of the solvent to all regions of the hydrophobic tails except the methyl groups at their ends (see Fig. 3), then the observed scattering would be due to the methyl groups alone and would provide information about their spatial distribution. Achieving that situation is possible because of a fortunate quirk of nature—the relatively large difference between the scattering lengths of hydrogen and deuterium. Researchers substitute deuterium for hydrogen in the terminal methyl groups and disperse the micelles in a mixture of D_2O and H_2O whose scattering-length density closely

matches that of the tails. By this process, researchers can "see" the spatial distribution of the deuterated methyl groups alone. Typically, the radius of gyration found for the deuterated methyl groups in a micelle is less than that of the whole micellar core, indicating that the ends of the molecular tails are more likely to be found toward the center of the micelle than evenly distributed throughout the core. On the other hand the difference between the two values of R_g is typically less than 20 percent. So the ends of the chains are certainly not tethered to the center of the micelle as Fig. 2 indicates.

So far, I have talked only about experiments on generic colloidal systems. To develop colloidal systems for specific applications, we must improve our understanding of the relation between a sample's structure and its properties. As a simple example: it is easy to understand why soap works in terms of Fig. 2—dust or oil are imprisoned by amphiphilic molecules and washed away in a milky dispersion. But suppose, by chance, that even slight dilution of the system caused the soap molecules to release their captives. This structural change would mean that the very action of rinsing off the soap would redeposit all of the dirt! To understand this phenomenon—or develop a better soap—we would need to examine the stability of molecular arrangements as a function of molecular concentration. Neutron scattering can tell us the arrangements amphiphilic molecules assume, but the *free energy* of these arrangements determines their stability.

The free energy of a molecular arrangement is calculated by using the equation

$$F = H - TS. \qquad (2)$$

That is, the free energy, F, equals the sum of the particle-interaction energies, H, minus the temperature, T, times the entropy, S. The arrangement with the lowest free energy is the most stable. Entropy is a measure of disorder and is largest for a random distribution of amphiphilic molecules in solution and smaller for an ordered arrangement such as a micelle. Because T and S are always positive, free energy decreases as entropy and temperature increase. Thus, entropy considerations tend to favor random distributions of amphiphilic molecules over ordered systems such as micelles, especially at high temperatures. Also opposing micelle formation are the electrostatic forces that come into play when the polar heads of the amphiphilic molecules are constrained to lie on the surface of a micelle; the heads experience repulsive forces which increase the interaction energy (and thus the free energy).

On the other hand, what amounts to an attraction between the hydrophobic tails of the molecules counters the "bumping heads" effect and the effects of entropy and thus favors the formation of micelles. The attraction is thought to be an indirect effect of the interaction between the tails of the amphiphilic molecules and the molecules of the solvent. When amphiphilic molecules are randomly distributed in water, their hydrocarbon tails disrupt many of the water's hydrogen bonds, thereby increasing the solvent's free energy. The water molecules like to restore these bonds and reduce their free energy; so they tend to push the tails out of the way, causing them to group together and form micelles. For many amphiphilic molecules, the effective attraction between the hydrophobic tails is about equal to the repulsion between their polar heads if the concentration of amphiphilic molecules is high enough. The free energy is then lowest when the molecules self-assemble into micelles. However, at high enough temperatures entropy always wins and the molecules are randomly distributed in the solvent.

Vesicles: A Novel Drug Delivery System. Aggregations of amphiphilic molecules can assume more complicated shapes, and the hydrophobic interaction can have a wider variety of effects than simple micelles. For example, commercial applications of the hydrophobic effect include nonwetting, nonstick pans and waterproof fabrics, and the semipermeable membranes of biological cells in our bodies are formed from bilayers of amphiphilic lipid molecules. In addition, one of the most technologically exciting actions of amphiphilic molecules is the formation of hollow, spherical shells called *vesicles* (see Fig. 4). Vesicles are composed of bilayers of amphiphilic molecules that are very similar in form to the membranes of biological cells and sometimes contain the same lipids. Such vesicles—called *liposomes*—occur naturally in the body and serve a variety of purposes, including the transport of fats during digestion.

The idea of using liposomes for drug delivery was born in the late 1960s as a possible solution to one of medicine's fundamental compromises. Because all drugs are poisonous at some level, the least possible amount should be administered. On the other hand, drugs are diluted in the blood stream, and large amounts are degraded, excreted, or adsorbed onto healthy cells; so the doses must be large enough to overcome this wastage. In some cases a compromise between these contradictory requirements is almost impossible—if a drug is administered in sufficient quantity to effect a cure, a side effect of the drug may kill the patient. However, drugs encapsulated in vesicles have been found to circulate in the blood for considerable periods without being degraded or adsorbed. Such liposomal drugs have provided effective therapy for several prevalent diseases—invariably with far smaller doses than usual. Most vesicles containing a drug

(a) Two-Component Micelle

Rod-like
Micelle

Lecithin
Bilayer

Hydroxyl
Group

Glycocholate
Molecule

(b) Vesicle

Lecithin
Bilayer

Glycocholate
Dimers

are recognized as foreign and eventually ingested by cells of the immune system that circulate in the blood stream. For this reason drugs encapsulated in vesicles—one may think of them as medical Trojan Horses—have proven effective in treating infections of the immune system itself. The vesicles may also be incorporated into body tissue in other ways; for example, the lipid bilayer composing the vesicle may become fused with a cell membrane. Such fusion allows liposomal drugs to be adsorbed by cells that are not part of the immune system and to provide effective therapies for diseases such as fungal infections. Finally, weakened pathogens concentrated on the surfaces of vesicles can induce the immune system to produce antibodies, making liposomal vaccines promising candidates for disease prevention.

We don't always understand why liposomal drugs work, and probably no single hypothesis will explain all successes. Nevertheless, an increased understanding may well come from a knowledge of the morphological changes that occur both when vesicles are diluted in the blood stream and when they come into contact with living matter. Clearly if the "wrong" phase change takes place when a liposomal drug is diluted in the blood—for example if the vesicles transform into open sheets and release their drug prematurely—encapsulation will have been in vain. For this reason neutron-scattering studies of vesicle structures are an important element in the development of liposomal drug-delivery systems.

Here at the Los Alamos Neutron Scattering Center, Rex Hjelm has made extensive studies of spontaneous vesicle formation in a mixture of the bile salt glycocholate and lecithin suspended in water. Such spontaneous liposome formation could greatly reduce the costs of drug-delivery systems. The morphologies of particles formed in the
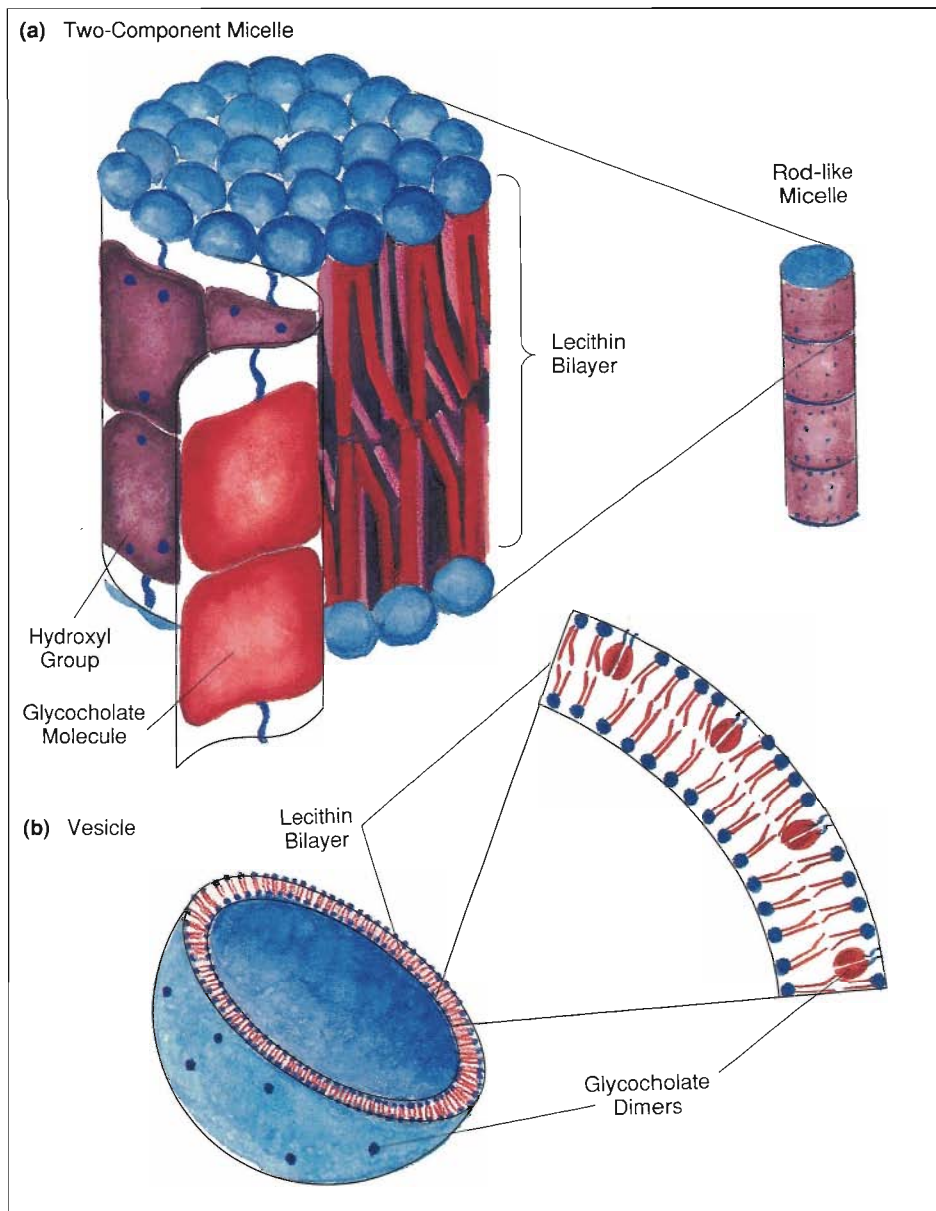
PARTICLES INVOLVED IN FAT DIGESTION

Fig. 4. Depending on their concentration in aqueous solution, molecules of the lipid lecithin and the bile salt glycocholate form a variety of structures. (a) At high concentrations and high molar ratios of glycocholate to lecithin, globular particles form. The lecithin molecules form bilayers with their hydrophobic tails (red) pointed inwards and their hydrophilic heads (blue) exposed to the aqueous environment. It is thought that the glycocholate molecules stabilize the bilayers by forming a "ribbon" two molecules wide that wraps around the particle, further isolating the lecithin tails from the water. The hydrophilic, hydroxyl-containing surfaces of the glycocholate molecules form the outside of the ribbon. Neutron-scattering data show that the particles are 50 angstroms high and 50 angstroms in diameter and suggest that they assemble into long rods as the system is diluted. Particles of this kind are thought to occur naturally in the liver and the gall bladder for the purpose of removing lipophilic products—cholesterol, for example—from the liver. (b) When the concentration of glycocholate drops below a critical level as a result of dilution, vesicles form. Less glycocholate is available to associate with the lecithin, and the two-component micelles in (a) are no longer stable. Instead it is thought that the glycocholate molecules form dimers that act as wedges within the lecithin bilayer, causing it to curve and form a hollow sphere, or vesicle. The precise details of this dramatic rearrangement are not yet known. Such vesicles are thought to occur naturally in the digestive tract, emulsifying dietary fats and thus enhancing the effects of digestive enzymes.

glycocholate-lecithin system depend on a number of factors, including concentration; thus a knowledge of how the particles evolve from micelles at high concentrations to vesicles at low concentrations is essential to the development of liposomal drugs. The glycocholate-lecithin system is all the more interesting as it is an essential component in the digestion of fat from mammalian diets. Although glycocholate and lecithin are both amphiphilic, the glycocholate is far more soluble, and the combination of the two seems more efficient for solubilizing fats in the digestive tract than either alone. The emulsification of fats by the glycocholate-lecithin combination appears also to enhance the action of digestive enzymes.

A peculiarity of the glycocholate-lecithin combination is that the particles it forms tend to grow larger as the system is diluted in water, down to a low concentration at which they transform to vesicles. Such behavior is opposite to the usual effect of dilution. As Hjelm and his collaborators added water to their glycocholate-lecithin system they observed that the form of the neutron scattering law had changed in a manner qualitatively consistent with a change from globular micelles to rod-like structures. However, attempts to fit the data to the scattering law of a population of uniform-sized cylindrical micelles were not very successful; evidently, cylinders of different sizes were present in the suspension. To find out the size distributions of the particles, Hjelm and Devinderjit Sivia used a method of analysis known as the maximum entropy method. They assumed the system contained cylindrical micelles whose radii and heights could take any value within a reasonable range and calculated the neutron scattering law for each combination of radius and height. Then they applied the maximum entropy method to the observed

scattering law to determine the best estimate of the number of micelles with each height and radius. (Their results are plotted as equi-population contours in Fig. 11 of "Bayesian Inductive Inference, Maximum Entropy, and Neutron Scattering.") At a concentration of 16.7 grams per liter, most of the particles had globular structures with heights centered at 50 angstroms and radii centered at 25 angstroms. At a concentration of 10 grams per liter, a new population of elongated particles emerged in addition to the first group. Remarkably, the second population had about the same mean radius as the first but twice the mean height. Finally, on dilution to 7.1 grams per liter, a third population with a height three times greater than the first began to emerge. Since a lecithin bilayer is about 50 angstroms thick, these data strongly indicate that the elongated particles are built up of disks composed of lecithin bilayers. At present, Hjelm and his collaborators believe the glycocholate molecules wrap like a ribbon around the circumference of the lecithin disks (Fig. 4).

Polymer Colloids. Association colloids such as micelles and vesicles are not the only type of colloids in our arsenal of modern materials. In another type of colloid called a *latex*, the core of each particle consists of long polymer molecules. Latices were first produced by a process called emulsion polymerization as part of the development of synthetic rubber during World War II. In emulsion polymerization, droplets of monomers of an insoluble synthetic resin are kept in colloidal suspension by the addition of a suitable surface-active agent, or *surfactant*, composed of amphiphilic molecules. To this emulsion is added a small quantity of free-radical polymerization initiator, which diffuses through the surfactant shell and causes the monomers within to polymerize. The process produces a colloidal disper-

sion of polymeric, synthetic-resin microspheres, each surrounded by a surfactant layer that prevents their aggregation. At moderate particle concentrations, this dispersion has the appearance of a milky, white liquid.

Latex particles of different sizes and size distributions can be produced for a variety of industrial applications. For example, the particles may be dried to form a continuous film of latex paint, paper coating, carpet backing, or adhesive. They may also be used as pigments, as length calibration standards for electron microscopy, and as carriers for antigens in diagnostic immunological tests. To improve the latices—to design a better paint, for example—it is essential to have a detailed understanding of the process of emulsion polymerization and an accurate characterization of the structural and chemical properties of the latices that are produced. Without such knowledge design of new materials is a hit-and-miss affair.

Different characteristics of latices are important for different technologies, and no single investigative technique can provide a complete description of all particle properties. However, many of the properties of importance are structural. Examples include the mean particle size, the morphology of particles with more than one component ("avocado" or "current bun," for example), the type of force between particles, the thickness of the amphiphilic stabilizing layer that surrounds each particle, and the texture of the particle surface (rough or smooth). Once again, much of this information can be obtained by small-angle neutron scattering. Over the last decade, university and industrial researchers have performed a number of small-angle neutron-scattering experiments to determine the distribution of particle sizes in dilute latices containing one type of polymer. This information may be obtained either by fitting the scattered-neutron intensity to an as-

sumed distribution of particle diameters or by using the maximum entropy method to determine the most probable distribution. In either case, neutron scattering provides a simple alternative to the laborious counting of latex particles in electron micrographs. Results from these two techniques usually agree well. Contrast matching has also been used successfully to determine both the composition and structure of copolymer particles, that is, of particles formed from two different types of polymer molecules. Such information cannot be obtained by any other means.

To determine the composition of the particles of a copolymer latex, researchers collect SANS data for suspensions of the latex in various mixtures of H_2O and D_2O. Because $I(0)$, determined by extrapolation from the Guinier plot described earlier, is proportional to $(\bar{\rho}_p - \bar{\rho}_s)^2$—the square of the difference between the average scattering-length density of the particles and the average scattering-length density of the solvent—a graph of $\sqrt{I(0)}$ versus $\bar{\rho}_s$ yields a straight line (Fig. 5). The intersection of that line with the $\bar{\rho}_s$ axis (the point at which $I(0) = 0$, or the contrast-matching point) is equal to $\bar{\rho}_p$. For example, the data in Fig. 5 lead to a value for $\bar{\rho}_p$ of 4.94×10^{10} cm^{-2}. The composition of a copolymer latex may then be obtained directly by solving the following equation for c:

$$\bar{\rho}_p = c\rho_A + (1 - c)\rho_B, \qquad (3)$$

where c is the fraction of polymer A in each particle and $(1 - c)$ is the fraction of polymer B. The quantities ρ_A and ρ_B are the scattering-length densities of polymer A and polymer B, respectively. They can be deduced from the scattering lengths of the atoms in the molecules and their known molecular volumes. Alternatively, experimental values for ρ_A and ρ_B can be obtained by measuring the contrast-matching points
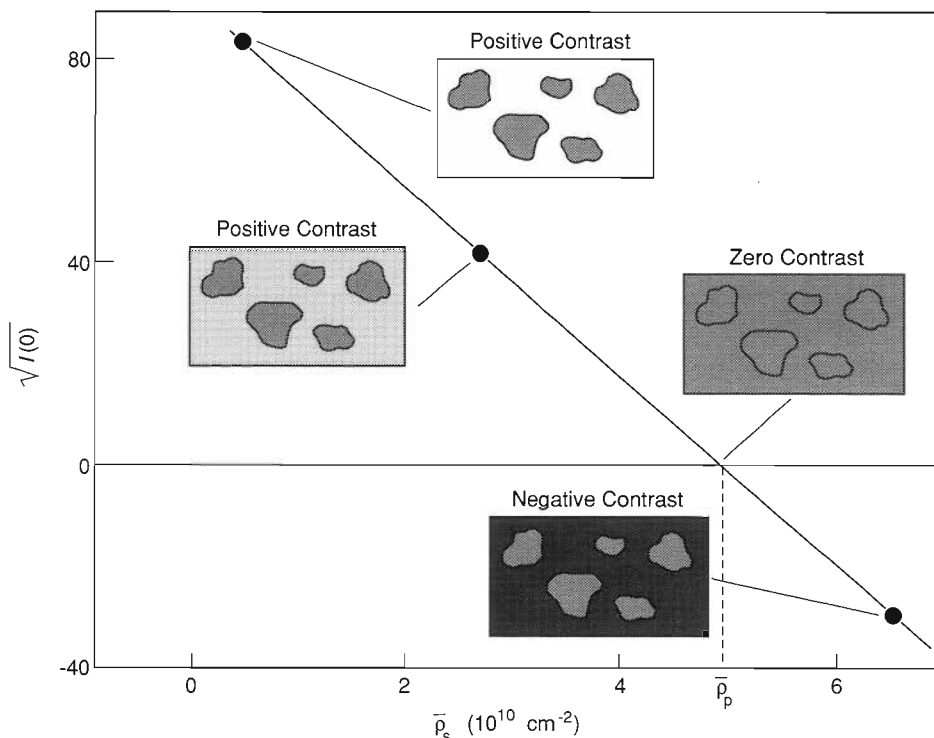
DETERMINING THE COMPOSITION OF A COPOLYMER LATEX

Fig. 5. Neutron-scattering data for copolymer-latex particles dispersed in water can reveal the relative amounts of the two polymers (here polystyrene and polyacrylonitrile) in the particles. The intensity of the neutrons scattered through zero degrees, $I(0)$, is proportional to the square of the scattering contrast, $(\bar{\rho}_p - \bar{\rho}_s)^2$, where $\bar{\rho}_p$ is the average particle scattering-length density and $\bar{\rho}_s$ is the average solvent scattering-length density. Thus, a graph of $\sqrt{I(0)}$ versus ρ_s should be a straight line, and the intersection of that line with the $\bar{\rho}_s$ axis (the point at which $\sqrt{I(0)} = 0$ and hence $\bar{\rho}_p = \bar{\rho}_s$) equals $\bar{\rho}_p$ for the latex particles. The three data points shown were obtained by varying the percentage of D_2O in the water; for two of the data points the contrast is positive $(\bar{\rho}_p - \bar{\rho}_s > 0)$, and for the third the contrast is negative $(\bar{\rho}_p - \bar{\rho}_s < 0)$. We can then use Eq. 3 in the text to determine the relative amounts of polystyrene and polyacrylonitrile in each particle. The data shown correspond to latex particles containing 63.6 percent polystyrene and 36.4 percent polyacrylonitrile. (Data courtesy of R. H. Ottewill of the University of Bristol.)

of latices prepared from the individual polymers.

In many cases it is possible to adsorb molecules—such as paint pigments or drugs—onto the surfaces of latex particles. To determine how many molecules are adsorbed under different chemical conditions, SANS measurements are first performed on bare particles in a series of solvents of different contrasts

in order to determine the scattering-length densities of the particles. Then the molecules of interest are adsorbed onto the latex particles, and the scattering law for the composite particles is measured in a solvent whose scattering-length density equals that of the latex particles alone. The observed scattering law is due to the shell of adsorbed molecules only. By fitting the observed

scattering law to that expected for a spherical shell, the total amount of adsorbed chemical may be determined. If the molecules are absorbed into the latex rather than *adsorbed* on the surface, the particles usually swell. Absorption can be discriminated from adsorption because a contrast-matching point can be found for uniform swollen particles but not for particles with a core-and-shell morphology.

Polymers

Of course, polymer molecules do not occur only as constituents of latex particles dispersed in a solvent. Most often in our everyday experience, an assembly of polymers is a solid lump—a telephone housing or a fishing pole—rather than a constituent of a milky liquid. Depending on the chemical composition of the polymer molecules and the method of preparation, a bewildering variety of mechanical, electrical, and optical properties can be achieved. In many cases the properties depend sensitively on the organization of the polymer molecules, that is, on the local molecular structure (Fig. 6). For example, a fiber made of Kevlar is very difficult to break because the polymer molecules in the fiber are stretched out along its length rather than being coiled around like tracks in a labyrinth. The strength of the Kevlar fiber is determined to a significant extent by the strength of carbon-carbon chemical bonds along individual polymer molecules. Fibers made of coiled polymer molecules, on the other hand, tend to stretch and break when pulled because the polymer molecules are only weakly linked and tend to slide over one another.

We know that individual polymer molecules are made of several thousand repeated subunits, called monomers, and that they are very long; an average polymer with a molecular weight of 200,000 daltons has a stretched length of several

(a) UHMW PE Gel
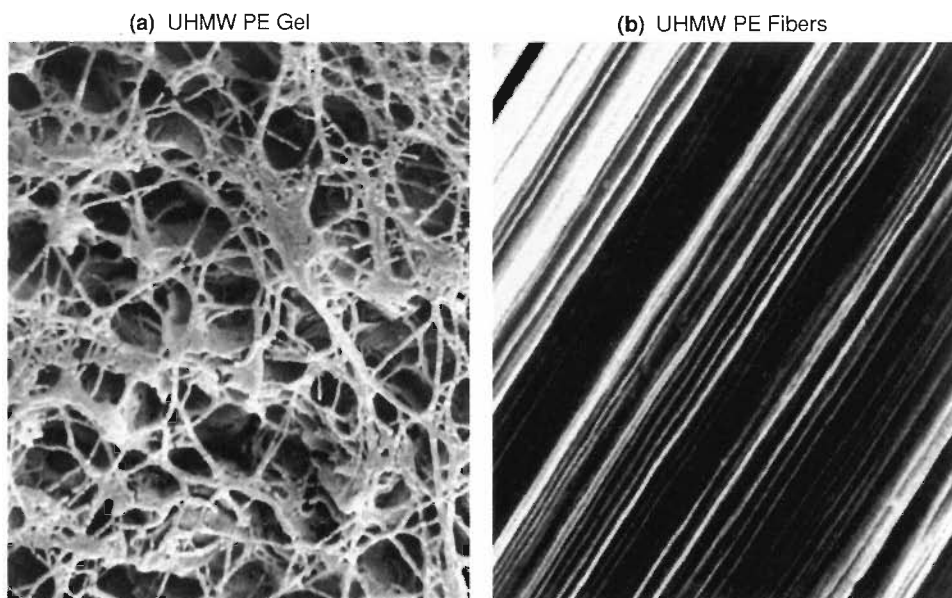
(b) UHMW PE Fibers

ULTRA-HIGH-STRENGTH FIBERS

Fig. 6. The tensile strength and stiffness of fibers made from ultra-high-molecular-weight polyethylene (UHMW PE) depend on the extension and alignment of individual polymer molecules—properties that depend, in turn, on the degree of polymer cross-linkage in the gels from which the fibers are formed. (a) A 2-percent solution of UHMW PE molecules was cooled to form the gel shown in this scanning electron micrograph (magnification = 1000). The structure of the gel reflects agitation of the solution prior to cooling. (b) This scanning electron micrograph shows a dried gel from a 0.6-percent solution of UHMW PE that has been drawn to 130 times its original length. (The draw direction was diagonal.) Note the extremely high degree of orientation. Over the last decade advances in techniques for manipulating the structures of UHMW PE gels by changing conditions and concentrations of the original solutions have led to rapid increases in the tensile strength of commercial fibers. (Micrographs courtesy of Paul Smith, University of California, Santa Barbara.)

microns (1 micron = 10,000 angstroms). But the long molecules are not usually stretched out unless special techniques have been used for synthesis, so how are they usually arranged? Before neutron scattering answered this question in the early seventies, there were several competing theories. One was based on a random coiling of the polymer molecules—much like a plate of well-mixed spaghetti. Another held that a large fraction of the polymer molecules were quasiparallel over some large fraction of their length.

According to the random-coil model, the radius of gyration of a polymer

molecule—which is a measure of the length of the molecule just as the diameter of a loosely wound ball of wool is a measure of the length of wool in the ball—ought to be proportional to the square root of the molecular weight of the polymer if the polymer is suspended in an *ideal solvent*, one that supports molecules rather than changing their conformation. Neutron-scattering researchers tested this prediction using small-angle scattering. Once again, the technique of selective deuteration allowed them to see a few select polymer molecules winding through a conglomeration of many others—essentially

100

the deuterated polymer molecules stood out from the background of their hydrogenated brethren just as a strand of spinach spaghetti would stand out in a plate full of the normal variety. The radii of gyration obtained by neutron-scattering experiments of this type for many different polymers were in impressive agreement with the prediction of the random-coil model. However the argument did not conclusively prove the model correct because it could not rule out a significant degree of local molecular alignment. To investigate this point, neutron-scattering data had to be obtained at greater values of Q, between about 0.1 and 0.6 inverse angstroms, where the conformation of the molecules is probed over distances smaller than their size, that is, between 10 angstroms and 50 angstroms. The idea was to see if polymer molecules line up with each other along segments within that size range. Comparisons of such data and calculations based on the random-coil model showed very good agreement for a variety of amorphous polymers and no indication of quasiparallel packing.

Fractal Objects

Many of the objects encountered in sludge science are disordered—polymers and ceramics, for example. Such materials seldom display the translational and rotational symmetries encountered in crystals and so cannot be characterized geometrically by such concepts. Fortunately a type of geometry called fractal geometry can be applied to the description of many disordered materials. Moreover, neutron scattering can be used to study objects that exhibit this type of geometry.

Fractal geometry was introduced almost two decades ago by Benoit Mandelbrot to describe objects—such as trees, clouds, mountains, and lakes—that cannot be described by normal Euclidean geometry because their surfaces or coastlines are too "bumpy." Mandelbrot's central observation was that many of the objects whose structure he sought to describe are self-similar when dilated (Fig. 7). For example, if we magnify a piece of coastline on a map, the image we obtain looks very much like the original. Small inlets now look much like fjords, and fjords have become ocean. Similarly, under dilation small islands in an archipelago look like the larger ones on the original map, but the overall appearances of the enlarged and original maps are very similar. This type of self-similarity is usually referred to as *statistical* to distinguish it from the exact dilational symmetry displayed by geometrical patterns such as the Sierpinski gasket. Statistically self-similar objects—clouds for example—look approximately the same under dilation.

Self-similarity is an appealing concept, and one can easily list examples of self-similar, or fractal, objects from many fields of study: blood vessels, lungs, tree bark, lightning, and galaxies are obvious candidates. But how do we say something exact about them? The answer was provided many years ago by Hausdorff and other pure mathematicians who had no idea that their theorems would be applied to natural phenomena. They examined how some measure of an abstract object, such as the distance between elements of the object, changes with the length scale on which this property is examined. This basic idea is easily applied to concrete examples such as the coastline of Norway. Suppose we try to measure the length of this jagged coastline. Clearly, if we use a long "ruler" we will jump across fjord entrances and get the wrong answer. To avoid this problem we could use a shorter ruler. But since the coastline is self-similar we will still jump over some inlets—they will just be smaller than the ones we missed with the larger ruler. Of course

we could continue this process *ad nauseam*, using smaller and smaller rulers, but because the coastline is self-similar, the total length we measure will always depend on the length of the ruler we use. Such a result clearly has no place in Euclidean geometry! However, if we use enough different rulers, we eventually find that the length of the coastline is proportional to the ruler length raised to the power $(1 - D)$, where D is called the *fractal dimension* of the coastline. For the coastline of a Euclidean object, such as a triangle, the fractal dimension is unity—the usual Euclidean dimensionality of a line—so its length is (fortunately for Euclid) independent of the length of the ruler used to measure it. The fractal dimension of the Norwegian coastline is about 1.2, illustrating that D is not usually an integer for fractal objects.

The concept of fractal dimension can also be applied to "solid" objects. The mass m of a sphere of uniform density scales as the cube of its radius r: $m = \frac{4}{3}\pi\rho r^3$. What about the mass of a spherical hunk of foam or a porous rock? Such a self-similar structure contains a great deal of "air"; so its mass fractal dimension—the power of a single dimension that is proportional to the object's mass—will be less than three. By the same token, a fractally rough surface has a larger area than a smooth surface; so its fractal dimension is greater than 2, the Euclidean dimension of smooth surfaces.

Materials scientists use fractal geometry because many growth and aggregation processes tend to form disordered systems that can only be characterized as fractal. Even simple polymers are fractal: according to the random-coil model of polymers mentioned earlier in this article, the mass of a polymer in an ideal solvent scales as the square of its radius of gyration; its mass fractal dimension is 2. Other examples range from dendrites in metals to glasses. Of-

(a) Sierpinski Gasket

(b) Fractal Coastline

$\times 2^1$

$\times 2^2$

$\times 2^5$

$\times 2^8$

$\times 2^{12}$

$\times 2^{16}$

$\times 2^{20}$

$\times 2^{24}$

$\times 1$
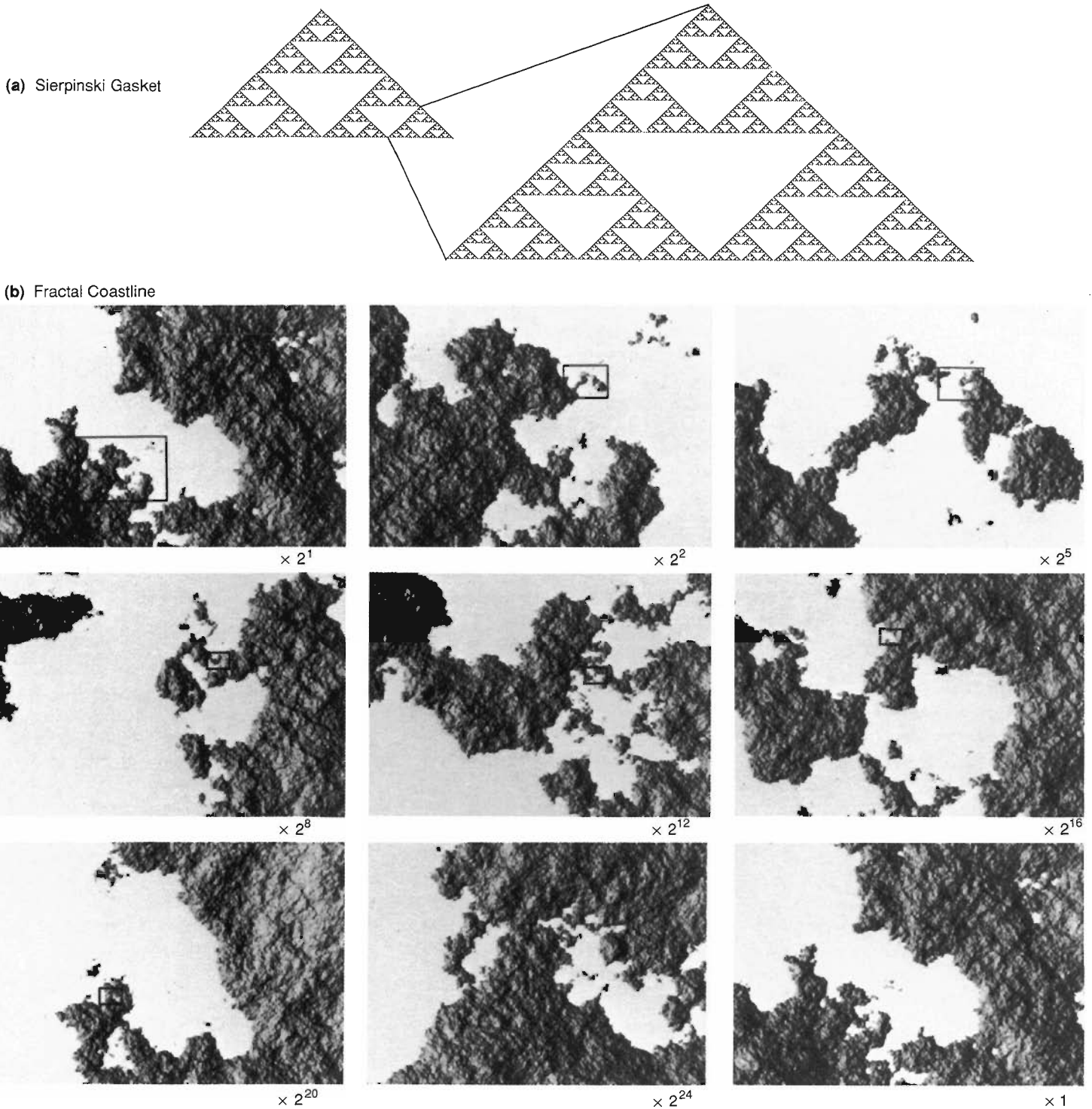
SELF-SIMILARITY

Fig. 7. The concept of self-similarity is essential to fractal geometry. (a) In the exactly self-similar pattern known as the Sierpinski gasket, any upright, triangular section of the pattern can be be made identical with any other upright, triangular section by dilation or shrinking. (The pattern is assumed to repeat indefinitely.) Obviously the maps shown in (b) do not display such exact dilational symmetry—it is impossible to expand one and overlay it exactly on another. On the other hand, enlargement of the part of each figure inside the small box gives the suceeding figure, and one can easily imagine that both figures are part of the same map. (The original map appears at the end of the series of enlargements to show that this similarity exists between coastlines of any magnification.) Such shapes are called statistically self-similar. The algorithm used to generate these pictures, called fractal Brownian motion, is similar to that employed by film-makers who wish to draw realistic mountains and lakes. (Fig. 8b ©Richard Voss/IBM Research.)

ten fractal patterns appear in systems that develop far from thermal equilibrium. For example, under non-equilibrium conditions aggregating particles may stick to a growing cluster at positions other than those that would yield the lowest free energy, whereas at thermal equilibrium particles tend to bounce around until they find those positions. For this reason open, dispersed structures are formed rather than the compact crystalline structures typical of growth under equilibrium conditions (Fig. 8). The type of structure produced depends to a large extent on the details of the non-equilibrium growth process. Materials researchers have begun simulating such growth processes on computers (Fig. 9) and comparing the structures the simulations produce with the structures of samples produced in carefully controlled experiments. The essential grounds for comparison are the fractal dimensions of the structures. Neutron and x-ray scattering are ideal techniques for measuring the fractal dimension of macromolecular structures because they can probe length scales ranging from interatomic distances to the sizes of many molecular aggregates (Fig. 10). Furthermore the scattering laws for fractal systems are simply power laws— the scattered-neutron intensity is proportional to Q raised to some power. For mass fractals the exponent is equal to $-D_m$, where D_m is the mass fractal dimension of the scattering object, whereas for compact objects with fractally rough surfaces, the exponent is equal to $6 - D_s$, where D_s is the fractal dimension of the surface. Thus, for fractal objects encountered in sludge science, the slope of a plot of the logarithm of scattered-neutron intensity versus the logarithm of Q is simply related to D_m and D_s. By comparing experimental values of D_m with those predicted by different growth models, the growth process in each experiment can be inferred. Because no naturally
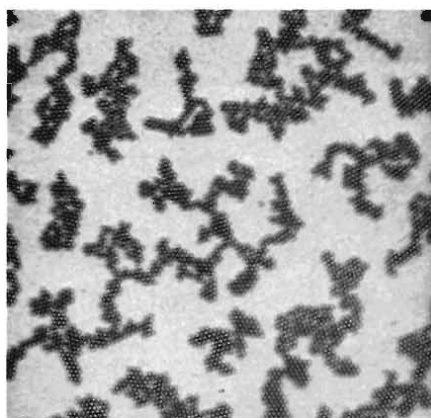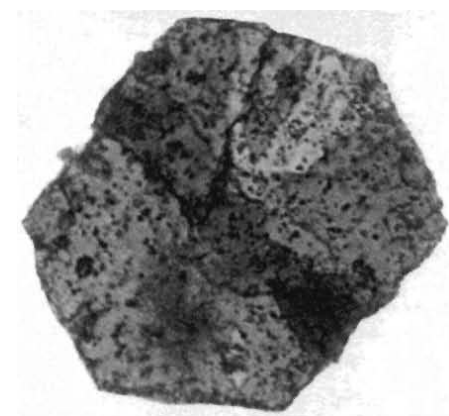

$$\frac{d\bar{r}}{dt} = 2 \times 10^{-2} \ \mu m/s$$
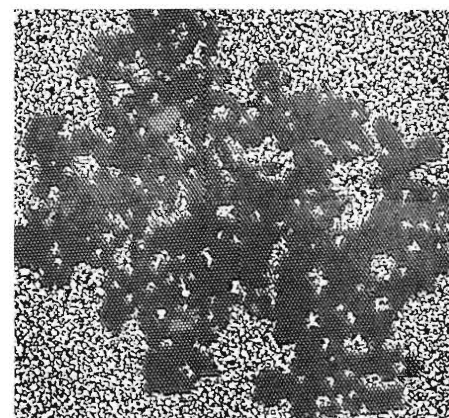
$$\frac{d\bar{r}}{dt} = 3 \times 10^{-4} \ \mu m/s$$
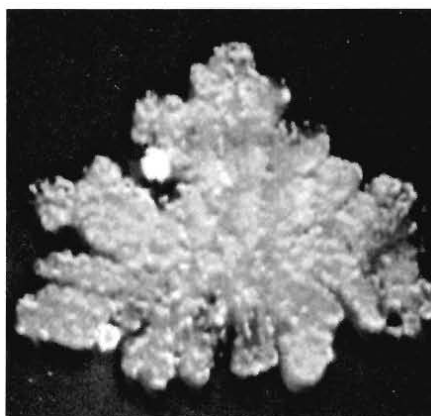
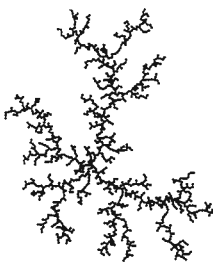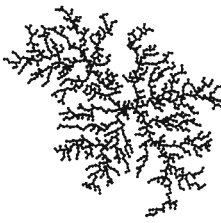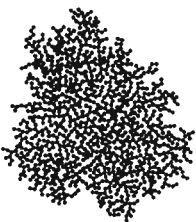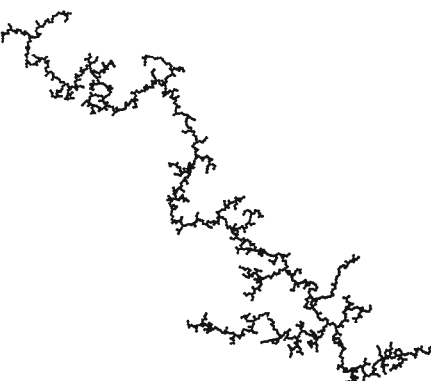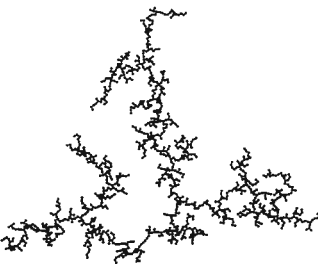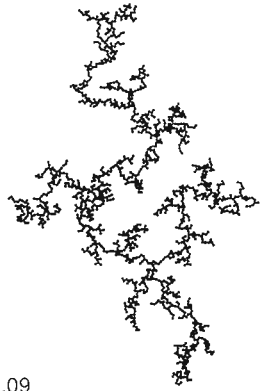$$\frac{d\bar{r}}{dt} \approx 1 \times 10^{-5} \ \mu m/s$$

$$\frac{d\bar{r}}{dt} \approx 5 \times 10^{-6} \ \mu m/s$$

PRODUCTS OF DIFFERENT AGGREGATION RATES

Fig. 8. In an elegant experiment with simple equipment, Skjeltorp was able to show that the form of aggregated matter depends sensitively on the speed of accretion. A colloid of negatively charged, 1.1-micron-diameter plastic spheres was suspended in water and placed on a microscope slide, together with a few 5.5-micron spheres. When a cover slip was placed over the drop, the larger spheres acted as spacers, defining a single layer within which the smaller spheres moved in a Brownian fashion as a result of collisions with water molecules. By adjusting the salinity of the colloid, Skjeltorp was able to control the repulsion between the spheres, and thus the rate of accretion. With a small repulsive interaction, particles tend to stick together when they collide; with a larger repulsive interaction, the particles undergo many collisions, exploring configurations of different total energy before finding one that lowers the energy of the system enough to allow accretion. Thus, larger repulsive interactions correspond to slow growth under conditions close to thermal equilibrium. The figure shows micrographs of the aggregates produced and the rate of change of average radius for each aggregate. (The crystalline aggregate took over six months to form.) Experiments like this one provide a visual framework for interpreting the fractal dimensions of molecular aggregates, data which can be obtained by neutron-scattering experiments. This experiment was reported by Arne Skjeltorp in "Visualization and characterization of colloidal growth from ramified to fractal structures" (*Physical Review Letters* 57, p. 1444 (1987)).

Diffusion-Limited Aggregation	Ballistic Aggregation	Reaction-Limited Aggregation
Monomer-Cluster Growth Witten and Sander $D = 2.50$	Vold $D = 3.00$	Eden $D = 3.00$
Cluster-Cluster Growth DLCA $D = 1.80$	Sutherland $D = 1.95$	RLCA $D = 2.09$

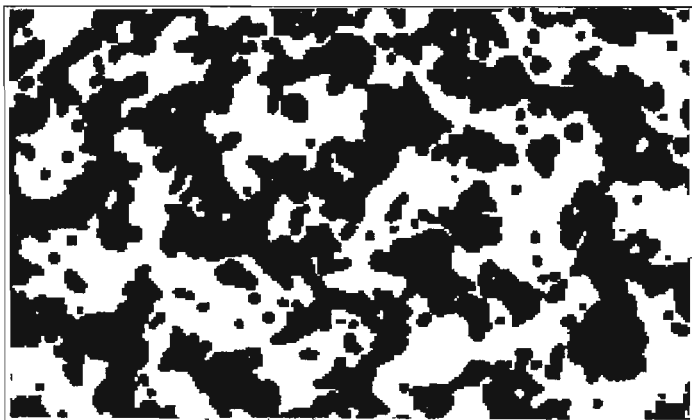## COMPUTER MODELS OF NON-EQUILIBRIUM GROWTH PROCESSES

Fig. 9. Computer simulated non-equilibrium growth processes produce three-dimensional structures with characteristic mass fractal dimensions (their two-dimensional analogs are shown here). Diffusion-limited aggregation (DLA), invented by Witten and Sander, simulates growth by placing a seed particle, or monomer, at a point on a grid and allowing other, similar particles to diffuse inwards from the edges of the grid in a series of random steps. The cluster, initially represented by the seed, grows if a particle lands on a site next to it. Because the particles move randomly, few of them ever penetrate to the interior of a growing cluster and accretion takes place at the tips. This so-called monomer-cluster growth results in a highly ramified structure with a mass fractal dimension of 2.5. An even more open structure, with a fractal dimension of 1.8, is formed by cluster-cluster growth, that is, if the growing DLA clusters are allowed to move around and aggregate with each other. The ballistic growth model invented by Vold—which is applicable to some gas-phase growth processes—assumes the same aggregation conditions as DLA, except that the particles travel in straight-line, or ballistic, paths before sticking to a cluster. Thus some penetrate to the center of the growing cluster, and a more compact aggregate is formed. Again, if clusters are allowed to move and aggregate, a more open structure is formed. Reaction-limited aggregation, which was invented by Eden, simulates growth in which some chemical barrier to aggregation, rather than the type of particle motion, determines when a particle sticks to a cluster. Thus a particle must land on sites next to the growing aggregate several times before sticking, and extremely compact particles are formed. Again, if clusters are allowed to move and aggregate, a more open structure is formed. Most branched polymers in solution follow reaction-limited kinetics. The important point to note is that the fractal dimensions of the clusters, and the visual impressions they give, vary considerably with the growth conditions. (Figure courtesy of Paul Meaken; reprinted with permission from Kluwer Academic Publishers.)

occurring object can be self-similar under dilation of arbitrary magnitude, a straight-line plot is obtained only over a range of Q values corresponding to the specific length scale of self-similarity. For example, the aggregates shown in Fig. 9 can be self-similar only on length scales between the size of an individual particle and the size of a cluster.

Fumed Silica. Alan Hurd and his collaborators have successfully used neutron- and light-scattering techniques combined with computer growth models to investigate the growth of fumed silica particles. This material is prepared commercially by burning $SiCl_4$ in hydrogen and oxygen to produce highly ramified silica clusters, which are used to modify the flow properties of foods and paints. (Next time you visit your local supermarket check the ingredients of common foods. It is not unusual to find that one of the ingredients is silica—sand by another name!) A similar process is used to make the starting materials for optic fibers. By conducting scattering experiments on samples obtained from

Phase-Separated Material

Dominant Structures at Decreasing Length Scales

10,000 Å	1000 Å	100 Å	10 Å
Individual Clusters	Mass Fractal	Surface Fractal	Atomic Lattice

Guinier Regime — Porod Regime — Bragg Regime

Slope $= -D_m$

Slope $= -(6 - D_s)$

$\log I(Q)$

Q (Å^{-1})

SCATTERING LAWS AT DIFFERENT LENGTH SCALES

Fig. 10. The black and white regions represent single-phase fractal clusters that are assumed to have separated from a highly disordered material by a non-equilibrium process. The shape of the scattering law for the phase-separated material (plotted as $\log I(Q)$ versus $\log Q$) changes as Q increases because the material's nonuniformity is being probed at different length scales. Small values of Q probe long length scales, whereas large values of Q probe short length scales. In the Guinier region, where $QR \approx 1$, the scattering law measures the overall cluster size R of each phase. As Q increases, the internal fractal structure of the clusters is probed, and a scattering law of the form Q^{-D_m}, where D_m is the mass fractal dimension, is observed. When Q exceeds a value close to the inverse size of the holes in the fractal clusters, the surfaces of the aggregates are probed, and the scattering law changes to the form $Q^{-(6-D_s)}$, where D_s is the surface fractal dimension of the clusters. Finally, at large values of Q, the scattering pattern evolves to a form containing Bragg peaks, which are characteristic of structural variations on atomic length scales. (Adapted from a figure in an article by Dale Schaefer. *Proceedings of the Royal Society of London* A423:37, 1989.)
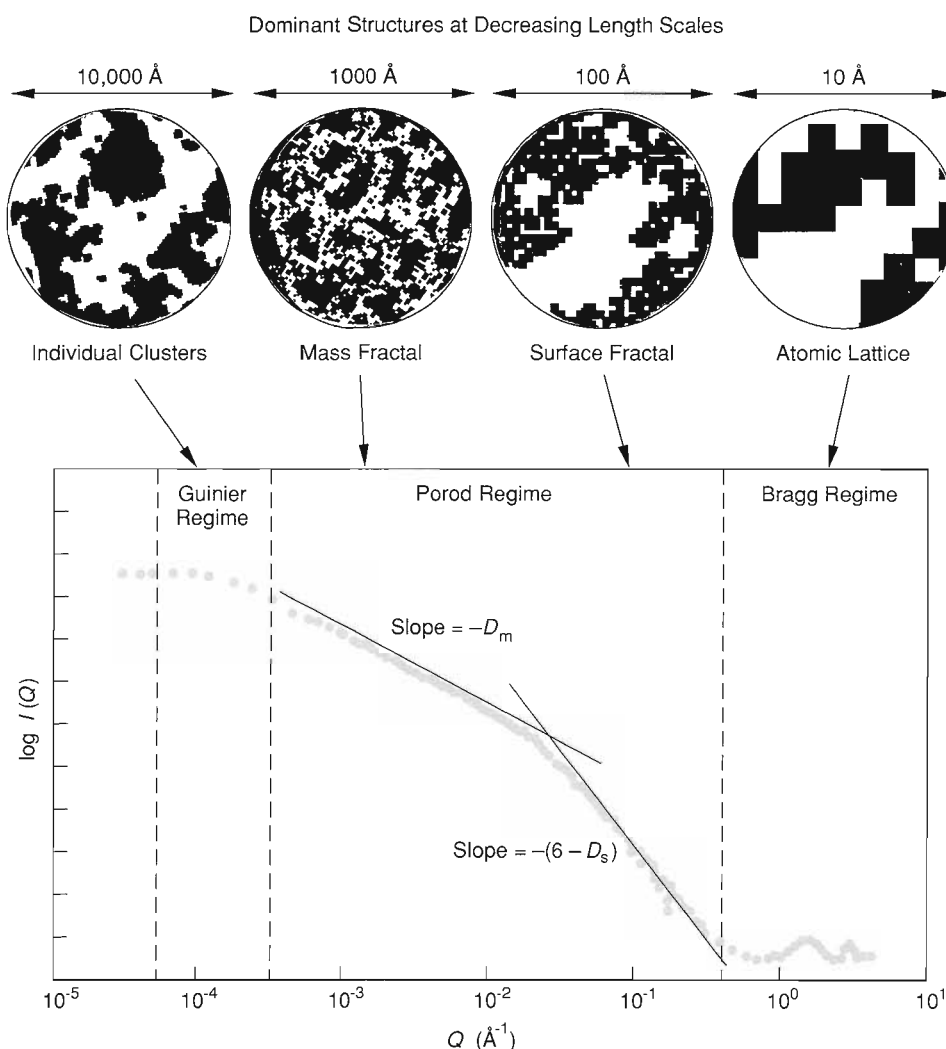
$SiCl_4/H_2/O_2$ flames, Hurd and his collaborators deduced that three processes are involved in the formation of fumed silica particles. In the early stages of growth SiO_2 monomers aggregate to form compact, rough-surfaced clusters resembling those produced by the ballistic monomer-cluster aggregation process (Fig. 9) and grow to sizes approaching 100 angstroms. Because the flame temperatures are near the melting point of silica, annealing competes with aggregation and the particles become smoother as they grow. Later in the growth process, when no more monomers are present, these smooth particles tend to coalesce, forming micron-sized aggregates whose fractal dimensions characterize

Sludge Physics
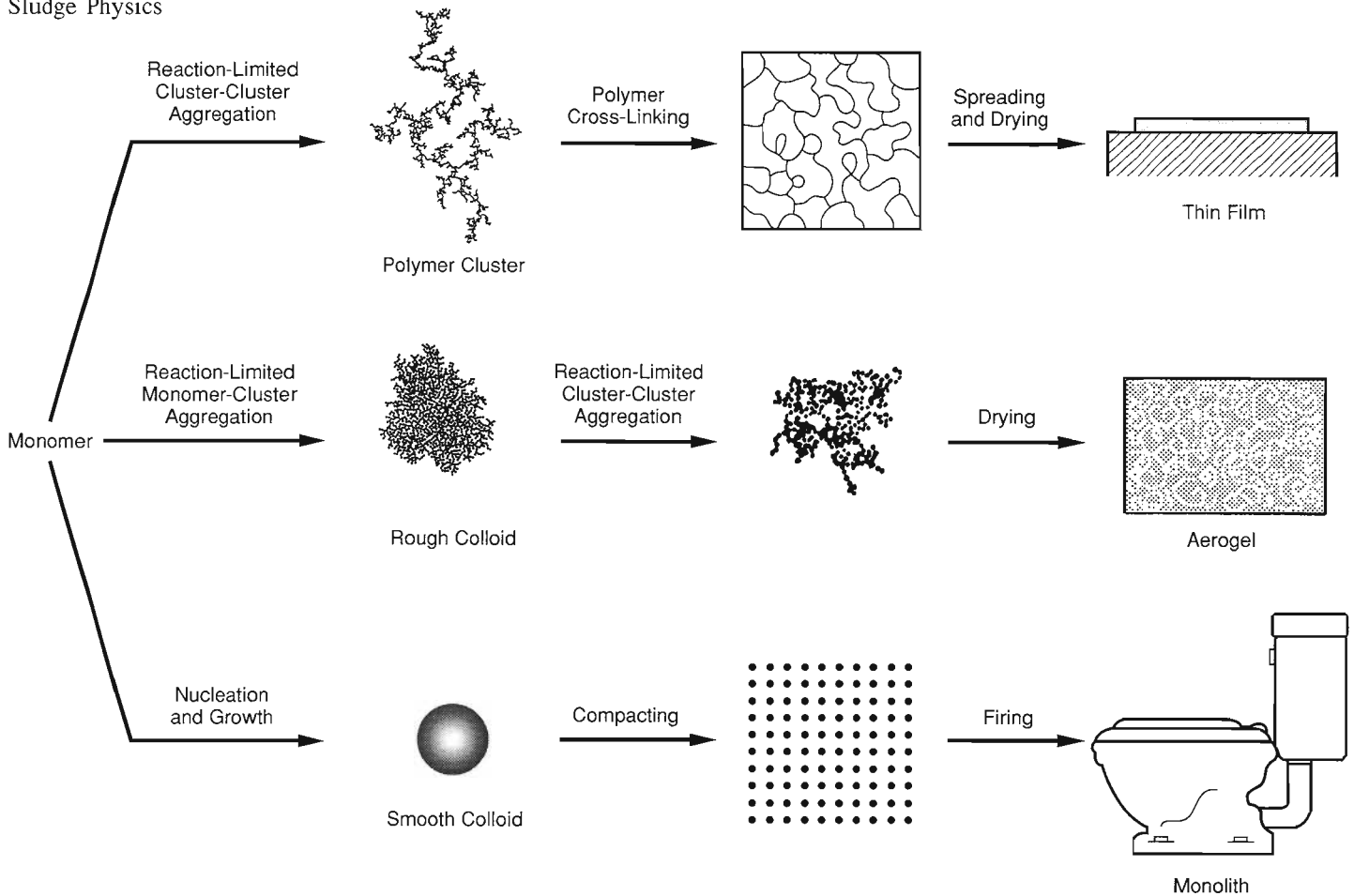
CERAMIC PRODUCTS AND THEIR PRECURSORS

Fig. 11. Depending on the initial conditions of growth, precursor particles made from a single monomer can lead to final products with widely varying properties. Reaction-limited cluster-cluster aggregation under acidic catalysis conditions yields polymerized precursors with a fractal dimension of 2.09. The highly ramified polymers then cross-link to form gels that collapse on drying into hard, high-density films. Such films are used as insulating layers in integrated circuits, and research is under way to use them as ferroelectric, nonvolatile computer memory-elements. Reaction-limited monomer-cluster aggregation forms compact, rough-surfaced colloidal particles, which then aggregate through cluster-cluster growth. Because the gels so formed are denser and much stronger at short length scales than those already discussed, they do not collapse when dried but rather form extremely porous, diaphanous materials called aerogels. Researchers hope to exploit the high surface area of aerogels to make transparent materials for insulating passive solar walls; however, present production costs are too high. In the traditional method of producing hard, durable, ceramic material for everyday use, smooth colloidal particles are grown under near equilibrium conditions, compacted, and then fired. (Adapted from a figure in *Science* 243 (1989): 102.)
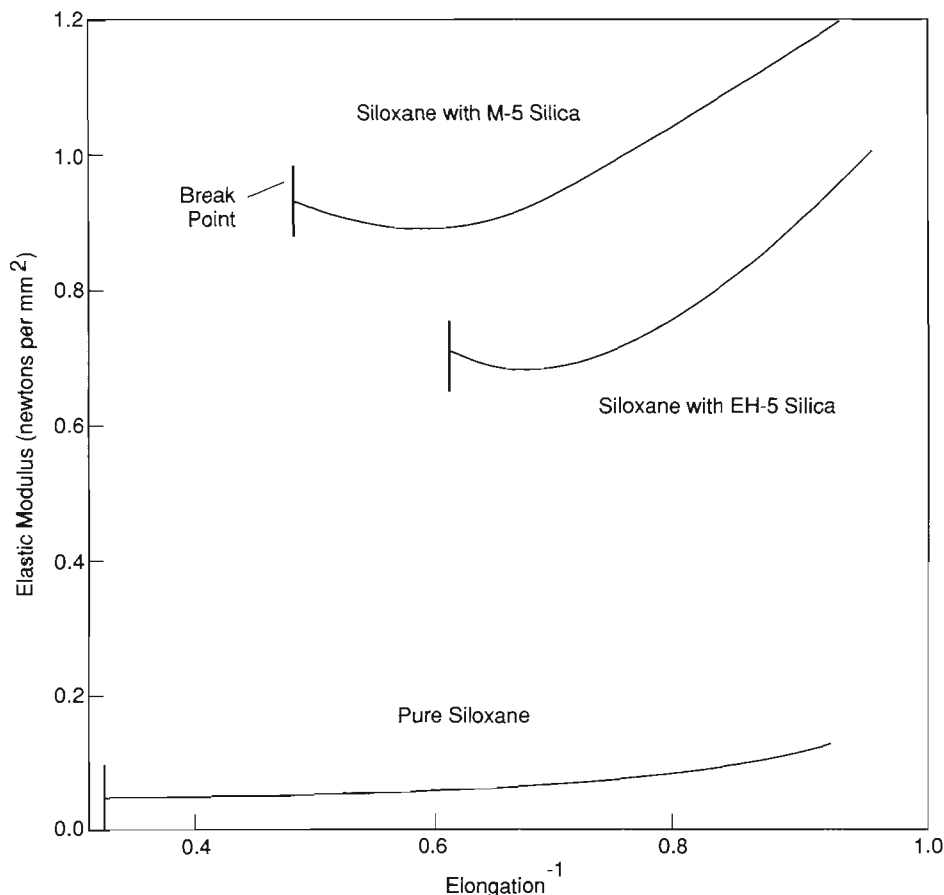
them as products of the diffusion-limited cluster-cluster aggregation process.

The importance of experiments like those carried out on fumed silica is that they provide an understanding of the relationship between growth conditions and the structures of the aggregates produced (Fig. 11). Materials scientists can use this information to tailor growth conditions and produce the structures required for particular applications. For example, silica aggregates, similar in chemical composition to the fumed silica described in the previous paragraph, can be produced in solution by polymerization of silicic acid. Depending on the pH of the reaction bath, how-

ever, quite different structures are produced. Catalysis of the reaction under acidic conditions produces a highly ramified, polymeric silica with a mass fractal dimension of about 2 by the reaction-limited cluster-cluster aggregation process. This silica collapses easily on drying and produces high quality protective films. On the other hand, a rough-surfaced, compact, colloidal silica, which is formed under basic catalysis conditions by the reaction-limited monomer-cluster process, is resistant to collapse and much more suitable for the production of porous, high-surface-area materials. Once again we see that an understanding of structure, which can

be provided by neutron scattering, is a prerequisite for the development of strategies for material preparation.

Composite Materials. Composites are multi-component materials in which an attempt is made to retain the useful properties of each component. For example, fumed silica is often combined with silicone elastomers to increase the strength of the elastomer while retaining its elastic properties. Figure 12 shows some typical experimental results obtained by Dale Schaefer and his collaborators at Sandia National Laboratories. The lower curve shows that the elastic modulus of pure polydimethyl-

Fig. 12. The properties of siloxane composites containing fumed silica as a reinforcing agent vary widely depending on the structure of the fumed silica particles. Pure siloxane rubber can be stretched to approximately four times its original length before it breaks. It also has a low elastic modulus, so little force is required to stretch it. (Elastic modulus is defined as the ratio of the applied force per unit cross-sectional area to the elongation, where the elongation equals the ratio of the change in length to the original length.) Pure siloxane rubber also grows progressively weaker as it stretches. Siloxane reinforced with rough-surfaced EH-5 fumed silica has a higher elastic modulus and breaks at a much lower elongation than pure siloxane. Smooth-surfaced M-5 fumed silica is a better reinforcing agent; it forms a composite that has a somewhat higher elastic modulus and breaks at an intermediate elongation. Researchers had expected composites of siloxane and the rough-surfaced EH-5 to perform best. They now think that particle size may be a more important property than surface texture because M-5 particles are larger than EH-5 particles. In addition, the higher-temperature growth process that smooths the surfaces of the M-5 particles may also strengthen them and hence the composite rubber. (Figure courtesy of Dale Schaefer, Sandia National Laboratories.)

STRENGTH OF SILICA-SILOXANE COMPOSITES

phous forms of carbon—and the properties of the rubber produced when carbon black is used as a reinforcing agent. Information obtained in these investigations is potentially of great economic importance—a decrease of 1 percent in the fuel consumption of automobiles due to improvement in tire rubber is worth a billion dollars or more every year!

Molecular composites differ from the conventional variety described above in that the filler is generated *in situ* during production rather than by mixing. Multiphase materials of this type, in which the phases are dispersed at the 100-angstrom level, offer great possibilities for the enhancement of material properties, but it is difficult to design materials until the relationship between synthetic protocol, structure, and properties has been established. A case in which these relationships have been investigated is that of composites made by swelling a polydimethylsiloxane rubber with tetraethylorthosilicate and polymerizing the swelled material to produce SiO_2—that is, glass—precipitates. This process was pioneered by J. E. Mark at the University of Cinncinati. The structure of the precipitated silica turns out to be very sensitive to chemical conditions prevailing during the synthesis, and the resulting silica and rubber composites vary significantly in elasticity and strength. Under some conditions, scattering experiments show that a bicontinuous network of glass and rubber is formed by a process called spinodal decomposition, yielding brittle composites. Under different chemical conditions the glass filler appears to have a polymeric mass-fractal character that enhances the extensibility of the resulting composite. Those who have studied molecular composites of this sort admit that they are often wrong in their intuitive guesses about the relationship between filler morphology and the mechanical properties of the resulting composite. There is no substitute for a

siloxane is small and becomes smaller as the material is elongated (that is, as we move left along the horizontal axis). A small elastic modulus means the material stretches with little applied force. After three-fold elongation the material breaks—there are no further data points to the left. When the elastomer is reinforced with two types of fumed silica, M-5 and EH-5, much higher elastic moduli are found, but the composite materials break at smaller elongations

than does the pure "rubber." Scattering data show that the M-5 fumed silica has a smooth surface whereas the EH-5 fumed silica has a rough surface. Therefore, the elastic modulus data indicate that smooth silica particles produce better reinforcement than rough ones—a somewhat counterintuitive result. Similar studies are underway at LANSCE to examine the relationship between the structures of carbon blacks—usually thought of as amor-

careful cataloguing of the relationships between synthetic protocol, structure, and properties.

Surfaces

Earlier in this article I described a way of using small-angle neutron scattering to examine particle-like structures formed by amphiphilic molecules. Complementary information concerning surface structures formed by amphiphilic molecules can be obtained from neutron reflectometry by adsorbing the molecules on a liquid and reflecting neutrons from the air-liquid interface (see "Neutron Scattering—A Primer" for an account of neutron reflectometry). As is so often the case, substitution of deuterium for hydrogen in both the molecules and the liquid increases the power of neutron scattering for such studies. One can even make the liquid a mixture of H_2O and D_2O that does not reflect neutrons at all, so that only the amphiphilic molecules at the surface are visible to neutrons! Both the amount of material adsorbed at the surface and many of the structural properties of the surface layer can be determined by reflectometry. For example, measurements on a variety of amphiphilic molecules show that the adsorbed layer usually consists of three principal regions. The first, nearest the vapor, contains only hydrocarbon chains; the second contains a small fraction of chains, the negatively charged head groups, solvent, and a few cations such as Na^+; the final region contains solvent and a diffuse atmosphere of ions. The thickness of each region can be established by reflectometry, as can the fractions of the various components in each region. When the thickness of the first region is less than the length of the hydrocarbon chains, an average tilt angle for the chains can sometimes be determined. The strong resemblance these layers bear to one side of the lipid bilayer discussed ear-

lier suggests that neutron reflectometry can be used to study model biological membranes; indeed, such studies have already started at several facilities.

Block copolymers—polymers made up of a string of monomers of one type (block A) linked to a string of monomers of another type (block B)—are used as surfactants, compatibilizing agents in polymer blends, and adhesives in biomedical and microelectronics applications. Very often the individual polymer blocks segregate, giving rise to heterogeneous arrangements consisting either of alternating layers (*lamellae*) or of rods of one block embedded in a matrix composed of another block. Such structures are often disorganized in bulk samples, making them very difficult to study. For example, the lamellae may be broken up and disorganized with respect to their neighbors. In copolymer films, on the other hand, one or other polymer block usually has an affinity for the film surface; thus lamellae are formed in an ordered way, parallel to the surface of the film. Neutron reflectometry is an ideal tool for investigating such copolymer films because the contrast variation technique, implemented through selective deuteration, can be used to highlight one or the other of the polymer blocks.

The copolymer films studied so far at LANSCE (and other neutron-scattering centers) are formed by placing a drop of solvent containing the copolymer on the surface of a silicon wafer and spinning the wafer at high speed to produce a uniform, thin film of copolymer as the solvent dries. The film is a random alloy of its polymer components at this stage, with both polymer blocks thoroughly mixed and entangled (see Fig. 13). When the polymer film is annealed, however, the different polymers segregate into alternating layers parallel to the plane of the film. The layers tend to form first at the top and bottom surfaces of the film, and the

sequence of layers propagating from each surface depends on the affinity of each polymer for the interface in question. Thus, if the system has a lower free energy when polymer-block A attaches to the silicon substrate than when polymer-block B does, the order will be A,B,A,... away from the substrate surface. The fully annealed film produces a complicated reflectivity profile from which it is possible to deduce the layer thicknesses as well as the definition and shape of the interfaces between adjacent polymer blocks. What happens during the annealing process to resolve the competition between the order of layers propagating from each face of the film is still somewhat of a mystery and the subject of continuing investigation at LANSCE and elsewhere.

Such studies of polymer compatibility bear on the crucial economic and environmental problem of recycling plastics. In 1984 the U.S. produced 46.3 billion pounds of plastics and discarded 18 billion pounds of plastic wastes. Recycling of the wastes would not only save disposal costs but would also reduce energy consumption; it is estimated that producing a fabricated plastic from waste instead of virgin resin would result in an 80-percent saving of energy. Unfortunately, it is not easy to make different plastics compatible. Various strategies such as using block copolymers to knit together incompatible homopolymers or using "radiation grafting" to form new materials from incompatible plastics have been proposed. To find out how well those techniques work, one may use neutron reflectometry to study the molecular structure of interfaces so formed between different polymers. The same message that we have seen in other areas is once again clear, but in a slightly different form. In this case, we need to understand the relationship between structure and properties in order to invent a new way of processing plastics.
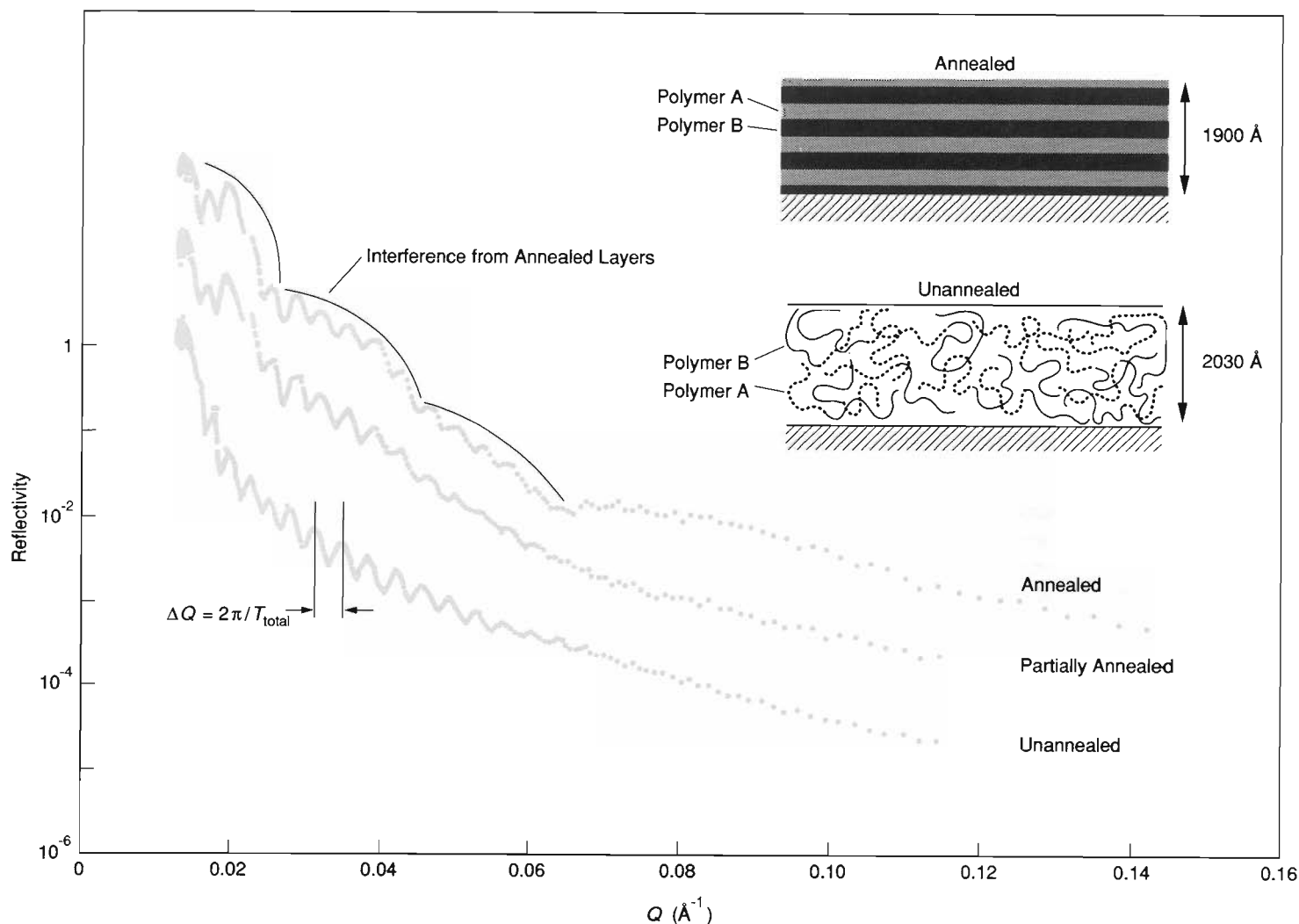
NEUTRON-REFLECTOMETRY MEASUREMENTS ON A BLOCK-COPOLYMER FILM

Fig. 13. Neutron-reflectivity profiles at three stages of the annealing of a diblock-copolymer film reveal the gradual formation of alternate layers of polymer A (polystyrene) and polymer B (polymethylmethacrylate). (Two of the profiles have been shifted vertically for easier reading.) The reflectivity profile of an unannealed, homogeneous film is a series of evenly spaced peaks of decreasing amplitude (an interference fringe pattern) produced by the interference of neutron waves reflected from the film's top and bottom surfaces. The difference between the Q values at which successive peaks occur, ΔQ, is inversely proprtional to T_{total}, the thickness of the homogeneous film. The data shown here indicate a thickness of 2030 angstroms for the unannealed film. The reflectivity profile for the fully annealed film (heated for 24 hours at 170°C) indicates a series of alternating microphase layers, each made up of a single polymer. The closely-spaced ripples due to interference from the whole film are superimposed on a new series of broader interference peaks (suggested by hairlines above the curve) due to the layers formed during annealing. Secondary-ion-mass-spectroscopy measurements indicated that each of these layers contained only one polymer and had a thickness of about 250 angstroms. Neutron-reflectometry measurements helped elucidate the long-range order and the degree of interpenetration of polymer layers. The alternating layers of polymer A and polymer B form because like polymers have an affinity for each other and tend to segregate together. (The bottom and top layers are half the size of the intermediate layers because they are formed from only one polymer block.) The reflectivity profile for a partially annealed film shares features of the other two. These measurements are part of a series of experiments to elucidate the processes and spatial relationships that dictate the final order. (The measurements were taken at LANSCE for Tom Russell of IBM.)
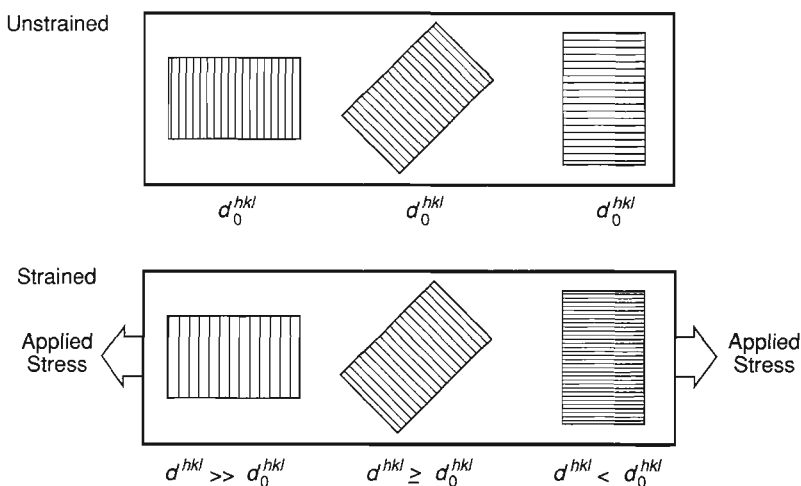
(a) Experimental Setup

Incident Beam

Applied Stress

Copper Bar

Detector −90°

Gauge Volume

Scattered Beam

Detector +90°

Enlarged View of Gauge Volume

Applied Stress

Bragg Diffraction from Planes Parallel to Stress Measures d Perpendicular to Applied Stress

d

d

Bragg Diffraction from Planes Perpendicular to Stress Measures d Parallel to Applied Stress

Residual Stress—A New Application of Neutron Scattering

Residual stresses—or residual forces per unit area—can be introduced into mechanical components both by fabrication processes and by deformations that occur during use. Any operation involving nonuniform deformation or thermal gradients—forging, welding, bending, and machining, for example—can give rise to residual stress. Whether the stresses are detrimental or beneficial to the strength and durability of a component, however, depends on the direction and magnitude of forces applied during use. In general, a residual stress that adds to an applied stress is detrimental to performance, whereas one that reduces the applied load is beneficial. For example, thick-walled pipes for a variety of applications in the chemical, nuclear, and armament industries are often subjected during use to a variety of cyclic internal pressures. Such tubes are frequently fabricated with compressive residual stresses on the inside, close to the bore, to inhibit crack propagation and extend lifetimes.

Anyone trying to predict the behavior of a mechanical component would want to know the distribution of residual stresses; however, most of the methods available for obtaining this information involve destruction or modification of the component—rather obvious drawbacks. An alternative, nondestructive method uses x-ray or neutron diffraction to determine local values of strain—that is, changes in the spacings between planes of atoms in the crystal. Stresses are then deduced from well-known relationships between stress and strain. At a spallation neutron source such as LANSCE, the deformation is determined by measuring the difference between the wavelengths of neutrons that diffract through a given scattering angle from a given set of planes in both stressed and unstressed samples. The differ-
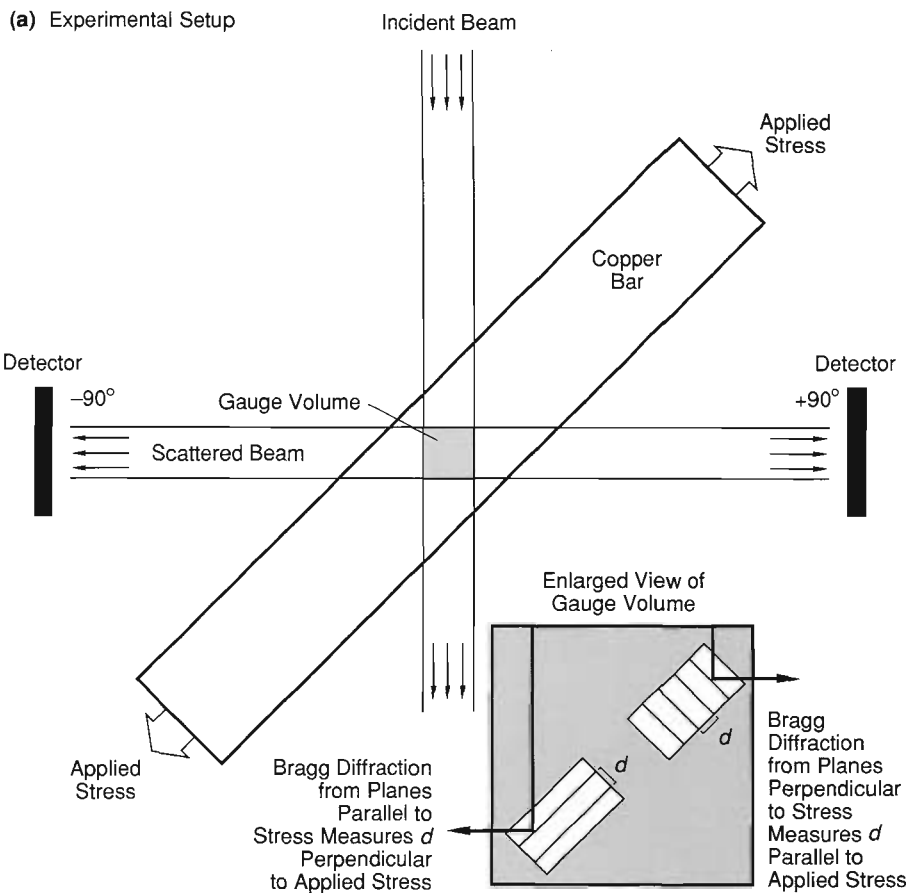
(b) Expected Changes in Lattice Spacing Resulting from Applied Stress

Unstrained

d_0^{hkl} d_0^{hkl} d_0^{hkl}

Strained

Applied Stress

Applied Stress

$d^{hkl} \gg d_0^{hkl}$ $d^{hkl} \geq d_0^{hkl}$ $d^{hkl} < d_0^{hkl}$

ent wavelengths are directly related to changes in the spacings of the atomic planes and thus to strains in various directions. Since the materials of interest are polycrystalline, Bragg peaks can be observed for any sample orientation unless the individual crystal grains exhibit considerable preferred orientation.

Depending on their application, the x-ray- and neutron-diffraction techniques each have unique advantages. X rays

are very useful for measuring surface strains; however, they do not penetrate most engineering materials sufficiently to be useful for bulk measurements. Neutrons, on the other hand, penetrate well in most cases—20,000 times farther than x rays in iron, for example—and so provide a nondestructive method of determining lattice distortions, or strains, throughout relatively large engineered components. Because the distor-

(a) Neutron-Diffraction Data

$\Delta d^{111} = 0.00248(6)$ Å
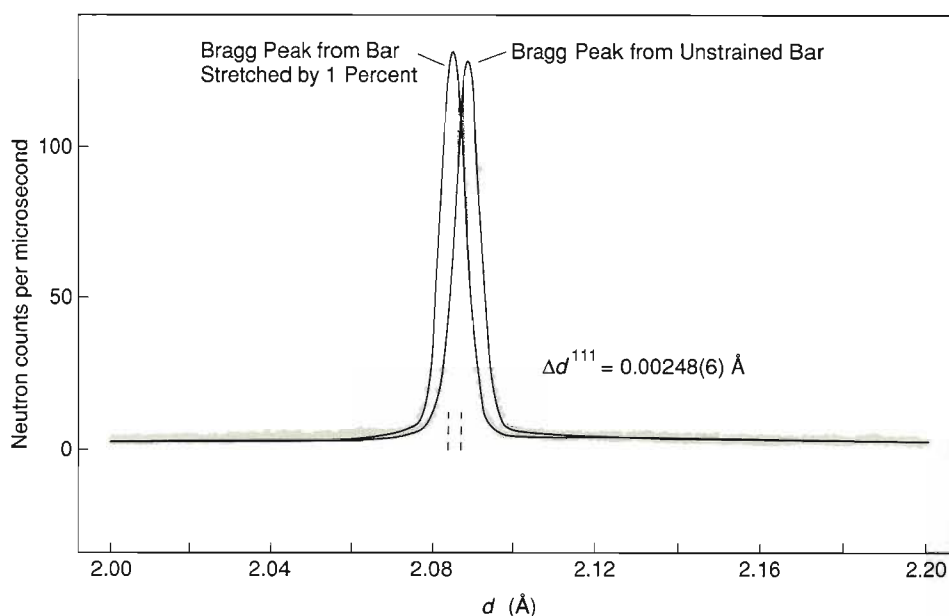
MEASURING STRAIN WITH NEUTRON DIFFRACTION

Fig. 14. Neutron diffraction can determine the strain in a polycrystalline material by measuring the stress-induced changes in the lattice spacings of crystallites, that is, in the spacings between sets of parallel planes of atoms. (a) Shown here is the experimental setup for measuring the strain in a copper bar stretched along its length. The bar is oriented at 45 degrees relative to the incident neutron beam and neutrons diffracted through ±90 degrees are detected. As shown in the blowup of the gauge volume (the volume in which lattice spacings are measured), this setup implies that the Bragg peak from a certain set of planes parallel and perpendicular to the applied stress are recorded in the -90 degrees and the +90 degrees detectors respectively. The lattice spacings d between those planes are related to the wavelength, and hence time of flight, of the neutrons that make up the Bragg peak. (b) Applied stress along the copper bar is expected to increase the lattice spacing between planes perpendicular to the applied stress and decrease the lattice spacing between planes parallel to the applied stress. Here d_0^{hkl} is the lattice spacing of the (hkl) planes in three variously oriented crystallites of an unstrained bar, and d^{hkl} is the lattice spacing of the same set of planes after the sample has been strained. The indices (hkl), called Miller indices, uniquely identify the set of lattice planes being measured.

(b) Strain-Induced Changes in Lattice Spacing

| Set of Lattice Planes (hkl)	$(d - d_0)/d_0$ for Planes ⊥ to Stress	$(d - d_0)/d_0$ for Planes ‖ to Stress
(111)	+0.119(3) %	−0.097(7) %
(200)	+0.209(3) %	−0.150(5) %
(220)	+0.130(4) %	−0.117(3) %
(222)	+0.113(5) %	−0.112(13) %

MEASUREMENT OF STRAIN IN A STRETCHED COPPER BAR

Fig. 15. To determine the microscopic elastic strains in a copper bar stretched by 1 percent, researchers compare the Bragg peaks for a certain set of planes before and after applying stress. (a) The position of the Bragg peak for the (111) planes shifts by an amount corresponding to a change in the (111) lattice spacing of 0.00248(6) angstroms. (The number in parentheses is the uncertainty in the last digit(s).) (b) The table summarizes the percentage change in lattice spacing, or $(d - d_0)/d_0$, for four sets of planes. The change is positive when the sets of planes are perpendicular to the applied stress and negative when the sets of planes are parallel to the applied stress, as expected. Note that the percentage changes in the lattice spacings of parallel sets of planes, the (111) and (222) planes, are equal within the experimental error, as they should be. Sets of planes that are not parallel to each other show different strains because copper is elastically anisotropic. The strain measurements were made on the Neutron Powder Diffractometer at LANSCE.

tions occur on the scale of interatomic distances—that is, on a much smaller scale than the 100 to 1000 angstroms covered by small-angle neutron scattering and neutron reflectometry—and because the materials of interest are polycrystalline, powder diffraction is the technique of choice.

Figure 14a shows the arrangement for measuring strains in a stressed copper bar with a neutron powder diffrac-

tometer. The intersection of the incident and detected neutron beams defines the so-called gauge volume, the volume of material within which the average interplanar spacings are measured. By using suitable neutron-absorbing masks the gauge volume can be varied. A large gauge volume yields a high neutron count rate and thus high accuracy in the measurements, but it also yields an average value of the strain over that

volume, an average that will not reflect the true strain in regions of steeply changing stress. Typically, gauge volumes of a few tens of cubic millimeters provide a reasonable compromise between those two factors.

The intersection of the incident and detected beams also determines the direction in which strain is measured. For example, as shown Fig. 14a, the detectors at +90 degrees and −90 degrees

1 inch
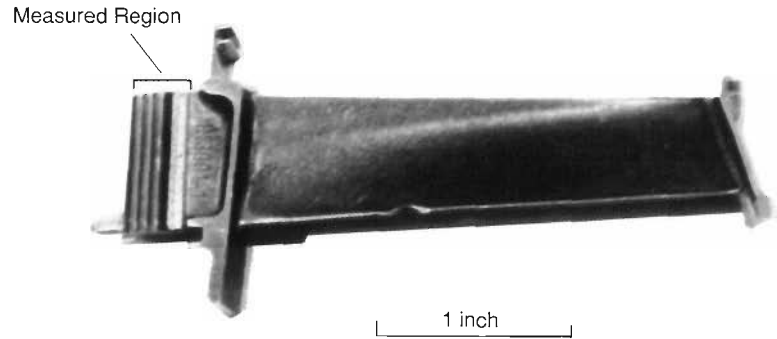
measure scattering from planes at +45 degrees and —45 degrees. If the bar is oriented as in Fig. 14, these directions correspond to lattice spacings parallel and perpendicular to the stress applied along the length of the bar. Because the copper bar is polycrystalline, there are crystallites in which a given set of lattice planes is oriented parallel to the applied stress and others in which the same set of planes is oriented perpendicular to the stress. In general, one would expect the lattice spacings to increase in a direction parallel to the applied load and to decrease in the perpendicular direction. Figure 15a shows a Bragg peak shift resulting from stretching the copper bar by 1 percent, and Fig. 15b lists the changes in lattice spacing for several sets of lattice planes. As expected, the lattice spacings parallel to the stress increase while the lattice spacings perpendicular to the stress decrease. The change in lattice spacing— 0.1 percent— is much less than the macroscopic change because the bar has been plastically deformed. Neutron scattering measures only elastic strain, which is much smaller. These experiments were performed on the LANSCE Neutron Powder Diffractometer, which can measure lattice spacings with an accuracy of 1 part in 100,000 or better.

On a powder diffractometer at a reactor source it is usual (although not mandatory) to work with fixed neutron wavelength and to record different Bragg peaks by changing the position of the detector (see "Neutron Scattering—A Primer"). However, the gauge volume changes shape slightly as the detector position is changed, and correcting for such effects is not always easy, especially in cases where strain varies rapidly with position in the specimen. At a pulsed spallation source, on the other hand, researchers work with neutrons of many different wavelengths simultaneously and record several Bragg peaks in a sin-
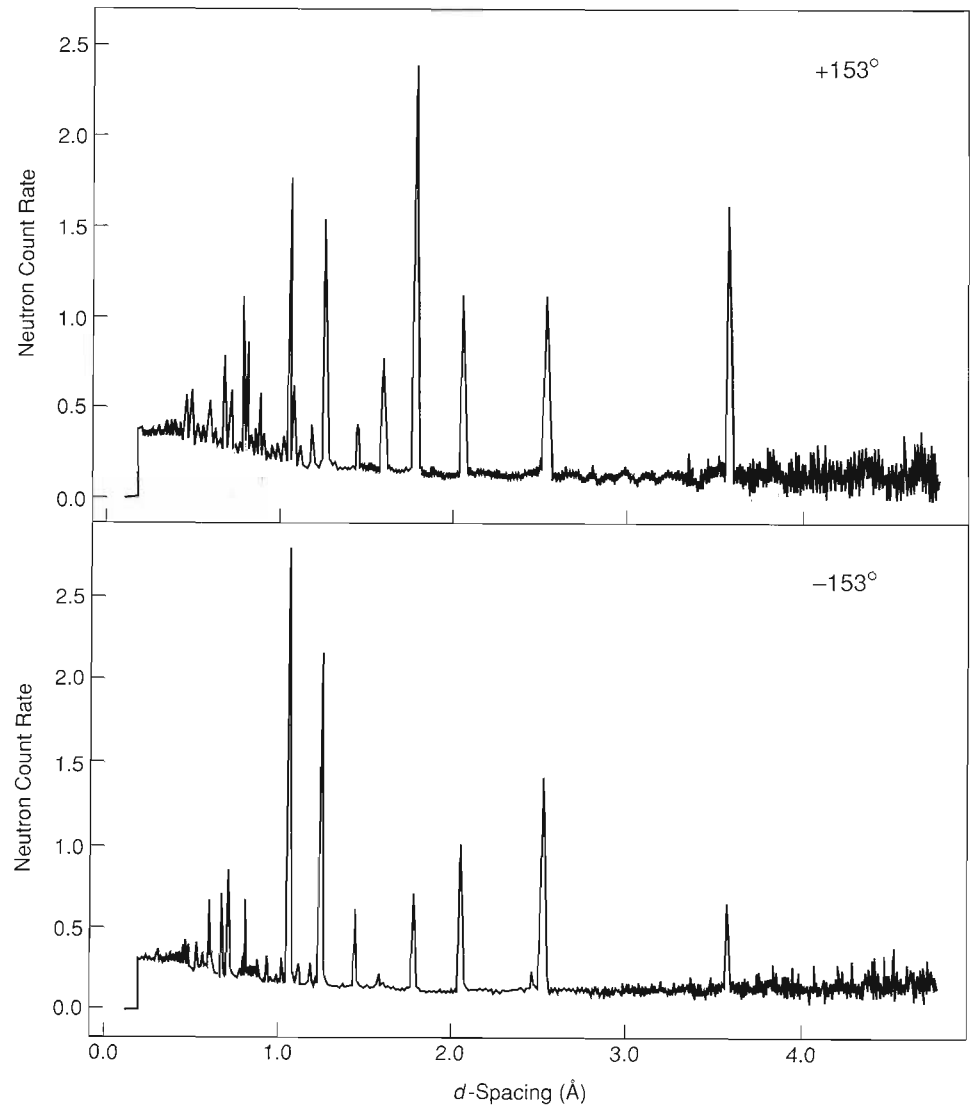
PREFERRED CRYSTALLITE ORIENTATION IN A TURBINE BLADE

Fig. 16. Bragg peaks measured at +153 degrees and —153 degrees from the incident beam of the High-Intensity Powder Diffractometer at LANSCE indicate a high degree of preferred crystallite orientation in the mounting end of a jet-engine turbine blade. If the crystalline grains were oriented randomly in the blade, the intensities of the corresponding peaks in the two patterns would be identical because equal numbers of grains would contribute to each pattern. Neutron scattering provides a simple method of checking preferred crystallite orientation.

gle detector. Because the scattering angle is constant during a measurement, the shape and size of the gauge volume are also constant. Furthermore, the simultaneous measurement of several Bragg peaks from parallel sets of atomic planes—for example, the (111) and (222) planes in Fig. 14—must correspond to the same microscopic strain if the measurement has been performed accurately. (The data in Fig. 14 clearly satisfy this consistency check.) Once this check has been made it is easier to be confident that the strains measured for nonparallel planes—for example, the (111) and (200) planes—are actually different, and that this difference is due to elastic anisotropy of the sample rather than an experimental mistake.

The Neutron Powder Diffractometer at LANSCE uses several detectors positioned symmetrically about the incident beam direction, thus allowing researchers to measure strain in several directions simultaneously. Also, by measuring a Bragg peak from the same set of planes but at several different scattering angles, researchers can determine whether the sample is textured, that is, whether the individual crystallites that make up the specimen have a preferred orientation. The data in Fig. 16 reflects the preferred orientation of the crystalline grains in a jet-engine turbine blade. The same Bragg peaks measured on the High-Intensity Neutron Powder Diffractometer at +153 and −153 degrees ought to have identical intensities if the crystallites in the sample are randomly oriented. Instead results for the turbine blade show different intensities for each of the Bragg peaks. This occurs because turbine blades are deliberately fabricated so that all crystalline grains have the same orientation with respect to the length of blade. This direction is chosen to enhance the blade's strength and resistance to permanent deformation at high temperatures. From the variation of the intensity of

the Bragg peaks with scattering angle, one can determine the degree of crystallite orientation.

Until fairly recently experiments to determine residual stress by neutron scattering have been demonstrations only. For example, researchers have measured the stress distributions in thick-walled steel tubes, compared their results with those produced by more conventional methods of measurement, and found good agreement. In the past year or two, however, the technique has been "used in anger" for a number of measurements, including residual stress resulting from the deformation of welds, residual stress in railway lines, and residual-stress fields at the tips of propagating cracks. Future plans for work at LANSCE include measurement of residual stress in ceramics and composites, studies of stresses close to shot-peened and bearing surfaces, investigations of metals formed by compressing powders, and measurements of strain distributions around welds that have been vibrated during the welding process. The application of neutron diffraction to the determination of residual stress is a technique in its infancy and one that holds great promise for the future. ∎

Acknowledgments

Further Reading

Sow-Hsin Chen and Tsang-Lang Lin. 1987. Colloidal solutions. In *Methods of Experimental Physics*, volume 23, part B, edited by Kurt Sköld and David L. Price, pp. 489–543. Orlando, Florida: Academic Press Inc.

Marc J. Ostro. Liposomes. *Scientific American*, January 1987, 102–111.

R. Hjelm, T. Hiyagarajan, D. S. Sivia, P. Lindner, H. Alkan, and D. Schwahn. 1990. Small-angle neutron scattering from aqueous mixed colloids of lecithin and bile salt. Accepted for publication in *Progress in Colloid and Polymer Science*.

R. H. Ottewill. 1987. Characterization of polymer colloids. In *Future Directions in Polymer Colloids*, edited by M. S. El-Aasser and R. M. Fitch, pp. 253–275. NATO ASI Series E, volume 138. New York: Plenum Press.

G. D. Wignall. Neutron scattering. 1986. In *Encyclopedia of Polymer Science and Technology*, 2nd Edition, volume 10, pp. 112–184, Jaqueline I. Kroschwitz, executive editor. New York: John Wiley and Sons.

Benoit B. Mandelbrot. 1977. *The Fractal Geometry of Nature*. New York: W. H. Freeman and Company.

Richard F. Voss. 1985. Random fractals: Characterization and measurement. In *Scaling Phenomena in Disordered Systems*, edited by Roger Pynn and Arne Skjeltorp, pp. 1–11. NATO ASI Series B, volume 138. New York: Plenum Press.

Dale W. Schaefer. 1989. Polymers, fractals, and ceramic materials. *Science* 243: 1023–1027.

D. W. Schaefer, B. C. Bunker, and J. P. Wilcoxon. 1989. Fractals and phase separation. *Proceedings of the Royal Society of London* A423: 35–53.

Roger Pynn and Tormod Riste, editors. 1987. *Time Dependent Effects in Disordered Materials*. NATO ASI Series B, volume 167. New York: Plenum Press.

Paul Smith. 1989. Some industrial applications of soft condensed matter in high-performance polymer processing. In *Phase Transitions in Soft Condensed Matter*, edited by Tormod Riste and David Sherrington. NATO ASI Series B, volume 211. New York: Plenum Press.

A. Stacey, H. J. MacGillivary, G. A. Webster, P. J. Webster, and K. R. A. Ziebeck. 1985. Measurement of residual stresses by neutron diffraction. *Journal of Strain Analysis* 20: 93–100.

G. A. Webster. 1989. Propagation of fatigue cracks through residual stress fields. In *Fatigue and Stress*, edited by H. P. Lieurade. Gournay-sur-Marne, France: IITT International.

D. J. Smith, R. H. Leggatt, G. A. Webster, H. J. MacGillivary, P. J. Webster, and G. Mills. 1988. Neutron diffraction measurements of residual stress and plastic deformation in an aluminum alloy. *Journal of Strain Analysis* 23: 201–211.

Samuel A. Safran and Noel A. Clark, editors. 1987. *Physics of Complex and Supermolecular Fluids*. New York: Wiley-Interscience.
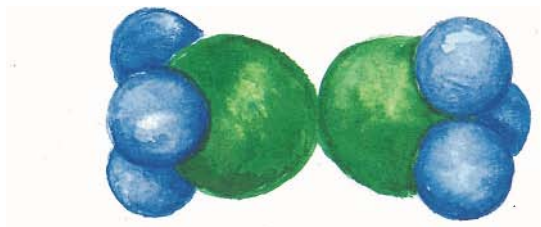
Chemical reactions catalyzed on the surface of a metal (red) include the dissociation of hydrogen molecules (blue) and the hydrogenation of C_2H_4 (ethylene) to C_2H_6 (ethane). An intermediate in the hydrogenation, C_2H_3, is also shown. The carbon atoms are green.
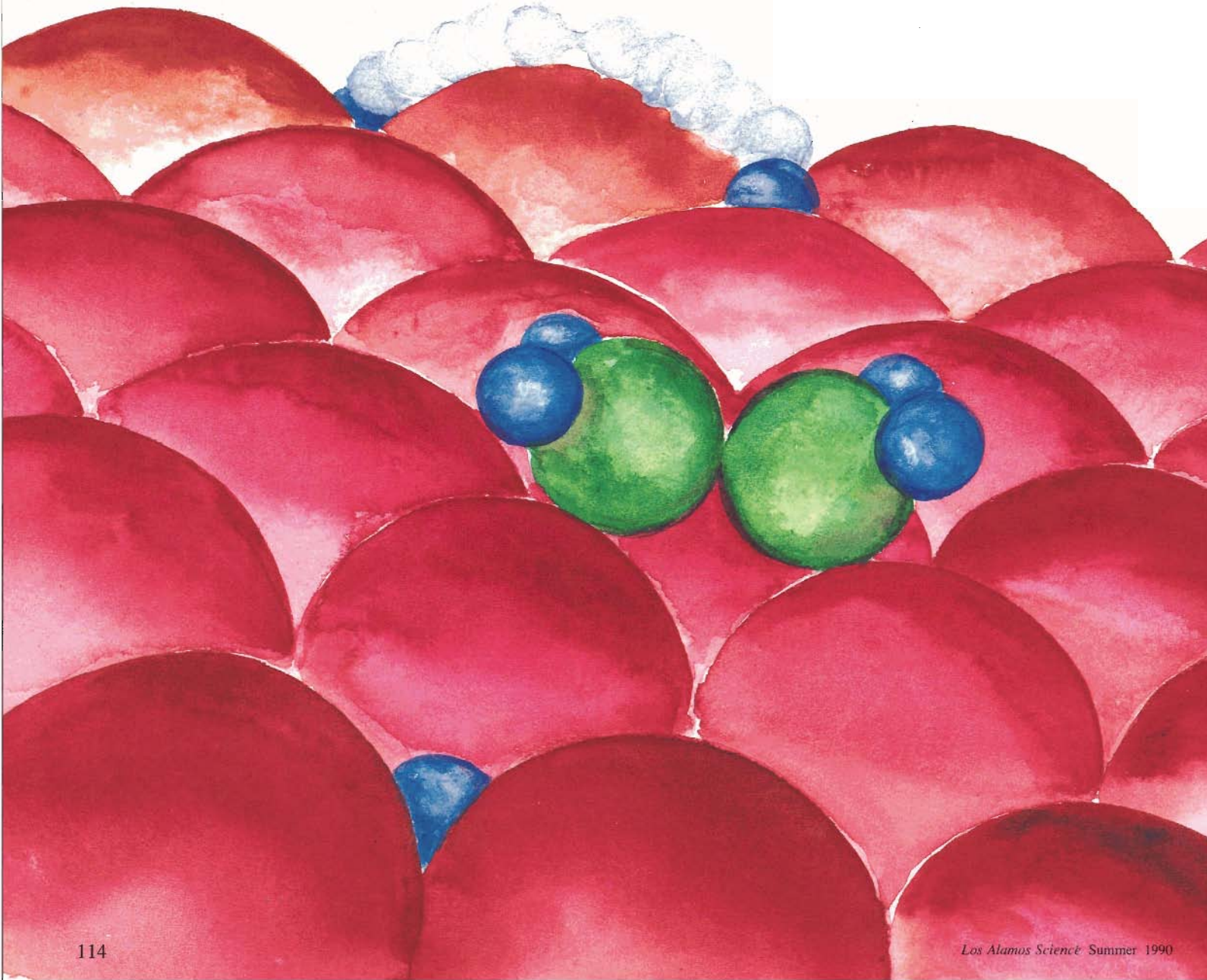
NEUTRONS & CATALYSIS

by Juergen Eckert and Phillip J. Vergamini

Catalysis—the ability of some substances to alter the rate of chemical reactions without being consumed—was recognized more than 150 years ago and has been applied on an industrial scale since the beginning of this century. Modern industrial chemistry, especially petroleum processing and the manufac-

ture of chemicals, could not function without catalysts. It has been estimated, for example, that catalysts are involved at some point in the production of 60 to 70 percent of all industrial chemicals. Yet the store of knowledge about how catalysts work is surprisingly small. The search for a catalyst for a particular reaction, or for ways to improve existing

catalysts, is still, as it always has been, largely empirical.

In the last few years, however, sophisticated new analytical and computational techniques have helped stimulate a renaissance in catalysis research. Powerful economic forces have motivated the study of catalysis as well: the need for new sources of energy and
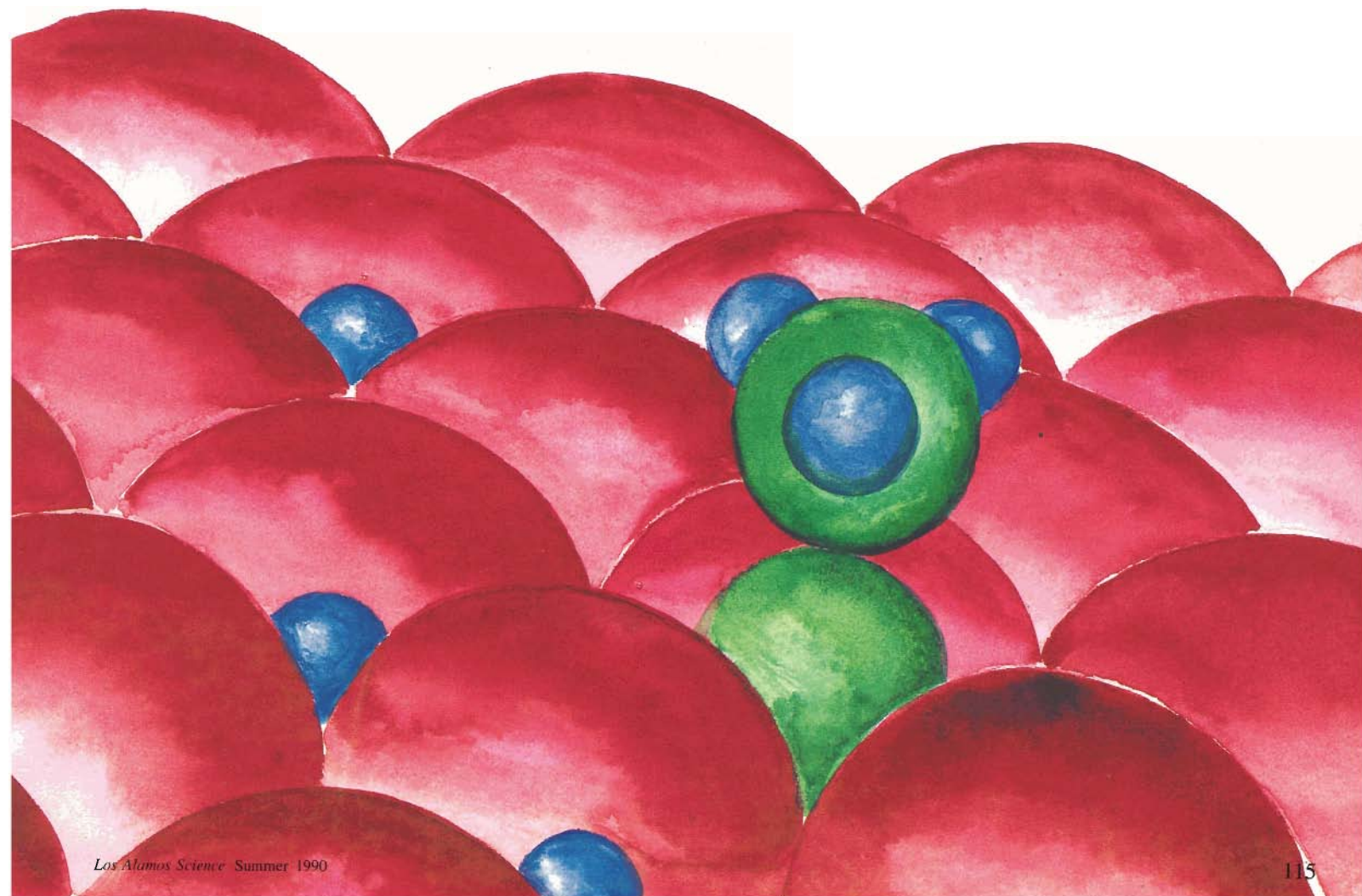
chemicals, changes in the availability of raw materials, potential restrictions on the availability of noble-metal catalysts, and the desire for new products have pointed up the need for a clearer understanding of catalytic processes.

Much of the research into catalysis is directed toward metals, because they catalyze many important reactions. Metals may be catalytically active in the form of finely divided particles, organometallic compounds in solution, or ions bound to large biologically active molecules, such as enzymes. The catalysis may be *heterogeneous* in the sense of involving more than one phase (solid metal and gaseous reactants, for example) or it may be *homogeneous* in the sense of involving only one phase (such as a solution). Whatever the form, when the metal binds to a reactant molecule, it almost always alters the chemical bonding in the reactant. If that alteration is favorable to some particular reaction, then the metal is a catalyst for that reaction.

To understand catalytic activity, or to tailor a catalyst to do a specific job, we need to know the individual steps in the catalytic process in great detail. For example, consider the hydrogenation of ethylene to ethane,

$$H_2C=CH_2 + H_2 \rightarrow H_3C-CH_3.$$
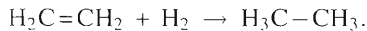
which can serve as a prototype of reactions used in producing synthetic fuels. The production of synthetic fuel from coal, for example, involves various series of reactions, including the stepwise hydrogenation of carbon to form acetylene (HCCH), ethylene, and ethane, as well as the stepwise hydrogenation of carbon chains with more than two carbon atoms. The hydrogenation of ethylene shown above is a particularly useful reaction to study because it can be carried out at moderate temperatures in the presence of a metal catalyst. The various steps to the reaction are represented schematically in Fig. 1, and a full understanding of the hydrogenation process requires knowing many details about each step. Knowledge of the spatial relationships of the adsorbed species and the metal atoms at the catalyst surface may enable us to identify reactive sites on the surface. Determination of the changes in bond angles
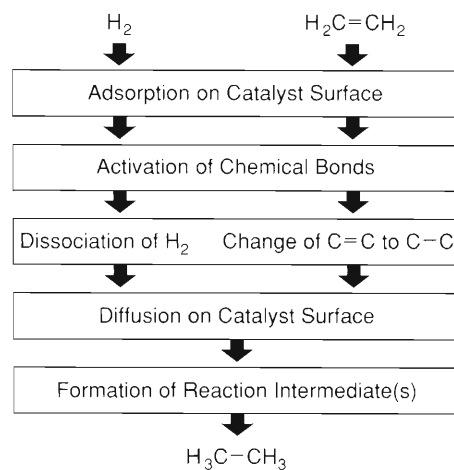
HYDROGENATION OF ETHYLENE

Fig. 1. The catalytic hydrogenation of ethylene to ethane involves various steps, all of which need to be examined in detail if we are to fully understand the reaction. Neutron scattering experiments can help provide some of the necessary detail.

and distances of the reactant molecules when they are adsorbed should make it possible to understand theoretically the changes in electronic structure that occur when the reactants are activated. Because the adsorbed species must diffuse on the surface to react and form new molecules, we need to know how this occurs. (We might wonder, for example, whether the adsorbed hydrogen diffuses only over the surface of the metal or also into its interior.) Finally, identification of reaction intermediates is crucial to understanding the entire process.

Unfortunately, these details of structure and dynamics cannot easily be determined in a "real-world" situation— that is, during an actual catalytic reaction. Catalytic processes usually proceed under conditions that preclude the direct application of many powerful analytical techniques—or at least make such application very difficult. Considerable effort has therefore been devoted to the study of so-called *model systems*, which are designed to reproduce the critical relationships as accurately as possible. One useful model system is a single crystal of a metal for which the surface arrangement of atoms is known. Others that have been widely used are synthetic molecules consisting of a metal atom (or a cluster of metal atoms) surrounded by stabilizing ligands, usually carbonyl groups (CO) or other more complex organic groups. When a reactant molecule such as ethylene or benzene binds to such a synthetic molecule, we can assume that, to some degree, the configuration of the resulting complex resembles that of the same reactant adsorbed on a metal surface. The complex can be studied with several spectroscopic techniques, and its crystalline form can be characterized by x-ray and neutron diffraction, which reveal details of its architecture with great accuracy.

The more closely the properties of the model system approximate the properties of the real-world system, of course, the better. As a result, model systems are often structurally modified to refine their properties and bring them closer in line with the more complex system of interest. However, such modifications can complicate the structural characterization of the model system. For example, as the model system becomes larger and more complex, the chances increase that some portions of the molecule will be disordered or less easily defined. The necessity of modeling the disorder can decrease the precision of the results for

the metal-hydrogen interaction, which is the feature of most interest. In effect, the results become slightly fuzzy and less precise.

Besides being useful in the study of catalysis, metal complexes are highly suitable for theoretical studies of chemical bonding between the bound molecule (ligand) and the metal atoms. They are therefore of fundamental interest to researchers studying chemical bonds from first principles. Finally, metal-cluster complexes can stabilize certain molecules that are unstable in pure form. For example, cyclobutadiene can be stabilized by binding to iron carbonyl, $Fe(CO)_3$; and ethylidyne, CH_3–C (a highly reactive intermediate formed in the hydrogenation of ethylene), can be isolated by reacting with cobalt carbonyl to form the metal-cluster complex CH_3C–$Co_3(CO)_9$.

The kind of information available through the study of model compounds is illustrated by the case of the cluster compound $HFeCo_3(CO)_{12}$. Diffraction studies show that the single hydrogen atom is located at a site of threefold symmetry, that is, just outside the triangle formed by the three cobalt atoms (Fig. 2). The vibrational spectrum of hydrogen in this compound is very similar to that of hydrogen atoms chemisorbed on a nickel or a platinum surface. Since the vibrational spectrum of a molecule or atom strongly reflects the way in which it is bound to other atoms, the similarity here allows the inference that hydrogen chemisorbed on a catalyst surface is located at a site of threefold symmetry. We can further infer that the catalytically active surface is the so-called (111) plane of the metal, because that is the only crystal plane having threefold symmetry. This information could not have been easily obtained in any direct way.

How does one then study the model systems? There are many experimental techniques, each especially suited

for a particular aspect of the problem, and neutron scattering is one of these. However, even the most intense neutron sources produce fluxes far below those commonly available from sources of photons (x rays, ultraviolet, visible light, and infrared), and so neutron scattering is not one of the principal tools of surface science. Nevertheless, when the systems include hydrogen
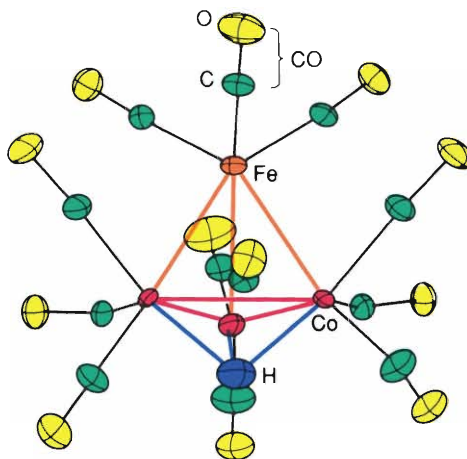
A CLUSTER COMPOUND

Fig. 2. The $HFeCo_3(CO)_{12}$ complex, which contains a single hydrogen atom (blue) located against an equilateral triangle of cobalt atoms (red), can serve as a model system for hydrogen atoms chemisorbed on a metal surface. In particular, comparison of vibrational spectra can help establish whether or not the hydrogen on the metal surface is also located at sites with threefold symmetry. (Adapted from a figure in an article by R. G. Teller, R. D. Wilson, R. K. McMullan, T. F. Koetzle, and R. Bau. *Journal of the American Chemical Society* 100: 3071, 1978.)

or molecules containing hydrogen—as do the more important types of compounds involved in industrial catalytic processes—neutron scattering is extremely useful.

The singular utility of neutron scatter-

ing is in locating the all-important hydrogen atoms and highlighting the vibrational and rotational motions associated with them. This strength is a result of the fact that neutrons scatter as strongly from hydrogen as from most other elements (see "Neutron Scattering—A Primer" by Roger Pynn). Although it is nearly impossible to "see" hydrogen atoms in the presence of heavy metals using x rays, x-ray diffraction can sometimes *implicitly* locate hydrogen atoms bound to or interacting with metal atoms. If a site in a metal complex is usually filled, an apparent vacancy at that site, together with other physical and chemical evidence, can lead to the inference that hydrogen occupies the position. Neutron scattering, however, is needed to confirm the actual presence of hydrogen. Thus, the structures of compounds of interest are typically determined by first applying x-ray diffraction to locate the heavier atoms and then neutron diffraction to obtain precise metal-hydrogen distances and bond angles.

Historically, single-crystal neutron diffraction has been more difficult than x-ray diffraction. Neutrons can travel large distances through material without being scattered, so neutron diffraction requires a much larger crystal. This problem has been partly alleviated by the availability of more intense sources of neutrons. Furthermore, the time-of-flight wavelength measurements possible at pulsed-accelerator-based neutron sources makes all neutrons in each pulse usable. Area detectors make it possible to collect large volumes of data at one time and make feasible full structural determination from polycrystalline material.

For the observation of molecular vibrations, optical techniques (infrared absorption and Raman scattering) are far more common and much easier to use than neutron scattering. Once again, however, the difference in the nature of

the interaction between the scatterer and the probe makes neutron-scattering vibrational spectroscopy advantageous in certain cases. First, absorption of photons in optical spectroscopy depends on the symmetry properties of the vibrational mode being excited in the sample, whereas no such symmetry-based selection rules apply to *inelastic* neutron scattering. (We use inelastic to refer to the fact that the neutron loses or gains energy during the scattering process. The change in energy corresponds to a change in the vibrational energy of the scattering molecule.) Hence, in principle (though not necessarily in fact), all vibrations of a molecule can be observed by inelastic neutron scattering.

The factors determining the intensity of a given excitation are a second, and perhaps more important, difference between neutron-scattering and optical vibrational spectroscopy. Large-amplitude vibrations by nuclei with high neutron-scattering cross sections (such as hydrogen) typically give rise to intense inelastic neutron-scattering bands; whole-body librations of molecules are typical examples. Such motions, however, usually involve only small changes (if any) in the polarizability or the dipole moment of the molecule, which are the factors that govern intensities in Raman scattering and infrared absorption. Thus, optical and neutron-scattering methods are remarkably complementary.

The utility of inelastic neutron scattering can be greatly enhanced by replacing certain atoms, whose vibrations are to be highlighted, with isotopes of different neutron-scattering cross sections. Such isotopic substitution is particularly valuable for studying hydrogen, because the neutron-scattering cross sections appropriate to inelastic neutron scattering for hydrogen and deuterium differ by more than an order of magnitude. For example, to distinguish the motions of the methyl group in toluene ($C_6H_5CH_3$, a benzene derivative in which one of

the ring hydrogens is replaced by a methyl group), the remaining ring hydrogens can be replaced by deuterium atoms. Then, as far as neutron scattering is concerned, the deuterium atoms are much less "visible" than the three hydrogen atoms on the methyl group. (In optical spectroscopy, isotopic substitution alters the frequency of vibration but leaves the intensity of absorbed or scattered photons virtually unchanged.) Another application of isotopic substitution in neutron scattering involves the differential spectra of isotopic species, examples of which will be described in the following sections.

We have chosen to describe three examples of neutron-scattering studies on metal complexes, each of which may serve as a model system for a particular step in the hydrogenation reaction shown in Fig. 1. The first example is a hydride ligand in an octahedral cluster of metal atoms, a model system that may help us understand the motion of hydrogen atoms between the surface and the region just below the surface (once the H_2 molecule has dissociated on the surface). The other two examples—an ethylene-diosmium complex and a set of molecular-hydrogen complexes—may be regarded as models for the bond activation that precedes the actual reaction on the surface. The complexes that bind molecular hydrogen are particularly important in this context because they represent a "capture" of the long-sought intermediate in perhaps the most fundamental reaction, the dissociation of hydrogen molecules. As we shall show below, elastic and inelastic neutron-scattering studies of these compounds have provided remarkably detailed information on the nature of the chemical bond between the dihydrogen, or molecular hydrogen, ligand and the metal center, including evidence for back-donation of electron density from the metal to the antibonding orbital of the hydrogen molecule.

The Hydride Ligand

The first example we want to discuss in detail is the interstitial hydride ligand—a single hydrogen atom bound to a metal atom or atoms. Hydride ligands are usually formed on metal surfaces when molecular hydrogen dissociates and are referred to as terminal, doubly bridging, triply bridging, and so forth, depending on whether they are bonded to one, two, three, or more metal atoms.

In large cluster complexes with many metal atoms, hydride ligands may also occupy interior, or interstitial, sites. Among the large metal-cluster compounds of this type that have been synthesized, two—the octahedral clusters of cobalt and ruthenium—stand out for their remarkable simplicity. Both these compounds have a cluster of six metal atoms that form the octahedral site, a single hydride ligand, and several carbonyl groups outside the metal cluster that serve to stabilize the molecule.

This kind of hydride coordination looks very much like that observed in bulk metals, where interstitial sites of octahedral or tetrahedral symmetry may be occupied by hydrogen. A hydrogen atom in a metal is of course surrounded by many more metal atoms than the six of a cluster compound. There are six atoms as *nearest* neighbors in an octahedral site, but further shells of metal atoms occur at ever increasing distances. However, if the hydrogen atom is located just below the surface of the metal, that is, between the first two layers of metal atoms, the number of more distant neighbors is minimized in one direction. The interstitial hydrogen in the cluster may therefore be a better model for "subsurface" hydrogen than for hydrogen within the bulk of the metal. Such a system may help answer a question raised earlier in the discussion of hydrogenation—that is, where the hydrogen is likely to move,

(a) The [HCo$_6$(CO)$_{15}$]$^-$ Anion

(b) The Central Octahedron
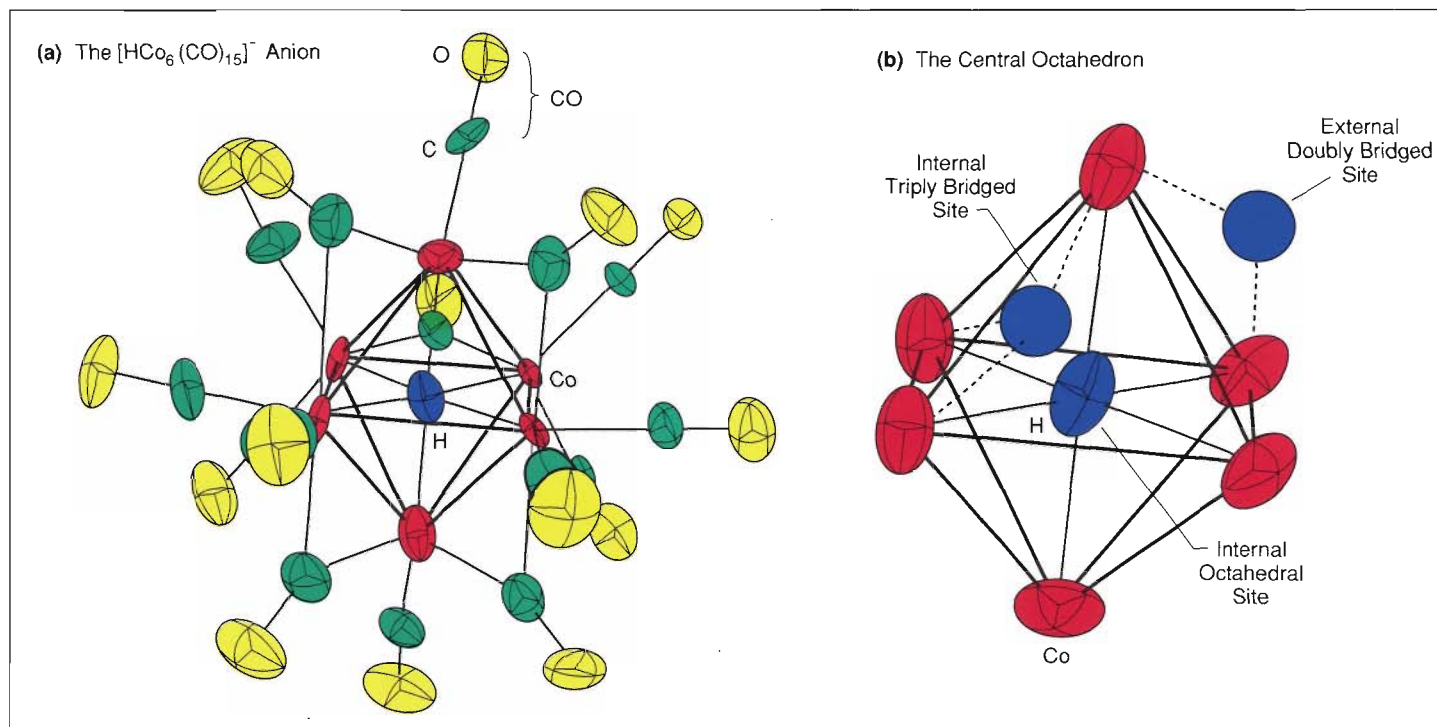
AN OCTAHEDRAL CLUSTER

Fig. 3. (a) The location of the hydride ligand in the anion [HCo$_6$(CO)$_{15}$]$^-$ has been determined with considerable certainty by neutron diffraction from a single crystal containing the cation [(Ph$_3$P)$_2$N]$^+$. The hydride ligand (blue) is located at the center of an octahedron of cobalt atoms (red); the cobalt atoms, in turn, are surrounded by twelve carbonyl ligands (green and yellow). The shapes at each atomic position are *thermal ellipsoids*, which indicate the extent and direction of the thermal motion of the atoms about their equilibrium positions. (The surface of each ellipsoid defines the volume in which the atom is contained 50 percent of the time.) (b) There are a number of alternative sites close to the central octahedron of cobalt atoms that may, in some compounds, serve as the location of the hydride ligand (two possibilities are shown in light blue). The alternative sites are either external or internal to the octahedron of cobalt atoms, and the hydrogen atom can be doubly or triply bridged to those atoms. However, such sites have considerably less symmetry than the central octahedral site, and the vibrational spectra of hydrogen when located at such a site would be quite different from the single excitation that is actually observed in the spectrum of [HCo$_6$(CO)$_{15}$]$^-$. (Adapted from a figure in an article by D. W. Hart, R. G. Teller, C.-Y. Wei, R. Bau, G. Longoni, S. Campanella, P. Chini, and T. F. Koetzle. *Journal of the American Chemical Society* 103: 1458, 1981.)

after dissociation, relative to the metal catalyst's surface.

The first step, however, in understanding the microscopic properties of hydrogen in a metal cluster is to attempt to locate it by diffraction studies. The only reliable way to do this is by use of neutron beams, for the reasons discussed earlier. Because hydrogen is a minor component of the rather large metal-cluster molecules, single crystals

must be used for the diffraction studies. Also, cluster compounds are most commonly ionic species, and a suitable *counter ion*—a large, oppositely charged ion—must be added to produce sufficiently large single crystals. In the present case, the complex of interest is [HCo$_6$(CO)$_{15}$]$^-$, and the counter ion used to produce the single crystal is [(Ph$_3$P)$_2$N]$^+$, in which Ph is the phenyl group, C$_6$H$_5$.

A neutron-diffraction study of a single crystal of [(Ph$_3$P)$_2$N][HCo$_6$(CO)$_{15}$], carried out on the high-flux-beam reactor at Brookhaven National Laboratory, showed quite convincingly that the hydrogen is located approximately at the center of a somewhat irregular octahedron of cobalt atoms (Fig. 3) even though other locations, such as threefold coordination inside or outside one of the triangular faces of the octahedron,

are possible. The vibrational spectrum of hydrogen in a regular octahedral site would show a single excitation, a triply degenerate hydrogen-metal stretching mode. If, however, the hydrogen were to also move significantly off center, additional peaks would appear in the vibrational spectrum. It is in just such cases that vibrational spectra are of great value in obtaining structural information. An inelastic neutron-scattering study of the cesium salt of the same cobalt cluster showed primarily a single excitation at a frequency of 1056 reciprocal centimeters (cm^{-1}), confirming the central location of the hydride ligand in the octahedral site.

Prompted by the results of some infrared spectroscopic studies that showed interesting changes in the spectra of the $[HCo_6(CO)_{15}]^-$ cluster as the crystalline environment was altered, we recently investigated the vibrational spectrum of the cluster combined with the very much smaller counter ion K^+. The data, shown in Fig. 4, were obtained by the differential technique on the Filter Difference Spectrometer at the Manuel Lujan, Jr. Neutron Scattering Center (LANSCE) at Los Alamos. Two samples were measured, one with hydrogen and one with deuterium as the ligand. Because the scattering cross section for deuterium is much smaller than that for hydrogen, the vibrational spectrum of the deuterated compound serves essentially as a "blank" to be subtracted from that of the protonated compound. The resulting differential spectrum is free of all the many vibrational modes of this large molecule that do not involve motion of the hydrogen and thus highlights the vibrational modes that do involve hydrogen.

The features shown in the differential spectrum can be immediately identified with hydrogen vibrations and suggest that all the hydrogen atoms are by no means located at the center of the octahedron of cobalt atoms. The broad
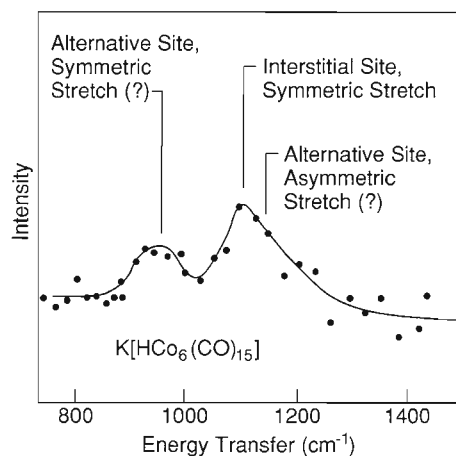
Fig. 4. The differential spectrum shown here for the $[HCo_6(CO)_{15}]^-$ cluster (Fig. 3)—only crystallized with the much smaller counter ion K^+—includes only those vibrations that involve the hydride ligand. The broad peak just below 1100 cm^{-1} has been identified as the symmetric stretching vibration of hydrogen located at the center of the interstitial octahedral site; this peak appears in various spectra regardless of which counter ion is present. The second peak at about 950 cm^{-1} is not present in the spectra of crystals containing Cs^+ as a counter ion and may be the stretching vibration of doubly or triply bridged hydrogen located at an alternative site (such as those shown in light blue in Fig. 3b). If the latter assignment is correct, the high-frequency shoulder just above 1100 cm^{-1} would correspond to the asymmetric stretching vibration of hydrogen at a triply bridged site. These data were obtained by using the Filter Difference Spectrometer at LANSCE.

band in the region between 1050 and 1100 cm^{-1} may certainly be assigned to the stretching vibration of hydrogen at the interstitial site, but the band at 950 cm^{-1} must then be indicative of hydrogen at a different site—one bridging either two or three cobalt atoms. In

either case, a second vibrational line at higher frequency would be expected. The data are not conclusive in this respect, but if the band at 950 cm^{-1} is the symmetric stretching vibration for the doubly or triply bridged hydrogen, then the high-frequency shoulder just above 1100 cm^{-1} has about the expected frequency for the asymmetric stretch of triply bridged hydrogen.

The spectrum thus appears to reveal an instance of the *fluxionality* of the hydride ligands in cluster compounds. Fluxionality—commonly detected in nuclear-magnetic-resonance studies—refers to the movement of hydrogen from one site to another. Because the movement occurs on a time scale that is many orders of magnitude greater than the time scale of a typical vibration, the hydrogen can be "caught" vibrating rapidly at more than one site. However, if the binding energy is much larger at one site than at others, such fluxionality is unlikely.

In any case, the remarkable result of our studies is that the position of the hydride ligand in these metal clusters apparently depends on the nature of the counter ion used to crystallize the compound. This fact suggests that the binding strengths for hydrogen at the various sites differ by only small amounts and may, in fact, be affected by the charge balance between the complex ion and its counter ion. Such a conjecture is needed to explain the observed change in fluxionality of the hydrogen atom in the cluster. Moreover, the conjecture is in agreement with nuclear-magnetic-resonance observations of the $[HCo_6(CO)_{15}]^-$ ion in solution, which show that the hydrogen can easily leave the octahedron and exchange with protons of the solvent molecules.

The factors that govern fluxionality of the hydride ligand in cluster compounds may, of course, differ considerably from those that determine the diffusion of hydrogen between the metal surface and,

VIBRATIONAL SPECTRUM OF THE OCTAHEDRAL CLUSTER
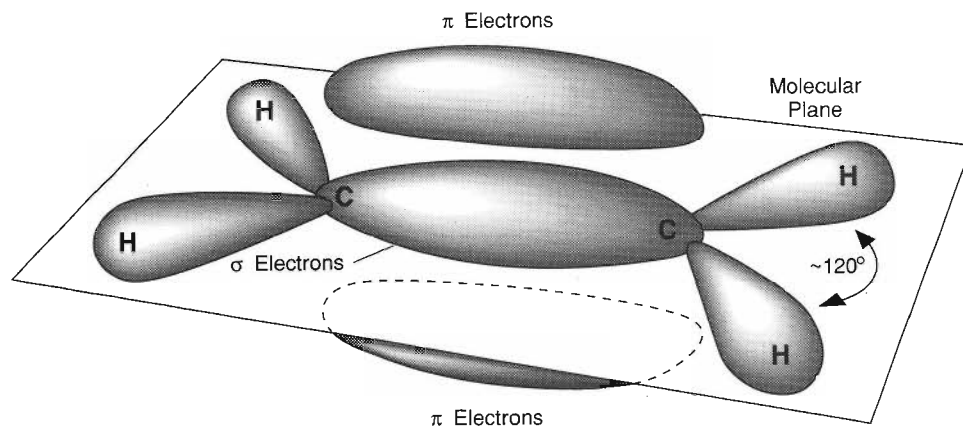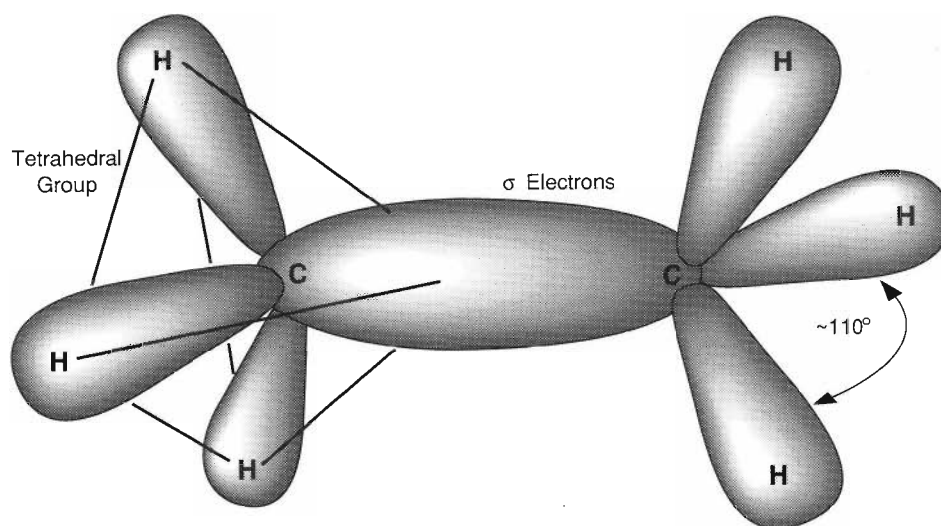
(a) Ethylene

(b) Ethane

GEOMETRY OF ETHANE AND ETHYLENE

Fig. 5. (a) All the atoms of ethylene lie in a single plane, and almost all of its electron orbitals are concentrated close to that plane. The exception is the double bond between the two carbon atoms, which includes two π electrons in an orbital that juts above and below the plane, as well as two σ electrons in the central bonding orbital in the plane. (b) The atoms of ethane, on the other hand, are arranged in two overlapping tetrahedral groups that surround each carbon atom (one tetrahedron is outlined in red), and it has a single σ bond between the carbon atoms. This arrangement allows the two CH_3 groups to rotate with respect to one another.

for example, the subsurface layer. If one wishes to hydrogenate, say, ethylene on a metal surface, it is important to know that the hydrogen atoms will stay on the surface long enough after dissociation to react with the ethylene rather than diffuse rapidly into the bulk metal. Despite the above caveats, the detailed structural picture, and in some cases the detailed dynamical picture of hydride motion, that can be obtained from model cluster systems is quite important for understanding the diffusion of hydrogen along a metal surface or between the surface and the bulk.

The Ethylene-Metal Complex

We now focus on another question in the hydrogenation reaction, namely the formation of a complex between ethylene and the metal and the resulting bond activation that is necessary for ethylene to take up hydrogen. Hydrogenation changes the ethylene molecule from an arrangement in which all the atoms are coplanar and the carbon-carbon bond is a short double bond that includes π-bonding electrons (Fig. 5) to an ethane molecule in which the atoms are grouped tetrahedrally and the carbon-carbon bond is the longer single bond based solely on σ electrons. Here again, we shall examine a model compound in which the precise arrangement of the ethylene molecule and the metal atoms can be studied by neutron diffraction.

Ethylene molecules can interact with a metal surface in several ways. Perhaps the most common configuration is the so-called side-on π complex shown in Fig. 6a, in which the planar ethylene molecule is parallel to the surface of the metal and the π electrons in the carbon-carbon double bond interact directly with the electrons on a single metal atom. A second species, the di-σ-bonded complex (Fig. 6b), has also been observed; in this case the double bond

between the carbon atoms is reduced to a single bond and a further σ bond is formed between each carbon atom and one of two adjacent metal atoms. Obviously, this last complex could be an intermediate species in the hydrogenation reaction.

The $Os_2(CO)_8(C_2H_4)$ complex (Fig. 7) does in fact have an ethylene ligand that shows the di-σ binding geometry. Further, the vibrational spectrum of this complex is very similar to that of ethylene chemisorbed on platinum at temperatures below 100 kelvins, and the complex can serve as a model for that system. X-ray diffraction studies of $Os_2(CO)_8(C_2H_4)$ show both that the carbon-carbon distance is longer than a normal ethylene double bond and that the four-membered ring formed by the two osmiums and two carbons is twisted and nonplanar. This last observation implies that the hydrogens have probably also twisted out of their plane with the carbons. However, as we already pointed out, it is very difficult to directly determine the positions of the hydrogen atoms with x rays. Spectroscopic evidence also suggests unusual structural features within the bridging ethylene ligand. This evidence may or may not be consistent with the x-ray observations but cannot be interpreted in an unambiguous way. Therefore, a knowledge of the detailed structure, particularly the positions of the hydrogen atoms, is necessary to resolve questions regarding the bonding in this compound.

Neutron-scattering measurements show that two hydrogens, an osmonium atom, and the other carbon atom are arranged approximately tetrahedrally around each ethylene carbon atom. This observation is consistent with the elongation of the carbon-carbon bond, as well as with the spectroscopic evidence and theoretical calculations that allow for simple σ bonding between each carbon atom and an osmium atom. In this example, then, the bonding of ethylene

(a) π-Bonded Ethylene
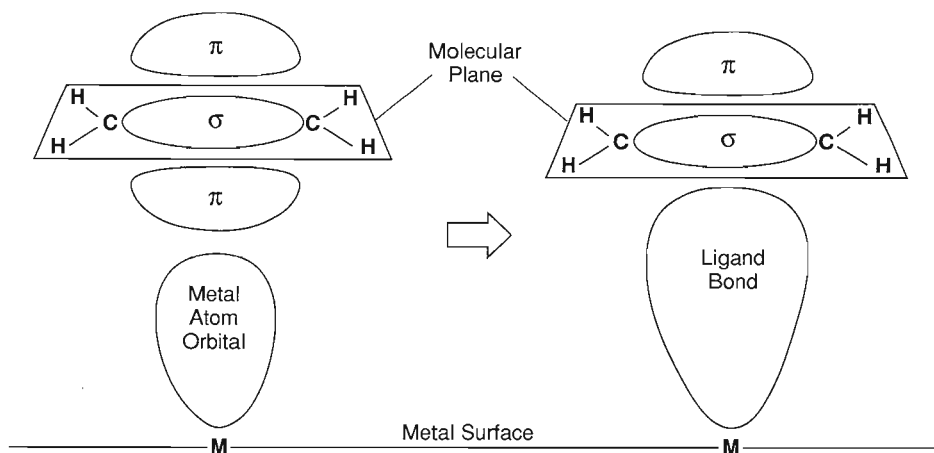
(b) Di-σ-Bonded Ethylene
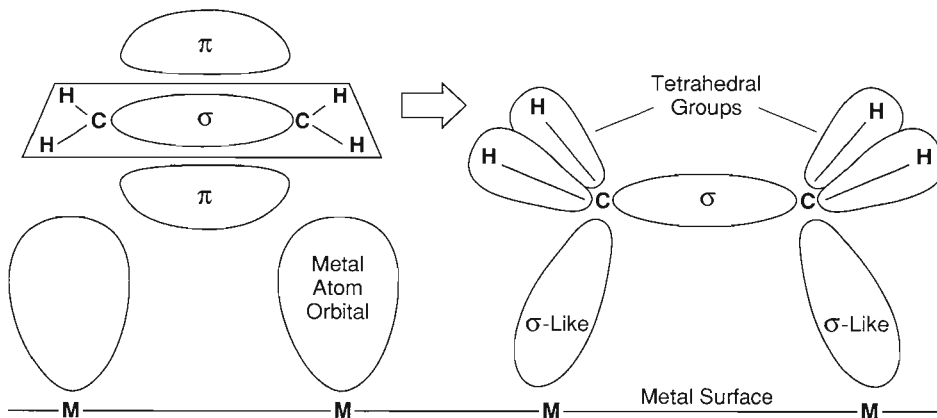
TWO KINDS OF ETHYLENE LIGANDS

Fig. 6. Ethylene can form two distinctly different types of ligand bonds with metal atoms. (a) The π-bonded type involves the formation of a symmetric sigma-like bond between a *single* metal atom and a π-bond orbital of the ethylene molecule, a configuration in which the planar nature of the ethylene molecule is retained. (b) The di-σ-bonded type involves the formation of *two* sigma-like bonds between the carbon atoms and two adjacent metal atoms. As a result, the carbon-carbon bond in the ethylene becomes a single σ bond and the hydrogen atoms move out of the molecular plane and assume an approximate tetrahedral arrangement about the carbon atoms.

to a metal-cluster compound results in rearrangement of the bonding electrons in the smaller molecule.

The planar configuration of the ethylene molecule is obviously drastically altered by its association with the os-

mium carbonyl cluster. Simply replacing the two osmium atoms with hydrogen atoms would produce the geometry found in ethane, the product of hydrogenation of ethylene. The twist observed in the cluster-bonded ethylene

is tending toward the bond angles normally found in ethane.

In this example, the osmium cluster can be considered a model of either an isolated fragment of a metal surface (as in heterogeneous catalysis) or an individual catalytic molecule (as in homogeneous catalysis). In either case, osmium is not necessarily unique in complexing with ethylene. Other metal atoms have different electrons at different energy levels, so the degree of activation and distortion may differ from one complex to another. However, all of them should have a tendency to activate ethylene to one degree or another by forming a complex of this kind.

Binding of Molecular Hydrogen to a Metal

In this final example we shall go back one step in the hydrogenation reaction and focus on the reaction of molecular hydrogen with a metal atom, the reaction that precedes its dissociation into hydrogen atoms. As mentioned earlier, most small molecules can bind chemically to complexes containing one or more metal atoms, often in ways that roughly resemble the chemisorbed state of the molecule. The coordinated molecule and the metal atom or atoms share electrons to some extent; as a result, some bond angles or bond lengths in the bound molecule are changed. Molecular hydrogen has always been a notable exception; until recently it was found to bind only dissociatively, that is, as two individual hydrogen atoms. Observation of chemically bound molecular hydrogen would offer enormous potential for understanding on the basis of first principles the process that eventually results in dissociative binding of hydrogen.

A few years ago, G. J. Kubas and collaborators isolated the tungsten complex $W(CO)_3(PCy_3)_2H_2$ (where Cy is cyclohexyl, a 6-carbon alkane ring) that
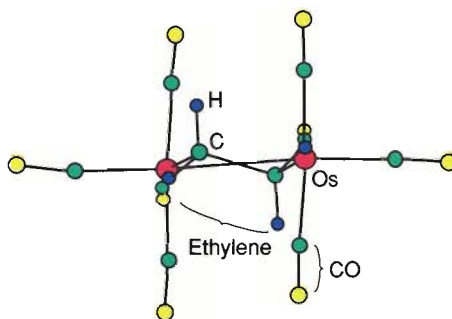
AN ETHYLENE-BRIDGED COMPLEX

Fig. 7. An example of the di-σ-bonded type of ethylene ligand (Fig. 6b) may be found in the osmium complex $Os_2(CO)_8(C_2H_4)$. Not only has the carbon-carbon bond in the ethylene ligand lengthened, but the ligand has twisted, allowing the hydrogen atoms (blue) and the two osmium atoms (red) to assume a more tetrahedral grouping about each central carbon atom (green).

may represent the long-sought intermediate in the oxidative addition of hydrogen to a metal. Since then, many additional molecular-hydrogen complexes with central metal atoms other than tungsten and ligands other than tricyclohexylphosphine have been identified. The hydrogen in these complexes is apparently reversibly bound to the metal, as can be demonstrated by passing hydrogen gas into a solution of, for example, the precursor of the above tungsten-tricyclohexylphosphine complex at room temperature. The solution, which is originally purple, turns yellow, and light yellow crystals of the H_2 complex can be precipitated from it. If the hydrogen stream is replaced by a chemically inert gas such as argon, the purple color returns, implying the dissociation of H_2 from the complex. An interesting feature of these compounds is that formation of a stable hydrogen complex apparently requires organophosphine ligands that are large and bulky. The structure of the purple precursor contains a clue to the role these ligands

may play: the P–W–P axis is distorted and the organic portion of one phosphine appears to fill the hole left by the absent molecule of hydrogen.

Neutron-scattering techniques have played a decisive role in characterizing the dihydrogen ligand of the complexes in terms of both its structural and dynamical properties. This information has then been used to work out a detailed quantitative picture of the bonding of the hydrogen molecule to the metal, as will be described in the following sections.

A Sigma-Bond Complex. In the early stages of the investigation of the complexes, it was absolutely essential to locate the dihydrogen ligand and ascertain whether, in fact, it retained its molecular identity. Although some initial evidence for the molecular-hydrogen binding came from x-ray diffraction, conclusive evidence required the use of neutron diffraction because of its sensitivity to scattering from hydrogen. The first structure determined for a molecular-hydrogen complex is shown in Fig. 8. The complex is the same as the one we have been discussing except the tricyclohexylphosphine ligands (PCy_3) have been replaced with less bulky triisopropylphosphine ligands ($P(i\text{-Pr})_3$, where i-Pr represents $CH(CH_3)_2$). Probably the most important features of this structure are the two equal W–H distances of 1.89 Å and the increase in the H–H distance by some 10 percent over that in free H_2 (0.82 Å versus 0.74 Å). These facts clearly suggest the formation of a three-center metal-dihydrogen bond (that is, some of the electrons are shared between the metal atom and the two hydrogen nuclei) and a substantially weakened H–H bond.

The dihydrogen ligand was also found to have a well-defined equilibrium orientation (Fig. 8c), one in which the H–H axis is parallel to the P–W–P axis of the complex. This fact might be ex-
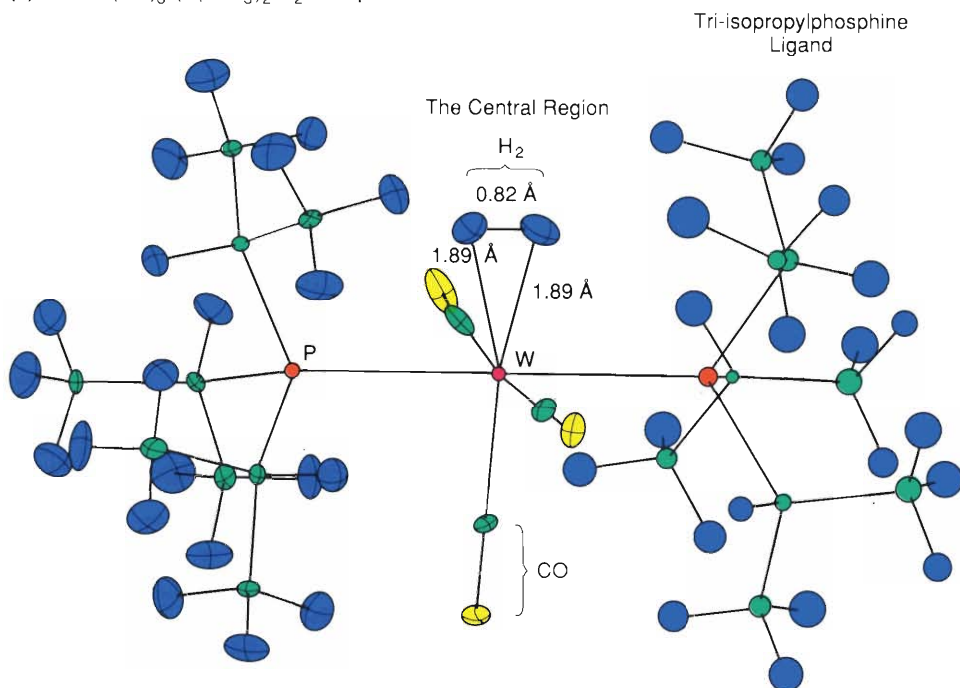
A DIHYDROGEN COMPLEX

Fig. 8. (a) W(CO)$_3$(P(i-Pr)$_3$)$_2$H$_2$, the first molecular-hydrogen complex to have its structure determined, has two bulky tri-isopropylphosphine ligands (orange, blue, and green) located on opposite sides of a tungsten atom (red). The central region between the opposing phosphorus atoms (orange) contains three carbonyl ligands (green and yellow) and the molecular-hydrogen ligand (blue). The fact that the H–H bond length (0.82 Å) is longer than in free H$_2$ (0.74 Å) and the fact that the two W–H bond lengths are equal (1.89 Å) suggest a three-center metal-dihydrogen bond and a substantially weakened H–H bond. (b) The preferred orientation of the H$_2$ ligand is parallel to P–W–P axis, suggesting that there is a barrier to rotation of the H$_2$ ligand about the W–H$_2$ axis. (c) A potential-energy curve for rotation of the H$_2$ ligand in a plane perpendicular to the W–H$_2$ axis with one degree of angular freedom (the angle θ) has minima for the identical orientations of 0 and 180 degrees from the P–W–P axis. Because the ground-state wave functions (dashed lines) for each potential well overlap (shaded areas), there is tunneling between potential wells and, as a result, the energy levels split.

(a) The W(CO)$_3$ (P(i-Pr$_3$)$_2$ H$_2$ Complex

(b) H$_2$-Ligand Dynamics
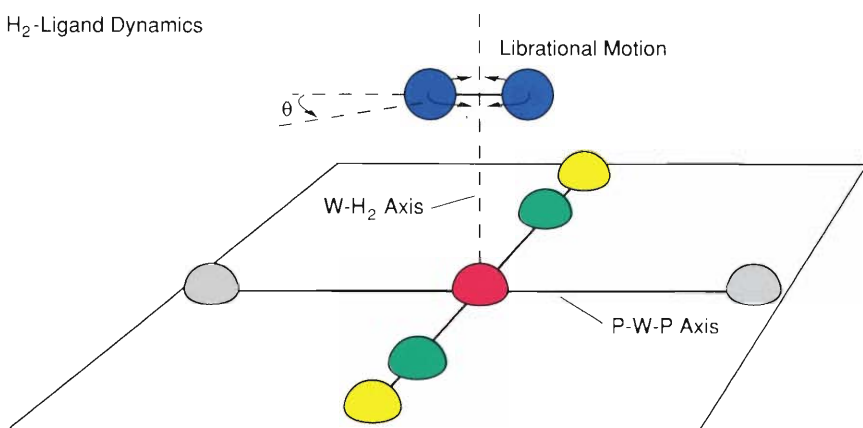

(c) Potential-Energy Barrier to Rotation

plained on the basis of interactions between the dihydrogen ligand and other ligands bound to the metal that would make alignment *perpendicular* to the P–W–P axis (and parallel to the OC–W–CO axis) energetically less favorable than alignment parallel to the P–W–P axis. We also observed that the hydrogen atoms of the bulky organophosphines formed a pocket around the region of the dihydrogen ligand, but the orientation of these organic groups is very accommodating and would not be expected to constrain the H$_2$ molecule.

Theoretical analysis is necessary to derive a more quantitative picture of the metal-ligand bonding than that indicated by the structural results. The most fundamental types of calculations can, in fact, derive structural parameters such
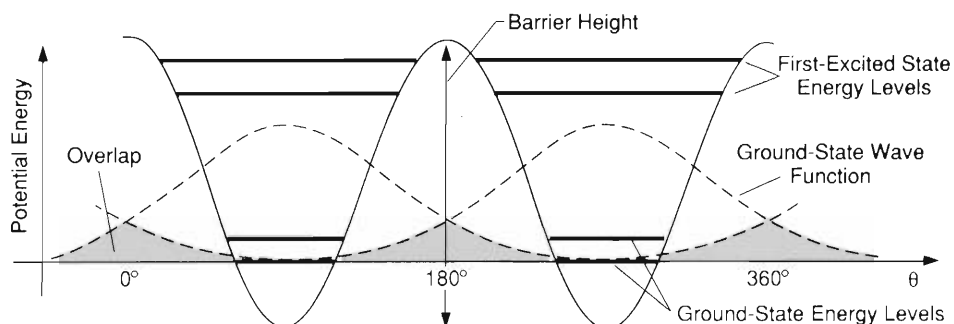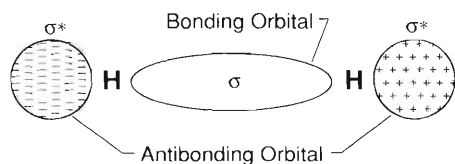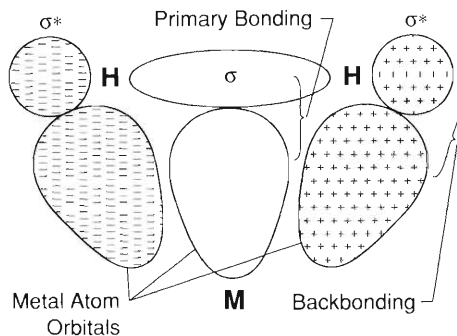
as the H–H or W–H distances; comparison with experimental values then serves as a check on the validity of the theory. The current problem, however, is sufficiently complex that structural information is used as input to simplified theoretical models. Whether or not the theoretical model is derived from first principles or from a combination of structural data and a theoretical model, it is highly desirable to have other experimental information on the nature of the chemical bonding that can be used to gauge the theoretical picture.

The nature of the bonding between the dihydrogen ligand and the transition metal is of major significance because the complex represents the first example of a *sigma-bond complex*, that is, a complex in which the ligand binds through interaction of a metal center with a σ-bonding electron pair. Theoretical studies of this three-center, two-electron bond indicate that both the bonding and antibonding orbitals of hydrogen (Fig. 9) may be involved. The

BONDING OF MOLECULAR-HYDROGEN LIGAND

Fig. 10. Although the main bonding between the dihydrogen ligand and the metal atom is due to an interaction between an empty metal-atom orbital and the σ orbital in the hydrogen molecule (here shown in a cross-sectional view), there is evidence for some *backbonding*, which is an interaction between an antisymmetric metal-atom orbital and the hydrogen σ^*-antibonding orbital. Both of these interactions weaken the H–H bond and strengthen the M–H bond. The former interaction donates electron density from the σ-bonding orbital of H_2 to the metal atom, whereas the latter interaction puts electron density from the metal atom into the H_2 antibonding orbital.

primary interaction between H_2 and the metal atom is donation of electron density from the H–H σ bond to an empty orbital in the metal atom (Fig. 10); however, the same studies indicate that, to a lesser degree, backbonding between a metal orbital and the H_2 antibonding sigma orbital (σ^*) also occurs. Backbonding stabilizes the side-on orientation shown in Figs. 8 and 10 rather than an end-on orientation (in which the H_2 molecule would have its bonding axis pointed straight at the tungsten atom with one hydrogen atom much closer to the metal atom than the other). The side-on coordination ultimately facilitates cleavage of the H–H bond to give dihydride complexes in oxidative addi-

MOLECULAR-HYDROGEN ORBITALS

Fig. 9. The usual theoretical picture of H–H bonding has the two electrons in the hydrogen molecule occupying a low-energy σ-bonding orbital whose electron-density map (here pictured schematically in a cross-sectional view) generally occupies the space between the two hydrogen nuclei. However, there is also a higher-energy σ^*-antibonding orbital that is usually unoccupied and whose electron-density map has a node between the two hydrogen nuclei. (The plus and minus patterns in the antibonding orbital are there to indicate the antisymmetric nature of the orbital.)

tion reactions. These theoretical predictions, of course, require experimental confirmation.

Rotational Dynamics. Hydrogen in the side-on coordination mode can undergo a remarkably wide variety of ligand dynamics, including torsional oscillations, or *librations*, about its equilibrium orientation and much slower 180-degree reorientations by tunneling through the rotational barrier. Establishing the presence of a significant electronic energy barrier to rotation would provide confirmation of metal-to-H_2 backbonding. Such a barrier is too small to be observed by standard nuclear-magnetic-resonance techniques. Inelastic neutron scattering, however, is highly sensitive to hydrogen motions because of the very large neutron-scattering cross section of protons and the typically large amplitude of the motions. In fact, this technique is routinely used to study rapid rotational motion (for example, of methyl groups and of solid or liquid hydrogen or molecular hydrogen in zeolites).

The nature of the rotational motion of the bound hydrogen molecule may be described with the aid of a diagram that shows the energy levels that the dihydrogen ligand may occupy as a function of the height of the barrier hindering the rotation. These levels are the solutions to the Schrödinger equation chosen to represent the rotational motion of the bound hydrogen molecule. In particular, the equation includes only one angular degree of freedom because we assume that the relatively strong three-center metal-dihydrogen bond keeps the hydrogen ligand essentially in a plane during its rotational motion. The complex may, in fact, be the first example of hydrogen rotation with only one degree of rotational freedom, a situation first described by Pauling as an approximation for solid hydrogen. If any mixing with vibrational modes can also be neglected,

the Schrödinger equation for the rotational motion is

$$\left[-B\frac{\partial^2}{\partial\phi^2} + \frac{1}{2}\sum_n V_{2n}(1-\cos 2n\phi)\right]\Psi = E\Psi, \quad (1)$$

where B is the rotational constant ($B = h^2/2I$, where I is the moment of inertia of the molecule for the rotation in question), ϕ is the angle of rotation about the O–C–W–H_2 axis, V_{2n} represents the barrier-height potential energy for a potential with $2n$ minima, and Ψ and E are, respectively, the wave function and energy of the allowed rotational states.

In the present case, as we've already pointed out, crystal-structure studies as well as theoretical calculations have shown the dihydrogen ligand to have a well-defined orientation parallel to the P–W–P axis. The hydrogen molecule in this complex should then have two equivalent orientations located at potential minima that are 180 degrees apart (Fig. 8c), and we may assume that the term with $n = 1$ (a simple double-minimum potential) will dominate. Equation 1 can then be reduced to the Mathieu equation, for which solutions are tabulated. The resulting energy-level diagram as a function of barrier height V_2 is shown in Fig. 11, in which both the energies and barrier heights are given in terms of B.

The energy levels corresponding to $V_2 = 0$ (left axis in Fig. 11) are those of a free rotor with one degree of freedom ($E_J = BJ^2$, where J is the rotational quantum number, yielding levels at energies of 0, B, $4B$, $9B$,...). Introduction of a barrier to this rotation ($V_2 > 0$) changes the level spacing drastically and removes some degeneracies. In the limit of very high barriers (suggested by the arrows on the right side of Fig. 11), the states approach a set of equally spaced energy levels characteristic of essentially harmonic torsional oscillations.
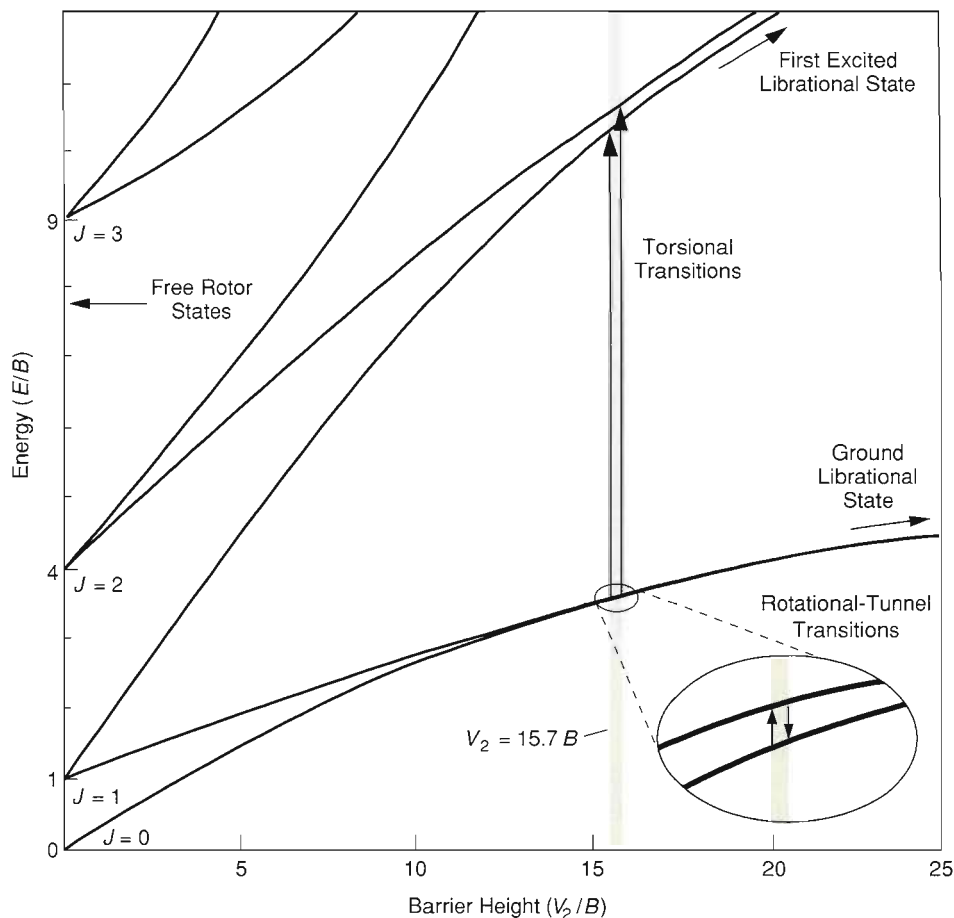

ROTATIONAL ENERGY-LEVEL DIAGRAM

Fig. 11. A dumbbell molecule (such as hydrogen) constrained to rotate in a plane has one rotational degree of freedom and rotational states J = 0, 1, 2, 3, ... at energies of 0, B, $4B$, $9B$,... if there is no barrier to the rotation (that is, if V_2 = 0). On the other hand, if V_2 is very high (that is, beyond the right side of the figure), the molecule will occupy a set of equally spaced torsional oscillator levels. For intermediate barrier heights we find a series of split librational states. The observed transitions (indicated by arrows) are of two types: transitions within the librational ground state called rotational-tunnel transitions (in which $\Delta I = \pm 1$) and transitions from the ground to the first excited librational state called torsional transitions. Because photons do not couple with nuclear moments, optical spectroscopy cannot be used to observe the tunneling transitions directly and can be used to observe only the torsional transitions between levels that have identical values of total nuclear spin ($\Delta I = 0$). The observed values for the transition energies are scaled by a value for B of 49.5 cm^{-1} rather than the 60 cm^{-1} value for free H_2 to reflect the increased H–H bond length (0.82 Å) relative to free H_2 (0.74 Å). The transitions shown are for a complex with a barrier height V_2 equal to 15.7 B.

The molecular hydrogen complexes being discussed here have intermediate barriers, and for these, we find a series of librational states, each of which is split relative to the torsional oscillator levels. This splitting arises from the fact that the barrier is *not* overly high, allowing the amplitude of the librations of the hydrogen molecule to be relatively large—large enough, in fact, for the wave functions that correspond to the molecule's being located in either of the two potential minima to overlap

(Fig. 8c). To satisfy the Pauli principle, the degenerate states corresponding to these two orientations must split into two states, each with a slightly different energy (Fig. 11 inset). The splitting is called *tunnel splitting* because it is due to the overlap of wave functions through a potential barrier. The size of the splitting decreases rapidly with increasing barrier height and is thus an extremely sensitive measure of barrier height.

The two resulting states are characterized by their symmetry. For exam-

(a) W-PCy₃ Complex

(b) W-P(i-Pr)₃ Complex
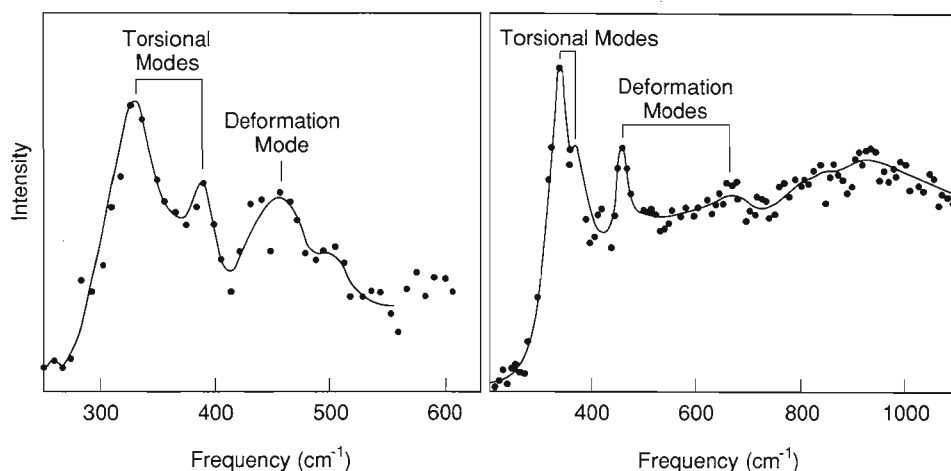
VIBRATIONAL-ROTATIONAL SPECTRA

Fig. 12. The high-frequency transitions associated with torsional, or rotational, motion of the dihydrogen ligand have been identified for the two complexes (a) $W(CO)_3(PCy_3)_2H_2$ and (b) $W(CO)_3(P(i-Pr)_3)_2H_2$ by using the Filter Difference Spectrometer at LANSCE. Unrelated frequencies in the spectra were eliminated by taking the difference between the scattering spectrum for the complex with a dihydrogen ligand and that for the complex with a dideuterium ligand. The deformation modes include rocking and wagging of the dihydrogen ligand with respect to the complex. One piece of evidence that the assignments are correct is the fact that, for inelastic neutron scattering, the modes with the largest-amplitude motions of the hydrogen atoms have the highest intensity and the rotational modes involve more motion of the hydrogen atoms than the rocking modes.

ple, 180-degree rotation corresponds to an odd permutation of identical spin-$\frac{1}{2}$ particles (the protons), with respect to which the total ground-state wave function must be antisymmetric. The low-temperature wave function can be constructed from linear combinations of nuclear-spin and rotational wave functions. Thus, a symmetric nuclear-spin wave function ($I = 1$, where I is the nuclear-spin-state quantum number) combines with an antisymmetric rotational wave function (J odd) and vice versa. These two cases correspond for zero barrier height to the two kinds of H_2 molecules referred to as ortho- and para-hydrogen, respectively. For finite barrier heights, J is no longer a "good" quantum number to describe the energy levels. The total nuclear spin of the molecule, however, must still change in a transition between the two lowest energy levels, that is, in a tunneling transition.

We note that transitions in which the total nuclear spin of the molecule changes cannot be observed in optical spectroscopy because photons do not couple to the nuclear spin. Neutrons,

however, do couple and are quite useful for studying rotational transitions of this type. The neutron has a nuclear spin of $\frac{1}{2}$, and a flip of the neutron spin during the scattering process will cause the total nuclear-spin state of the H_2 molecule to change also ($\Delta I = \pm 1$). A spin-flip neutron-scattering process then allows direct observation of the ortho-para transition in hydrogen—for example, para-hydrogen (with $I = 0$ and J even) changing to ortho-hydrogen (with $I = 1$ and J odd). For a *free* hydrogen molecule with two rotational degrees of freedom, the transition that changes the rotational state from $J = 0$ to $J = 1$ has an energy of $2B$, where B is the rotational constant. However, if the molecule is constrained to rotate in a plane with only *one* degree of rotational freedom, as is the case for our compounds, the transition has an energy of B for zero barrier height, that is, for free rotation. Moreover, as we discussed above, the energy for a tunneling transition rapidly becomes smaller with increasing barrier height until, at infinite barrier height, the splitting disappears and the two states become degenerate.

Experimental Confirmation. Now that we have selected an appropriate model for the rotational dynamics of the dihydrogen ligand in our system, it is a simple matter to relate the observed rotational transitions to the height of the rotational barrier. To observe both the high-frequency transitions to the excited librational state (the longer arrows in Fig. 11) and the very-low-frequency transitions associated with rotational tunneling (the two short arrows in the exploded portion of Fig. 11), we had to perform experiments on two spectrometers, each located at a different neutron source. The high-frequency torsional transitions were measured on the Filter Difference Spectrometer at LANSCE by using two samples for each complex, one of which had dideuterium instead of dihydrogen ligands. Vibrational modes involving mainly the dideuterium ligand cannot be "seen" in the presence of the many more modes that include hydrogen motion (that is, those of the organophosphine ligands). The deuterium-substituted sample thus served as a "blank" for subtracting all the various vibrational modes except those of interest—the motions of the dihydrogen ligand. Figure 12 shows the results for two tungsten complexes: one with tricyclohexylphosphine ligands (PCy₃) and one with the less bulky tri-isopropylphosphine ligands (P(i-Pr)₃).

The low-frequency rotational tunneling spectra for three complexes (Fig. 13) were obtained on a cold-neutron time-of-flight spectrometer at the High Flux Reactor of the Institut Laue-Langevin in Grenoble, France. No "blank" sample was necessary in this case, since the other ligands were not expected to have observable excitations in the frequency range of interest for this experiment (which is less than 10 cm⁻¹).

For the two tungsten complexes with PCy₃ and P(i-Pr)₃ ligands, this type of analysis yielded a significant barrier height—one that was roughly 15

ROTATIONAL-TUNNELING SPECTRA

Fig. 13. The spectra for the low-frequency transitions associated with rotational tunneling are shown here for the three complexes (a) Mo(CO)$_3$(PCy$_3$)$_2$H$_2$, (b) W(CO)$_3$(PCy$_3$)$_2$H$_2$, and (c) W(CO)$_3$(P(i-Pr)$_3$)$_2$H$_2$. The strong central peak in each spectrum is an elastic-scattering line, whereas the peaks to both sides of that line are the inelastic-scattering transitions associated with rotational tunneling. The fact that the inelastic peaks have a doublet nature is most likely due to structural disorder in the crystals.

times our derived rotational constant for bound H$_2$. Using the barrier heights and the energy-level diagram (Fig. 11), we were able to calculate the frequencies expected for the high-frequency transitions associated with the torsional motion of both complexes. The calculated values are in good agreement with the experimental values measured with the Filter Difference Spectrometer (Fig. 12), which suggests that the simple model of planar reorientaton in a double-minimum potential is a reasonable description for the hydrogen motion in these systems.

The crucial question at this point becomes what interactions give rise to the barrier to rotation. Two possible sources for the hindrance potential are *electronic* and *steric* effects. By an electronic effect we mean that the dihydrogen ligand may be constrained in its orientation because of the way the chemical bond is formed with the metal. In other words, the electron orbitals on the metal sharing electrons with those of the dihydrogen ligand have a symmetry that determines the orientation of the ligand.

Steric effects refer to the interactions of the dihydrogen ligand with the surrounding atoms of the other ligands. These are nonbonding interactions that may be described by van der Waals forces between pairs of atoms. They may be summed up for all pairs formed

(a) Mo-PCy$_3$ Complex
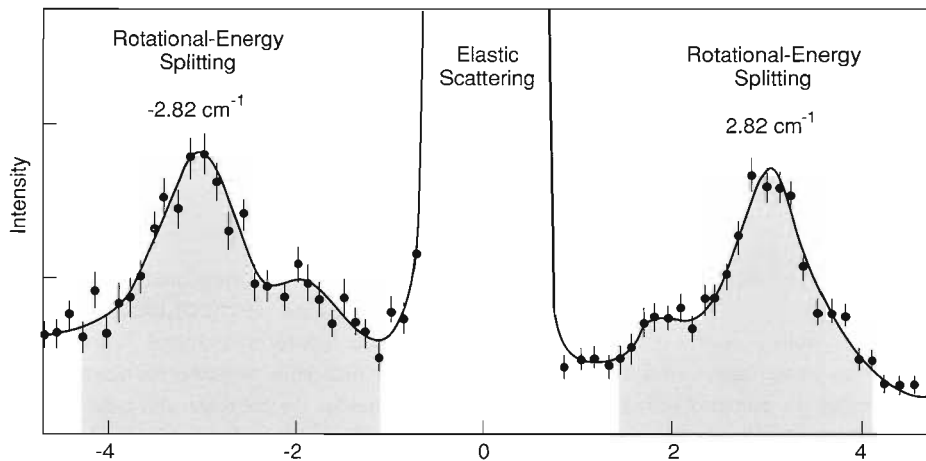
(b) W-PCy$_3$ Complex
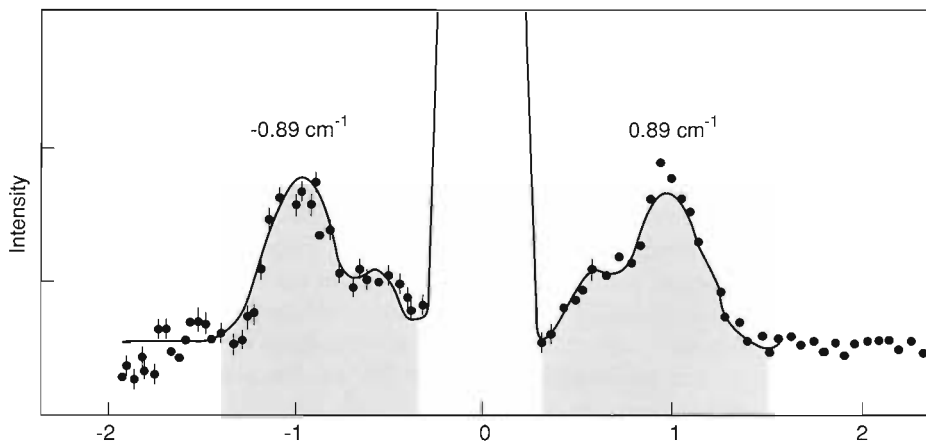
(c) W-P(i-Pr)$_3$ Complex
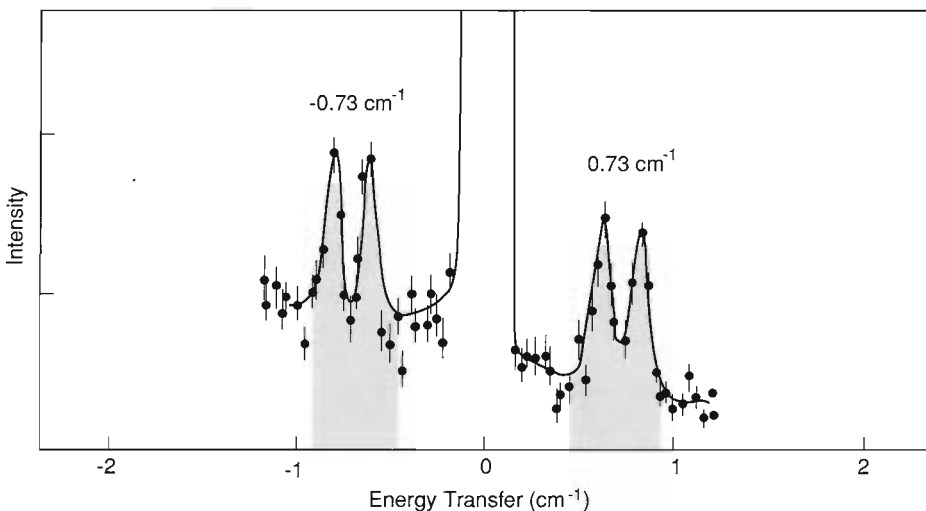
by using either one of the H atoms on the dihydrogen ligand and any one of the surrounding atoms. As the dihydrogen ligand is rotated, the sum of these interactions shows an angular variation, which gives rise to an effective "steric" barrier.

In an attempt to sort out the relative effects of these two types of interactions, we performed separate measurements on the two tungsten complexes with PCy_3 and $P(i$-$Pr)_3$ ligands; then we replaced the tungsten atom in the PCy_3 complex with a molybdenum atom and took measurements on this third complex. Thus, we hoped to gauge the effects of changing the central metal atom and of replacing the large, bulky PCy_3 ligand with the less bulky $P(i$-$Pr)_3$ ligand.

The peaks in the spectra of Fig. 13 to the left and right of the strong elastic line represent the rotational-energy splitting associated with the H_2 molecule tunneling through the barrier from one 180-degree orientation to the other. The position of these lines is extremely sensitive to the height of the barrier. A comparison of the three spectra shows that replacement of tungsten (Fig. 13b) with molybdenum (Fig. 13a) changes the tunneling frequency by a factor of just over 3, from 0.89 to 2.82 cm^{-1}. On the other hand, replacing the PCy_3 ligand in the tungsten complex with the less bulky $P(i$-$Pr)_3$ ligand (Fig. 13c) changes the frequency by less than 20 percent, from 0.89 to 0.73 cm^{-1}.

The change that occurs when the central metal atom is replaced may be taken as reflecting the metal-dihydrogen bonding directly; that is, it is essentially an electronic effect. Replacing the PCy_3 ligands with $P(i$-$Pr)_3$ ligands, on the other hand, probably has little effect on the electronic state of the metal and therefore on the metal-dihydrogen bonding. These ligands most likely produce a steric component of the barrier to H_2 rotation through direct, nonbonded in-

Table 1

Barrier Heights to Rotation for the Dihydrogen Ligand

Complex		Theoretical (kcal/mole)			Observed (kcal/mole)
Metal Atom	Ligand	Molecular Mechanics	ab initio (PH$_3$)	Sum	
W	PCy$_3$	0.6	1.8	2.4	2.2
W	P(i-Pr)$_3$	1.4	1.8	3.2	2.4
Mo	PCy$_3$	0.6	0.6	1.2	1.5

teractions. Thus, the experimental evidence, at least in these cases, strongly suggests that the barrier to H_2 rotation is determined more by electronic than steric effects. To test this conclusion, Jeff Hay, John Hall, and Caroline Boyle of the Theoretical Division at Los Alamos carried out two sets of calculations: an *ab initio* calculation—that is, from first principles—and a *molecular-mechanics* calculation.

The ab initio calculation treats primarily the electronic effects because a full set of one-electron wave functions for the whole molecule is used to compute the relative energy of a given configuration. The barrier to rotation was obtained from the difference in total energies for the structure with the H_2 aligned along the P–M–P axis and the structure with the H_2 aligned along the OC–M–CO axis. The calculation is rather extensive, and the bulky organophosphine ligands must be simplified to make it possible at all. When unsubstituted phosphine (PH_3) is used as a ligand instead of tricyclohexylphosphine or tri-isopropylphosphine, the calculation yields a barrier height of 1.8 kilocalories per mole for the tungsten complex and 0.6 kilocalorie per mole for the molybdenum complex.

The second type of calculation—the molecular-mechanics type—may be

viewed as representing mainly steric effects. In this case, the pairwise, nonbonding interactions between the hydrogen atoms of the dihydrogen ligand and each of the other atoms of the molecule are summed. The summation is repeated for each orientation of the ligand, generating a curve of potential energy as a function of orientation. This calculation is not sensitive to the type of metal atom at the center of the molecule. The results show a barrier height of 0.6 kilocalorie per mole for the complexes with PCy_3 ligands and 1.4 kilocalories per mole for those with $P(i$-$Pr)_3$ ligands.

If one makes the assumption that molecular mechanics treats only the steric effects and that the ab initio theory accounts primarily for the direct electronic interaction between H_2 and the metal, barrier heights from the two calculations may be added to arrive at an estimate of the effective total barrier. These assumptions are not unreasonable, because replacement of tungsten by molybdenum has no effect on the results of the molecular-mechanics case, whereas the ab initio theory uses the very small PH_3 ligands, rather than $P(i$-$Pr)_3$ or PCy_3 ligands, and thus essentially ignores steric effects. The sum of the two calculations for each of the complexes is shown in Table 1 along with the corresponding barrier height

calculated from the observed inelastic neutron-scattering data.

The calculated and observed barriers to H_2 rotation appear, at first glance, to agree only qualitatively. If, however, one takes into account the various limitations of the theoretical calculations, the agreement with experiment is remarkably good. For example, the necessary structural information is not known in detail for all three complexes in this experiment, and both types of calculations are normally used for barrier heights that are a factor of ten or so higher than the one in this study. Furthermore, comparison with experimental data does suggest that the molecular-mechanics calculation overestimates the steric part of the barrier, since replacement of the PCy_3 with $P(i\text{-}Pr)_3$ is found to change the barrier height by 0.8 kilocalorie per mole, which is four times the experimentally observed change. In view of these considerations, we can clearly conclude that the direct electronic binding of the dihydrogen ligand to the metal contributes significantly to the barrier, at least one-half to two-thirds of the experimentally determined value.

The rotational tunnel splitting is an extremely sensitive measure of the barrier height—in fact, it depends *exponentially* on the value of the barrier. Such sensitivity has clear advantages. For example, the observation of a higher barrier in the tungsten complex than in the molybdenum complex is an indication of stronger binding of the hydrogen molecule to the metal atom. This conclusion can be reached both by observing the higher $M–H_2$ infrared stretch frequency in the tungsten complex or by observing the difference in rotational tunnel splitting. However, the change in infrared stretch frequency is only 10 percent, whereas the change in rotational tunnel splitting is more than 50 percent. It is therefore clear that rotational tunneling spectroscopy of side-on

H_2 by inelastic neutron scattering can be used as a probe of the details of metal-to-H_2 binding.

Given the fact that the directional properties of the electron wave functions that help optimize the electron flow between the dihydrogen ligand and the metal atom also seem to be largely responsible for the barrier to H_2 rotation, we feel that establishment of a significant electronic component to the barrier height is convincing evidence of σ^* backbonding between the metal atom and molecular hydrogen.

The latter conclusion, in particular, is a truly remarkable result of our neutron-scattering studies and illustrates the very fundamental details to which these catalytic model systems can be studied with such techniques. Apart from our model systems, many more realistic catalytic materials are being investigated by the same techniques—studies that are often greatly aided by previous work on model compounds. These more realistic systems include molecules adsorbed on dispersed metal particles, inside cavities of zeolites, or attached to many other active substrates. Although the level of detail that can safely be inferred from the "real" catalytic systems is somewhat lower than for the simpler model systems, significant progress can nonetheless be expected in understanding the catalytic function of these materials on an atomic scale. Neutron scattering will certainly play an important role in these studies. ■

Acknowledgments

Juergen Eckert earned his B.S. at Yale University and his Ph.D. at Princeton University in 1975. Most of the research for his doctoral thesis, which involved neutron-scattering studies of the lattice dynamics of solid neon, was carried out as a member of the neutron-scattering group in the Physics Department of Brookhaven National Laboratory. After earning his doctorate, he remained at Brookhaven until 1979, when he accepted a staff membership at Los Alamos to initiate neutron-scattering research on the newly commissioned pulsed neutron source (WNR) at LAMPF. The focus of his work has steadily shifted towards applications of neutron scattering to chemistry, albeit from a physicist's perspective. He recently spent one year as a visiting scientist at the Institut Laue Langevin in Grenoble, France, where some of the work described here was carried out.

Phillip J. Vergamini received his M.S. in inorganic chemistry from the University of Minnesota, Minneapolis, in 1968 after earning his B.S. at the University of Wisconsin, Superior. He completed his Ph.D. in inorganic chemistry at the University of Wisconsin, Madison, in the fall of 1971 and then joined the Isotope and Nuclear Chemistry Division at Los Alamos. His research there involved application of spectroscopic and x-ray crystallographic techniques to studying the synthesis and structure of inorganic and organometallic compounds. In the spring of 1980, he joined the neutron-scattering group at Los Alamos and took responsibility for the construction and use of the Single-Crystal Diffractometer, one of the first two LANSCE instruments to be placed into the international users program. With that diffractometer he has studied the structures of materials ranging from molecular-hydrogen complexes to the crystalline mineral in dinosaur bones.

Further Reading

R. K. Thomas. 1982. Neutron scattering from adsorbed systems. *Progress in Solid State Chemistry* 14: 1–93.

C. J. Wright. 1985. Surface characterization by the inelastic scattering of neutrons from absorbates. In *The Structure of Surfaces*, edited by M. A. Van Hove and S. Y. Tong, pp. 210–218. Springer Series in Surface Science, volume 2. Berlin: Springer-Verlag.

T. J. Udovic and R. D. Kelley. 1988. Neutron scattering studies of hydrogen in catalysts. In *Hydrogen Effects in Catalysis*, edited by Z. Paál and P G. Menon, pp. 107–182. New York: Marcel Dekker.

G. A. Somorjai. 1986. Surface science and catalysis. *Philosophical Transactions of the Royal Society of London* A318: 81–100.

Norman Sheppard. 1988. Vibrational spectroscopic studies of the structure of species derived from the chemisorption of hydrocarbons on metal single-crystal surfaces. *Annual Review of Physical Chemistry* 39: 589–644.

E. L. Muetterties, T. N. Rhodin, Elliot Baud, C. F. Brucker, and W. R. Pretzer. 1979. Clusters and surfaces. *Chemical Review* 79: 91–137.

R. R. Cavanagh, J. J. Rush, R. D. Kelley, and T. J. Udovic. 1984. Adsorption and decomposition of hydrocarbons on platinum black: Vibrational modes from N15. *Journal of Chemical Physics* 80: 3478–3484.

R. Whyman. 1980. Metal clusters in catalysis. In *Transition Metal Clusters*, edited by Brian F. G. Johnson, pp. 545–606. New York: John Wiley and Sons.

Juergen Eckert, Gregory J. Kubas, John H. Hall, P. Jeffrey Hay, and Caroline M. Boyle. 1990. Molecular hydrogen complexes. 6. The barrier to rotation of η^2-H$_2$ in M(CO)$_3$(PR$_3$)$_2$(η^2-H$_2$) (M = W, Mo; R = Cy, i-Pr): Inelastic neutron scattering, theoretical, and molecular mechanics studies. *Journal of the Americal Chemical Society* 112: 2324.

Gregory J. Kubas. 1988. Molecular hydrogen complexes: Coordination of a σ bond to transition metals. *Accounts of Chemical Research* 21: 120–128.

J. Eckert, A. Albinati, and G. Longoni. 1989. Inelastic neutron-scattering study of K[HCo$_6$(CO)$_{15}$]: Implications for the location of the hydride. *Inorganic Chemistry* 28: 4055.

Pier Luigi Stanghellini and Guiliano Longoni. 1987. Vibrational studies of interstitial hydrogen in metal carbonyl clusters. *Journal of the Chemical Society. Dalton Transactions* 685–690.

Jacques Roziere and Antoine Potier. 1982. Liaison métal-hydrogène-métal et spectroscopie de vibration. *Bulletin de la Société Chimique de France. Partie I.* I-339–346.

Werner Press. 1981. *Single-Particle Rotations in Molecular Crystals*. Springer Tracts in Modern Physics, volume 92. Berlin: Springer-Verlag.

P. Jeffrey Hay. 1987. Ab initio theoretical studies of dihydrogen coordination vs. oxidative addition of H$_2$ to five-coordinate tungsten complexes. *Journal of the American Chemical Society* 109: 705–710.

Gregory J. Kubas, Robert R. Ryan, Basil I. Swanson, Phillip J. Vergamini, and Harvey J. Wasserman. 1984. Characterization of the first examples of isolable molecular-hydrogen complexes, M(CO)$_3$(PR$_3$)$_2$(H$_2$) (M = Mo, W; R = Cy, i-Pr). Evidence for a side-on bonded H$_2$ ligand. *Journal of the American Chemical Society* 106: 451–452.

George Edmund Bacon. 1977. *Neutron Scattering in Chemistry*. London: Butterworths.

R. E. Lechner and C. Riekel. 1983. Applications of neutron scattering in chemistry. In *Neutron Scattering and Muon Spin Rotation*. Springer Tracts in Modern Physics, volume 101. Berlin: Springer-Verlag.

Robert Bau, editor. 1978. *Transition Metal Hydrides*. Washington, D.C.: American Chemical Society.

H. Ibach and D. L. Mills. 1982. *Electron Energy Loss Spectroscopy and Surface Vibrations*. New York: Academic Press.

J. Eckert. 1986. Neutron vibrational spectroscopy: The use of hydrogen as a structural and dynamical probe. *Physica B+C* 136: 150–155.

x-ray and neutron

CRYSTALLOGRAPHY

a powerful combination

by Robert B. Von Dreele

Determining the structure of a crystalline material remains the most powerful way to understand that material's properties–which may explain why so many Nobel Prizes have been awarded in the field of crystallography. The standard tools of the crystallographer are single-crystal and powder diffraction, introduced earlier in "Neutron Scattering-A Primer." What was not mentioned was that until twenty years ago powder diffraction could not be used for solving a new crystal structure, but only for determining the presence of known crystalline phases in powders of unknown composition. At that time material had to be grown into large single crystals before crystallographers could unravel the positions of each atom within the repeating motif of a crystal lattice. This severe limitation disappeared after H. M. Rietveld developed a workable approach for resolving the ambiguities of most powder-diffraction patterns. The technique, known as Rietveld refinement, has opened up essentially all crystalline materials to relatively rapid structure analysis.

This Escher painting shows a square lattice with a complicated unit cell, illustrating in two dimensions several kinds of symmetries found in real crystals. (We have darkened lines of the original grid to emphasize the unit cell.) If the colors are ignored, this pattern has both fourfold and twofold rotational symmetry as well as a number of mirror symmetry operations. When the color is included, the fourfold rotation becomes a color-transformation operator. Similar changes occur in the nature of the other symmetry operators as well. Reproduced with permission: ©1990 M. C. Escher Heirs/Cordon Art, Baarn, Holland.

This article presents a further improvement in powder-pattern analysis–that of combining x-ray and neutron diffraction data. We used this combination to make the first unambiguous determination of the structures of certain high-temperature superconductors and have since produced a portable software package for use by all crystallographers who collect both x-ray and neutron data. Here we will discuss the concepts and techniques that make the combination so useful and some of our recent results, including the determination of fractional occupancies by different elements at single atomic sites in a crystal. First, however, we need to extend the concepts on diffraction introduced in the primer.

What Is a Crystal?

Most solids are crystals: They consist of very many repetitions of a single motif or "unit cell," of atoms. These repetitions occur at a regular array of points in three dimensions, a "lattice." The opening illustration is a two-dimensional analogue of a crystal. The unit cell there is square, and contains several objects each arranged in a particular way relative to the others. One question about this pattern is how much information one needs in order to reproduce it. Clearly, one need only describe a single object (a fish), the set of rules for positioning it and the other objects in the unit cell (the fish of other colors), and the dimension of the unit cell itself. With only this information the entire pattern can be laid out to infinity. The classification of how the objects are positioned in the unit cell (in most crystals these positions are symmetrical) and of how the unit cells repeat is the mathematical theory of spatial symmetry, which is a branch of group theory (see the sidebar "Crystal Symmetry Groups").

The crystallographer's goal is to measure the lengths and angles of the edges of the unit cell (the "lattice parameters") and, more important, the arrangement of the atoms within the unit cell. Many kinds of arrangements are possible, for example, the interlacing of long molecular chains in a crystallized protein, or the stacking of metal and oxygen atoms in a superconducting oxide, but in any crystal the arrangement is the same in every unit cell. Why should atoms and molecules form such orderly structures? A solid holds together because the atoms and molecules in it are attracted to each other. Thus the minimum-energy configuration of the solid occurs when its constituents are in as close contact as possible with their neighbors. This criterion is usually realized by a regular array, just as bricks in a neat stack are in closer contact and take up less space than bricks in a jumbled pile.

The unit cell of a crystal is extremely small, typically 10 angstroms (10^{-7} centimeter) on a side, whereas the sides of crystals in a powder may be 1000 to 100,000 times larger. An equivalent stack of bricks, each 20 centimeters on a side, would extend between 200 meters and 20 kilometers. The disparity in size between a unit cell and a crystal is so vast that we can model a crystal as if it contained an infinite number of unit cells in all directions. This approximation has an enormous simplifying effect on a mathematical description of a crystal because we need to describe only the unit cell and can ignore the crystal as a whole except to note that the unit cell repeats indefinitely in all directions.

With these ideas in mind, we can start with crystallographic mathematics and then connect it with the way a crystal scatters neutrons (reversing the plan of the primer). How do we mathematically describe a crystal? First, the description must reflect what we actually observe about a crystal. We "see" atoms in a crystal by scattering neutrons or x rays from them, so the mathematical model needs to describe the density of scattering power, $\rho(\mathbf{r})$, a function of position, $\mathbf{r}$, within the crystal. This scattering density is smooth and usually real and positive. (In some special cases it can be negative or even complex.) Second, the function needs to repeat infinitely in all directions to match the repetition of the unit cells. In one dimension $\rho(x)$ might look like the curve in Fig. 1a, which gives the x-ray scattering density along one di-

rection in molybdenum disulfide for two unit-cell repeats. The tallest peaks represent the scattering density around the molybdenum atoms; the smaller peaks on either side correspond to the sulfur atoms. Like any periodic function, the variation of the scattering density with position x along the repeat direction can be expressed as an infinite sum of sine and cosine functions, or in other words, as a Fourier series in one dimension:

$$
\begin{aligned}
\rho(x) &= \frac{1}{a} \sum_{n=0}^{\infty} F_n \left[\cos\left(-2\pi\frac{n}{a}x\right) + i \sin\left(-2\pi\frac{n}{a}x\right) \right] \\
&= \frac{1}{a} \sum_{n=0}^{\infty} F_n \exp(-2\pi i \frac{n}{a} x) = \frac{1}{a} \sum_{n=0}^{\infty} F_n \exp(-iQ_n x).
\end{aligned}
\tag{1}
$$

Here n is an integer, a is the length of the unit cell in the x direction, and $Q_n \equiv 2\pi n/a$. Figure 1b shows the first eight terms in the Fourier series for $\rho(x)$ of MoS$_2$. Each term represents a stationary wave, or "Fourier component," of scattering density whose wavelength is a/n, so that in the repeat distance a the wave undergoes exactly n oscillations. Thus the sum in Eq. 1 contains only waves that have a as a repeat distance. Each stationary wave has an amplitude F_n, which for the MoS$_2$ structure is either positive or negative. In the most general case F_n can be complex.

Just as the displacement, x, can be represented by a vector in one-dimensional real space, the inverse wavelengths n/a (= $Q_n/2\pi$) can be represented by vectors in one-dimensional "reciprocal space." These "reciprocal-lattice" vectors define a row of equally spaced points, labeled by the values of n. All the remaining reciprocal space is empty. The points are called the "reciprocal lattice" because their spacing is $1/a$, the reciprocal of the real-lattice spacing. (The name "reciprocal space" has the same origin.) Their locations depend only on a, the periodicity of the real lattice, and not on the contents of the unit cell. In Fig. 1c the amplitude F_n of the nth Fourier component of $\rho(x)$ for MoS$_2$ is plotted at the reciprocal-lattice point n/a.

Thinking of the Q_n's as one-dimensional vectors (the wave vectors of the Fourier components), we note from the definition of the Q_n's and the discussion of diffraction in the primer that when the momentum transfer in a diffraction experiment $(h/2\pi)\mathbf{Q} = (h/2\pi)\mathbf{Q}_n$, we observe a Bragg peak whose intensity is $S(\mathbf{Q}_n) = |F_n|^2$. In crystallographic terminology the F_n's are called structure factors; unfortunately the same name is used by the neutron-scattering community for $S(\mathbf{Q}_n) = |F_n|^2$. Whatever the nomenclature, crystallographers frequently describe crystals in reciprocal space because the quantities they measure directly are the reciprocal-lattice vectors $\mathbf{Q}_n$ and the intensities on the reciprocal lattice $S(\mathbf{Q}_n)/2\pi$. Figure 2 shows a variation of Fig. 1 whose significance will be discussed later.

In order to extend Eq. 1 to descriptions of three-dimensional crystals, we replace $Q_n = 2\pi n/a$ with a three-dimensional wave vector $\mathbf{Q_h}$. For simplicity we begin with a real lattice whose three axes are mutually perpendicular, as shown in Fig. 3a. Then the natural coordinates are orthogonal, and

$$
\mathbf{Q_h} = 2\pi \left(\frac{h}{a}, \frac{k}{b}, \frac{l}{c} \right).
$$

Here a, b, and c are the repeat distances along the three axes of the unit cell, or lattice spacings, and the integer triplet $\mathbf{h} = (hkl)$ gives the components of $\mathbf{Q_h}$ along the three axes of the reciprocal-lattice unit cell, measured in units of the reciprocal-lattice repeat distances $a^* = 1/a$, $b^* = 1/b$, and $c^* = 1/c$. Thus, in analogy with the one-dimensional case, the integer triplets $\mathbf{h}$ specify all the possible $\mathbf{Q_h}$ values, that is, all the wave vectors of Fourier components of the three-dimensional scattering-density distribution. Each $\mathbf{Q_h}$ is perpendicular to a stack of parallel planes in real space, and $2\pi/|\mathbf{Q_h}|$ (which has the dimensions of length) is the spacing between those planes, commonly called the "d-spacing". Each $\mathbf{h}$ labels a set of planes

(a) Scattering Density

(b) Fourier Components

A ONE-DIMENSIONAL LATTICE AND ITS RECIPROCAL-SPACE REPRESENTATION

Fig. 1. (a) The x-ray scattering density along one direction of molybdenum disulfide illustrates a one-dimensional lattice with a unit cell of length a. (b) The first eight Fourier components (n = 0 to 7) in Eq. 1 for the scattering-density function in (a). The wavelength of the nth component is $\frac{a}{n}$. Note that the phase of some of the waves is offset by 180° with respect to others. (c) The amplitudes F_n of the Fourier components from (b) plotted in reciprocal space. The reciprocal lattice is the set of points $\frac{n}{a}$ whose spacing is a^{-1}, the reciprocal of the lattice spacing in real space. Note that some amplitudes are negative (those of waves shifted in phase by 180° with respect to the waves with positive amplitudes). (d) The intensities $|F_n|^2$, or $S(Q_n)$, plotted in reciprocal space. This pattern of intensities would be obtained from a diffraction experiment. This pattern reveals the size of the unit cell, but as explained in Fig. 2 does not yield a unique determination of the contents of the unit cell.

(c) Fourier Amplitudes, F_n

(d) Fourier Intensities, $|F_n|^2$
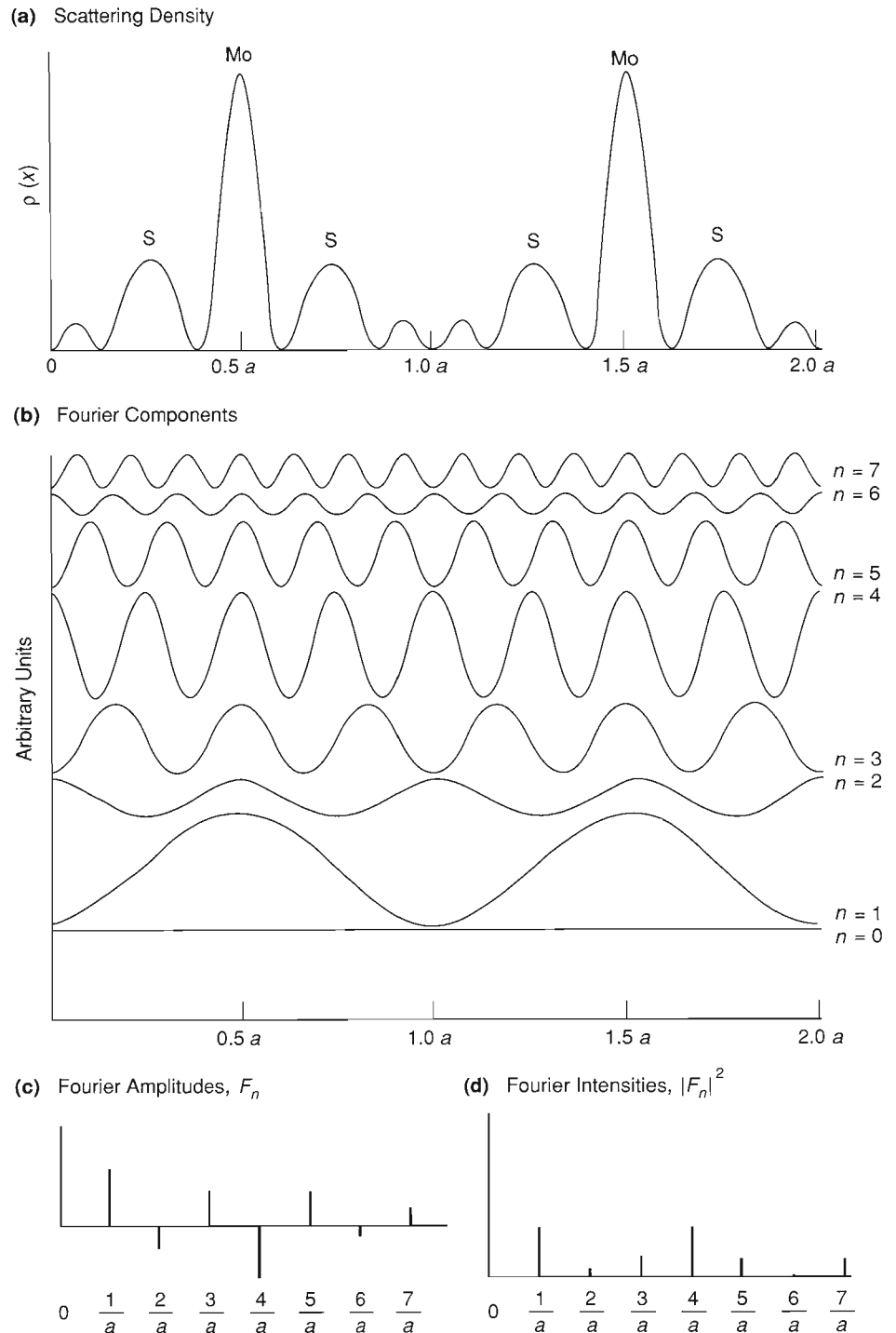
perpendicular to Q_h. Together the h's specify all the sets of planes that pass through unit cells in a periodic way. Therefore just as in one dimension a sum over the wave vectors $2\pi n/a$ with integer n sufficed to describe a periodic $\rho(x)$, in three dimensions a sum over the wave vectors Q_h, or over the h, is all we need to describe $\rho(r)$ for a crystal. The integer components hkl of h are identical to the Miller indices that crystallographers use to label faces along which crystals break. More important, the Q_h are the special wave vectors Q at which Bragg scattering can occur, as defined in the primer. In the general case, illustrated in Fig. 3b, the $Q_h(\equiv 2\pi(h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*))$ are still perpendicular to stacks of planes and their lengths are still equal to 2π divided by the d-spacings. Note that the Q_h must be defined in terms of the translation vectors of the reciprocal lattice, which are no longer simply parallel to the translation

(a) Scattering Density

(b) Fourier Components
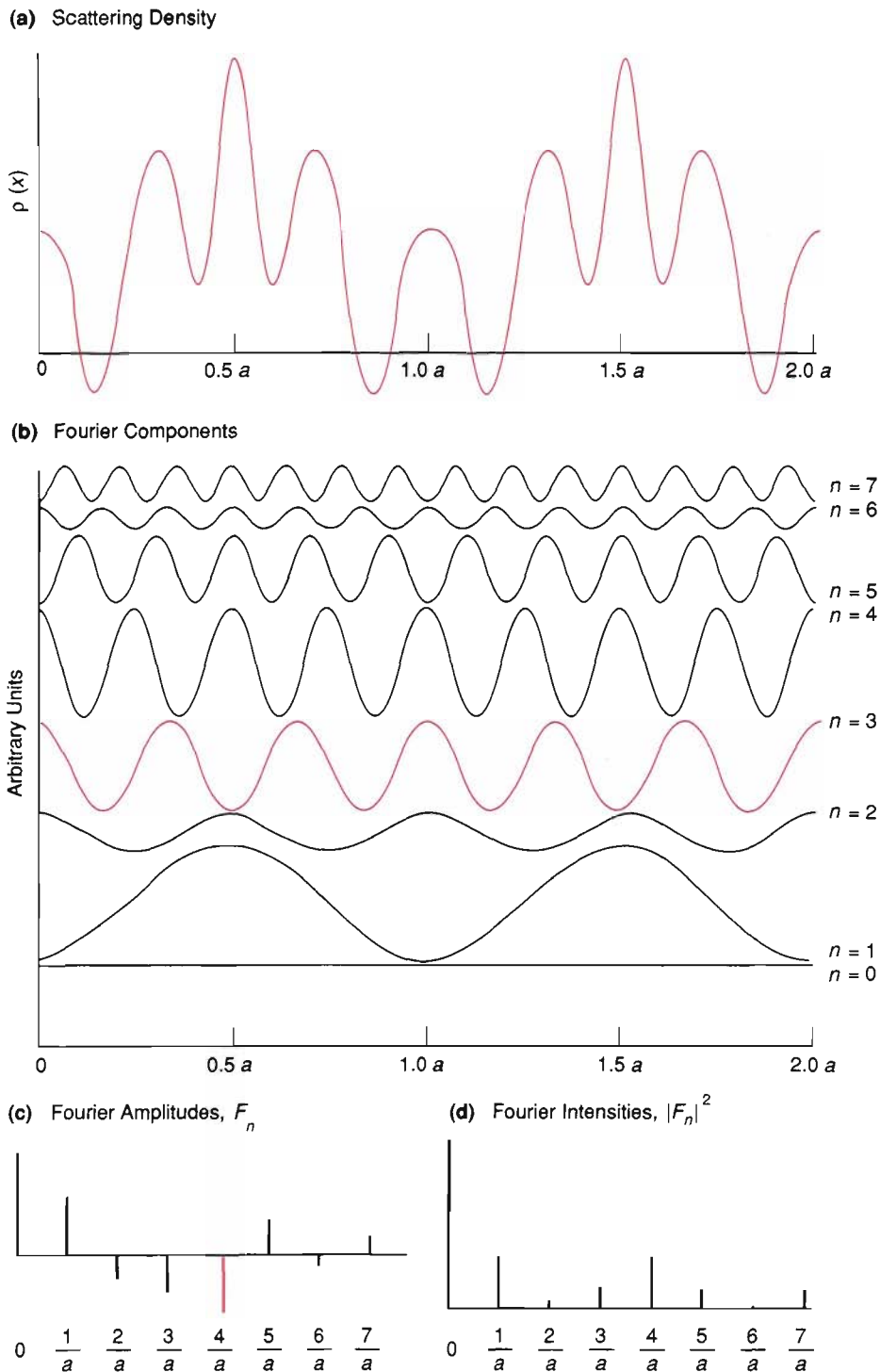
THE PHASE PROBLEM IN CRYSTALLOGRAPHY

Fig. 2. The same as Fig. 1, except that one Fourier component has been phase-shifted by 180° to produce an entirely different and fictitious scattering density for MoS_2. The shifted wave and its amplitude are shown in red. This example illustrates the ambiguity that arises in diffraction experiments from measuring the magnitudes of the F_n's but not their phases. Although the plot of the F_n's changes, the plot of $|F_n|^2$, which is analogous to a diffraction pattern, does not. Thus diffraction experiments can not distinguish the scattering density in Fig. 1 from that in this figure. Determining the phases is called "solving the structure" because only then can the contents of the unit cell be determined.

(c) Fourier Amplitudes, F_n

(d) Fourier Intensities, $|F_n|^2$

vectors of the real lattice but are more complicated functions of its parameters.

Returning to the three-dimensional version of Eq. 1, we replace the product $Q_n x$ by the dot product $\mathbf{Q_h} \cdot \mathbf{r}$ and normalize the Fourier series by the unit-cell volume V_c:
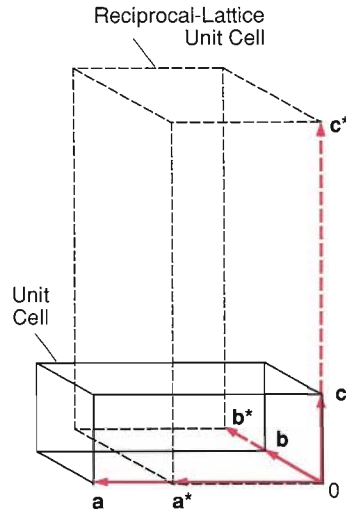
$$\rho(\mathbf{r}) = \frac{1}{V_c} \sum_{\mathbf{h}} F_\mathbf{h} \exp[-i(\mathbf{Q_h} \cdot \mathbf{r})]. \qquad (2)$$

(A specialist in this field would write equations such as this in crystallographic co-ordinates, using $\mathbf{h}$ instead of $\mathbf{Q_h}$ and defining $\mathbf{r}$ in terms of displacements along the crystal axes instead of along the Cartesian directions; furthermore the displacements

(a) Orthorhombic Unit Cell

(b) Trigonal Unit Cell

Reciprocal-Lattice
Unit Cell

Unit
Cell

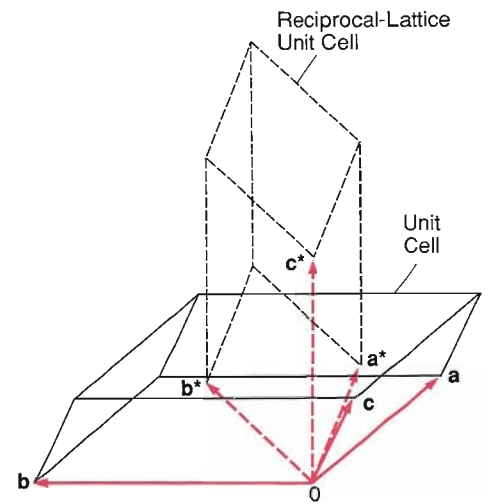Reciprocal-Lattice
Unit Cell

Unit
Cell

UNIT CELLS IN REAL AND RECIPROCAL SPACE

Fig. 3. (a) A unit cell in real space (solid lines) and its associated reciprocal unit cell (dashed lines), for a three-dimensional lattice whose translation vectors, a, b, and c, are mutually orthogonal. The reciprocal-lattice translation vectors a*, b*, and c* are collinear with those of the real-space lattice, but their lengths are the reciprocals of the lengths of the real-space vectors. Note that a*=Q_{100}/2π is normal to the real-space bc, or (1 0 0), plane; similarly the other reciprocal-lattice translation vectors are normal to their corresponding planes. (b) A unit cell and reciprocal unit cell for a lattice in which none of the translation vectors a, b, and c are orthogonal. The reciprocal-lattice translation vectors a*, b*, and c* are no longer collinear with their real-space counterparts but they are still normal to the planes that bound the unit cell, and their lengths are the reciprocals of the spacings of those planes. (c) The mathematical formulas describing the reciprocal-lattice translation vectors and the wave vectors Q_h. (d) An example of part of a set of planes and its d-spacing in a crystal whose unit cell is that shown in (b). The planes are labeled by h = (0 0 1) (parallel to the ab plane), and their d-spacing is equal to $|c^*|^{-1}$. In general $Q_h \equiv 2\pi(ha^* + kb^* + lc^*)$ and $Q_{001} = 2\pi c^*$ is perpendicular to the 001 planes.

(c) General Properties of Reciprocal-Lattice Translation Vectors and Q_n

$$a^* = \frac{b \times c}{a \cdot (b \times c)} \ , \quad b^* = \frac{c \times a}{b \cdot (c \times a)} \ , \quad c^* = \frac{a \times b}{c \cdot (a \times b)}$$

$$a^* \cdot a = b^* \cdot b = c^* \cdot c = 1$$
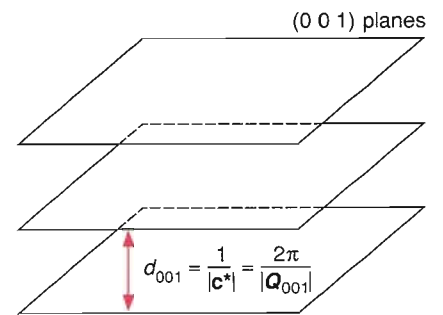
$$Q_h = 2\pi(ha^* + kb^* + lc^*)$$

$$Q_h \perp \text{ to } h \text{ planes}$$

$$Q_{100} = 2\pi a^*$$

$$Q_{010} = 2\pi b^*$$

$$Q_{001} = 2\pi c^*$$

(d) Example of Set of Planes Defined by Unit Cell in **(b)**

(0 0 1) planes

$$d_{001} = \frac{1}{|c^*|} = \frac{2\pi}{|Q_{001}|}$$

would be divided by the lattice spacings to give the components of a dimensionless vector. This shorthand is convenient for many purposes and is used in the references cited at the end of this article.) Equation 2 might appear extremely similar to the Van Hove equation for diffraction, Eq. 2 in the Primer, but it is in fact a kind of inverse. We will cover that relationship in a moment, but first we need to more fully consider what Eq. 2 implies about crystallographic mathematics.

From Eq. 2, we see that the reciprocal-space description of a crystal in three dimensions is based on an array of points defined by the vectors $Q_h/2\pi$ and extending to infinity in all directions from a single origin. These points are the corners of an infinitely repeating reciprocal unit cell. Each amplitude F_h (positive, negative or complex) is associated with the reciprocal-lattice sites $Q_h/2\pi$. A complex amplitude can be represented with the usual real and imaginary parts or as a "phase shift" of the structure factor:

$$F_\mathbf{h} = A_\mathbf{h} + iB_\mathbf{h} = |F_\mathbf{h}|e^{i\alpha_\mathbf{h}}, \tag{3}$$

where $\alpha_\mathbf{h} \equiv \tan^{-1}(B_\mathbf{h}/A_\mathbf{h})$. We will discuss the importance of phase shifts later in this section.

Figures 4 and 5 give examples of two-dimensional periodic scattering densities and their representations in reciprocal space (their Fourier transforms). In two dimensions the reciprocal-lattice vectors are perpendicular to sets of parallel lines (rather than planes) in real space. Larger **h** values correspond to more closely spaced lines.
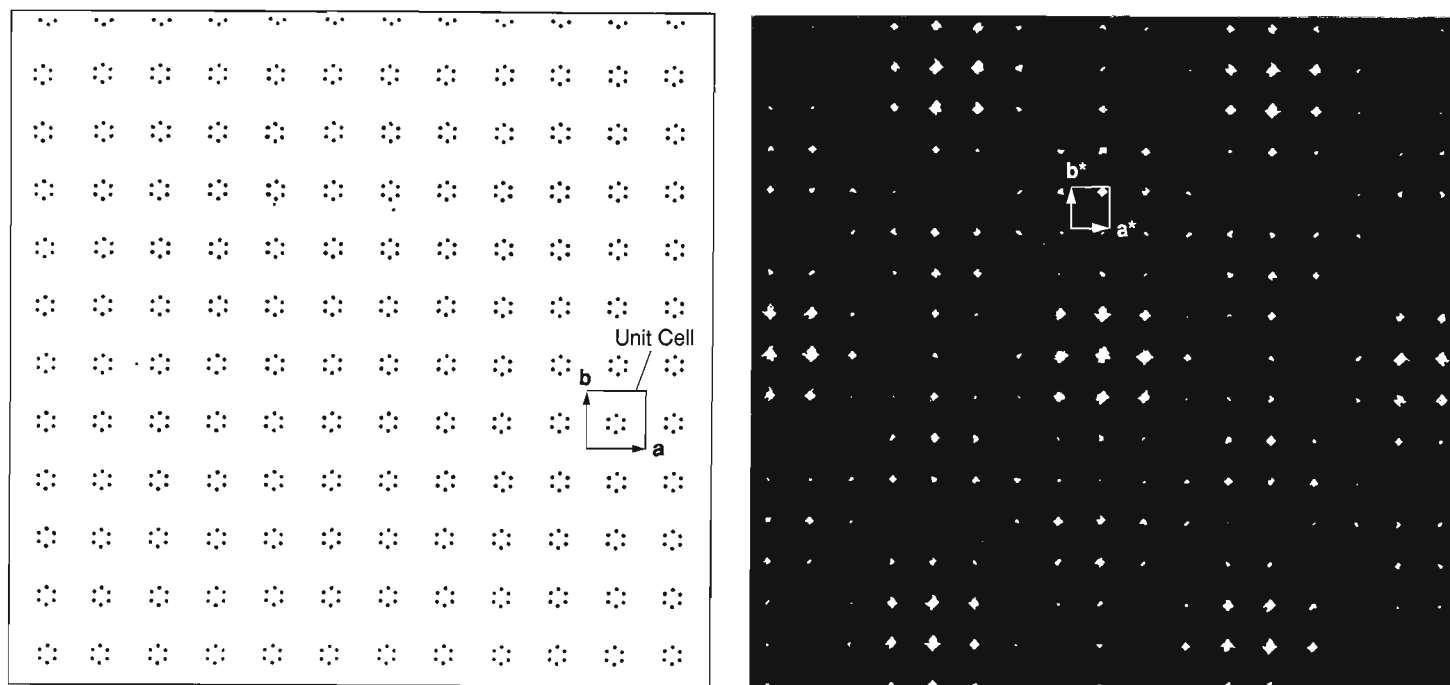
As Figs. 4 and 5b show, the locations of the reciprocal-lattice points are determined solely by the size and shape of the real unit cell, whereas the intensities $|F_h|^2$ reflect the unit cell's contents. The relatively few F_h with small $|h|$ (those closest to the origin) give only the gross features of the structure, that is, the features whose size is roughly on the order of the unit-cell dimensions. The much more numerous F_h with large $|h|$ contain information on the fine details within the unit cell, for example, exact atom locations and anisotropic features of the thermal motion.

Determining Crystal Structures

A property of Fourier series is that they can be "inverted." In crystallography, this transformation goes from the real-space to the reciprocal-space description. Thus, the inverse Fourier transform of Eq. 2 gives the amplitude in terms of the scattering density:

$$F_h = \int_{V_c} \rho(\mathbf{r}) e^{i(\mathbf{Q_h} \cdot \mathbf{r})} d\mathbf{r} = \sum_{i=1}^{n} f_i e^{i(\mathbf{Q_h} \cdot \mathbf{r}_i)}. \qquad (4)$$

The amplitude is first expressed as an integral to indicate that all of the space within the unit cell is used. It is then expressed as a sum by using the convolution properties of Fourier integrals and series. A convolution is a type of multiplication. In this case it is used to break up the integral into a sum over all the atoms in the unit cell; the coefficient f_i of the term for the ith atom is called the "scattering factor" of that atom. The scattering factor of an atom is the Fourier transform of the scattering density in its vicinity and in this expression includes the "smearing" effect of any oscillation (or thermally induced motion) of the atom about its average position. The so-called coherent scattering length $b_{coh,i}$ of neutron scattering is a scattering factor that does not include the effects of thermal motion. (This article deals only with coherent scattering, so in reference to scattering lengths the word "coherent" and the subscript "coh" will be suppressed from now on.) The Fourier transform represented in Eq. 2 implies that if the F_h are known, one can calculate the scattering density $\rho(\mathbf{r})$, which maps the locations and thermal motions of the atoms. Similarly, the transform in Eq. 4 implies that, if the atom positions and thermal motions are known, then the F_h can be calculated.

A SQUARE LATTICE AND ITS RECIPROCAL-SPACE REPRESENTATION

Fig. 4. A lattice in real space (left) with fourfold rotational symmetry and the corresponding reciprocal lattice (right), which has the same point symmetry (see "Crystal Symmetry Groups"). The intensities $|F_n|^2$ depicted on the reciprocal lattice differ from each other because the real-space unit cell has six point scatterers rather than one. Specifically, the pattern of intensities reflects the sixfold symmetry of the contents of the real unit cell. (Photos reproduced with permission from Cornell University Press.)
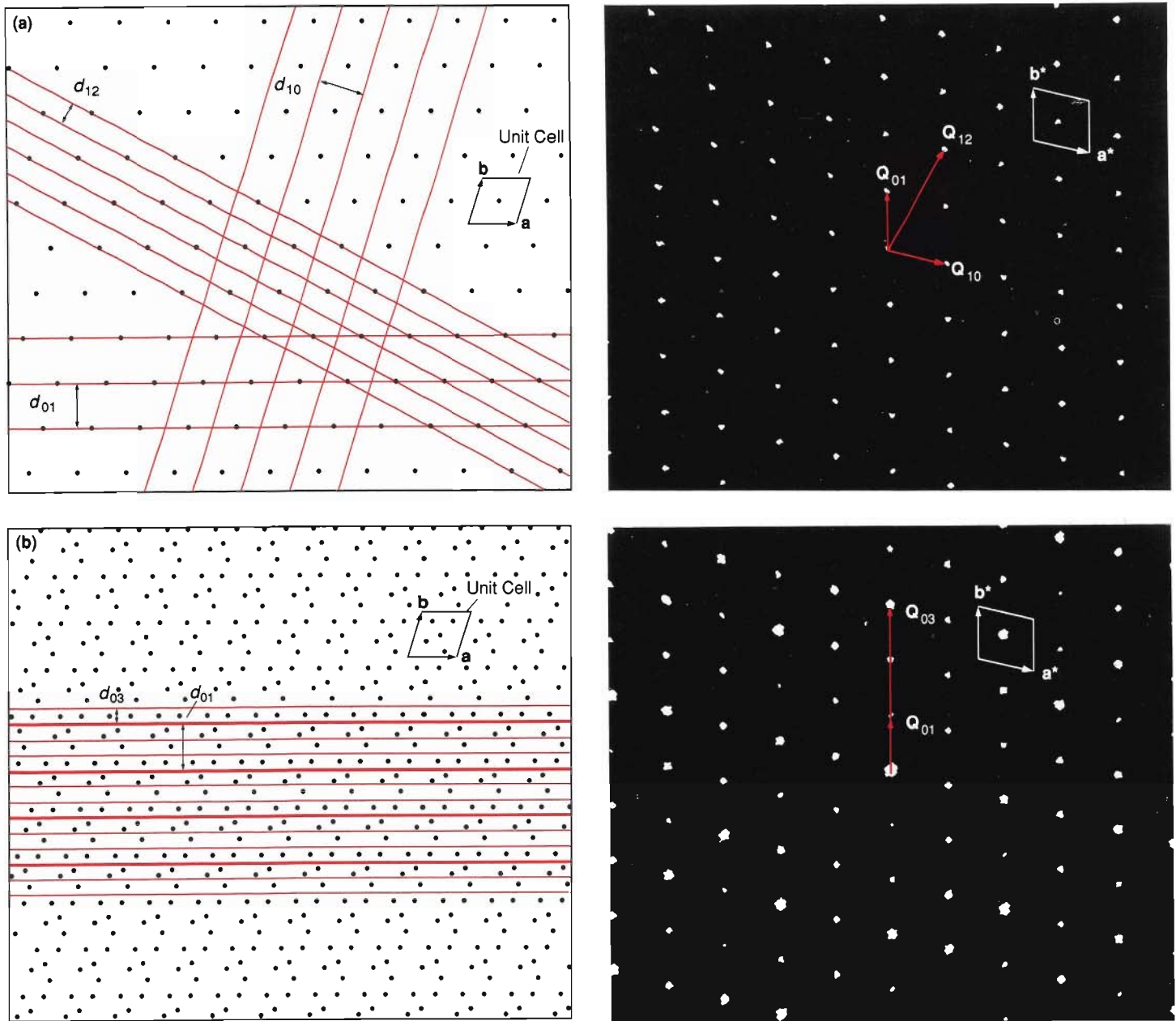
SIMPLE VERSUS COMPLEX UNIT CELL IN REAL AND RECIPROCAL SPACE

Fig. 5. Two-dimensional scattering densities (left) and their corresponding intensity representation in reciprocal space (right). The real-space lattice in (a) is less symmetrical than that in Fig. 4, having two-fold rotation and inversion. The real-space figure shows a unit cell and three sets of parallel lines (the two-dimensional equivalent of parallel planes in three dimensions) with the d-spacings of those lines. Note that the set of lines indexed h = (1 0) is not drawn through the point scatterers, because one need not think of those lines as locations of atoms in a crystal, but rather as defining periodic density variations with a definite orientation and spacing. The other sets of lines are drawn through the point scatterers for clarity. The Q_h vectors corresponding to the labeled sets of lines are shown in reciprocal space (right), along with a reciprocal unit cell. Each Q_h vector is perpendicular to the set of lines that h indexes, and its length $|Q_h|$ is inversely proportional to the d-spacings of that set of lines. Because the unit cell contains only a single point scatterer, the intensities $|F_h|^2$ on the reciprocal lattice are all identical. (b) A real-space lattice in which each point in (a) has been replaced by a five-point pattern or "molecule." This scattering density does not even have inversion symmetry. The points of its reciprocal lattice are in the same positions as those of the reciprocal lattice in (a), but here the intensities $|F_h|^2$ vary. Thus the locations of the reciprocal-lattice points provide information about lattice geometry whereas the intensities provide information about the contents of the unit cell. The heavy lines correspond to h = (0 1); they and the light lines together have h = (0 3). (The reciprocal-lattice figures were made by shining a laser beam through masks with holes punched out at the real-lattice sites and recording the diffracted light on film. With this method one can photograph much of the two-dimensional reciprocal lattice at once. Unfortunately, scattering neutrons analogously from a plane of atoms is not feasible. Neutrons interact so weakly with matter that a beam perpendicular to a single plane of atoms would pass through practically unaffected.) (Photos reproduced with permission from Cornell University Press.)

The Van Hove equation for elastic-scattering intensities given in the primer is the convolution of Eq. 4 with itself:

$$S(\mathbf{Q_h}) \equiv I_\mathbf{h} = F_\mathbf{h} F_\mathbf{h}^* = \int_{V_i} \int_{V_j} \rho(\mathbf{r}_i)\rho^*(\mathbf{r}_j)e^{i\mathbf{Q_h}\cdot(\mathbf{r}_i-\mathbf{r}_j)}d\mathbf{r}_i d\mathbf{r}_j = \sum_{ij} f_i f_j^* e^{i\mathbf{Q_h}\cdot(\mathbf{r}_i-\mathbf{r}_j)}, \quad (5)$$

where $I_\mathbf{h}$ is the measured scattering intensity at $\mathbf{Q} = \mathbf{Q_h}$. Equation 5 shows how elastic-scattering experiments directly measure the $|F_\mathbf{h}|^2$, the squares of the magnitudes of the structure factors. In the last form of the convolution, the double summation covers all interatomic vectors in the crystal, which matches our picture of coherent scattering as arising from interference effects between atoms. The double-integral form also reflects interference effects, since the largest contributions come from those $\rho(\mathbf{r}_i)\rho^*(\mathbf{r}_j)$ products such that $\mathbf{r}_i$ and $\mathbf{r}_j$ correspond to atomic positions. In this formulation of the Van Hove equation, we recognize that both f_i and $\rho(\mathbf{r})$ can be complex and that f_i has a Q dependence, so we keep them inside the integral and the sum. Although b_i and $\rho(\mathbf{r})$ are rarely complex in neutron diffraction, they are quite often complex in x-ray scattering because atoms can absorb x rays and because x-ray wavelengths are comparable to the size of the electron clouds from which they scatter, whereas neutrons scatter from the nearly point-like nuclei. In this discussion we have seen that by starting from a rather mathematical description of a crystal, we can interpret its coherent scattering properties for either x rays or neutrons in a particularly clean way.

The Van Hove equation gives us the relationship between the array of interatomic vectors and the observed intensities but, it also points out a major difficulty. Using Eq. 2 requires knowing both the real and imaginary parts of the $F_\mathbf{h}$, but a diffraction experiment yields only the magnitudes of the $|F_\mathbf{h}|^2$ and not the phases $\alpha_\mathbf{h}$. Without the phases we can not determine the positions of the atoms in the unit cell or even their number. The central problem of crystallography is recovery of the phases, so that the Fourier transform in Eq. 2 can be performed. The solution of this problem, known as solving the crystal structure, is the subject of considerable effort by crystallographers. The reader is encouraged to examine some of the references listed at the end of this article. Figure 2 is a one-dimensional illustration of the ambiguity. A new hypothetical scattering density $\rho(x)$ has been constructed by shifting the phase of the third Fourier component in Fig. 1a by π radians. This shift is equivalent to multiplying that wave's amplitude, F_3, by -1, as seen in the graph of the $F_\mathbf{h}$ in reciprocal space, Fig. 2b. A phase shift by some angle α other than 0 or π is equivalent to multiplying F_3 by the complex number $\exp(i\alpha)$—hard to depict on the page. (Mathematically inclined readers can convince themselves that the $F_\mathbf{h}$'s are real if and only if the unit cell is centrosymmetric, as defined in "Crystal Symmetry Groups.") In any case, the phase shift does not affect the value of $|F_3|^2$ (Fig. 2c). Since diffraction experiments provide only the $|F_\mathbf{h}|^2$, there is, in principle, no way of knowing whether the measured reciprocal lattice arises from the real-space scattering density of Fig. 1a or from the quite different density of Fig. 2a. In practice, the crystallographer realizes from his or her knowledge of physics and chemistry that the density in Fig. 2a makes no sense. Solving the structure of more complicated materials is not so easily done. Then the question is how to directly use the $|F_\mathbf{h}|^2$.

One possible use is to apply a Fourier transform to the Van Hove equation to get a mapping of the interatomic vectors:

$$P(\mathbf{r}_i - \mathbf{r}_j) = \frac{1}{V_c^2} \sum_h |F_\mathbf{h}|^2 e^{-i\mathbf{Q_h}\cdot(\mathbf{r}_i-\mathbf{r}_j)}. \qquad (6)$$

Equation 6 is effectively a map of the $\rho(\mathbf{r}_i)\rho(\mathbf{r}_j)$ product for all vectors $(\mathbf{r}_i-\mathbf{r}_j)$. The transform can be performed with no knowledge of the crystal structure apart from the unit-cell dimensions and point symmetry, which derive directly from the

diffraction data. In crystallographic parlance $P(\mathbf{r}_i - \mathbf{r}_j)$ is known as the Patterson function. It provides one of the routes to solving the crystal structure. Since the $\rho(\mathbf{r}_i)\rho(\mathbf{r}_j)$ product is largest for vectors between strongly scattering atoms, the highest features in the Patterson function correspond to vectors between pairs of such atoms and can generally be interpreted to give their locations. This technique for solving crystal structures, known as the heavy-atom method, is one of the oldest techniques known. It is generally applicable only to x-ray diffraction data for materials whose unit cells are composed of one strongly scattering heavy atom (typically a metal) and

NEUTRON SCATTERING LENGTHS

Fig. 6. Neutron scattering lengths for all the elements from hydrogen through xenon. Every fourth element is marked. Each element is made up of its natural mixture of isotopes. Unlike x-ray scattering factors, neutron scattering lengths do not increase linearly with atomic number. Instead they vary erratically, not only from element to element but from isotope to isotope.
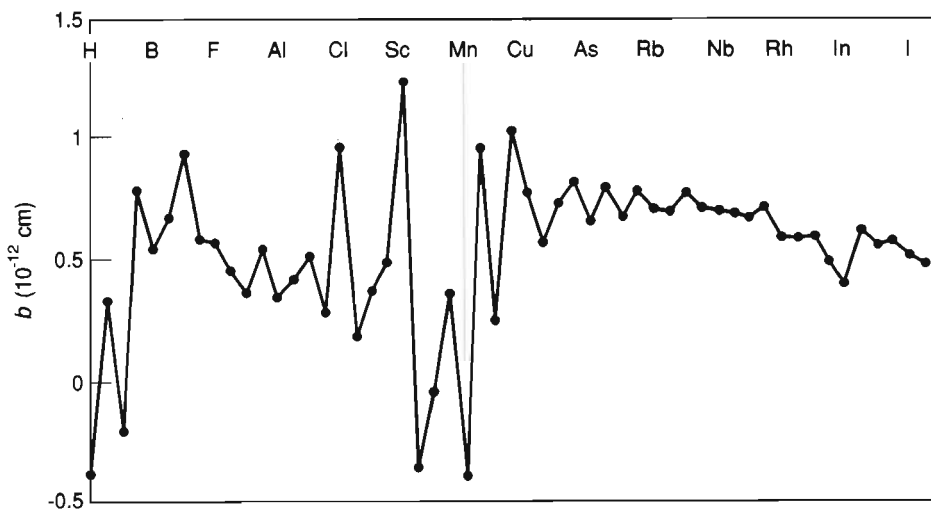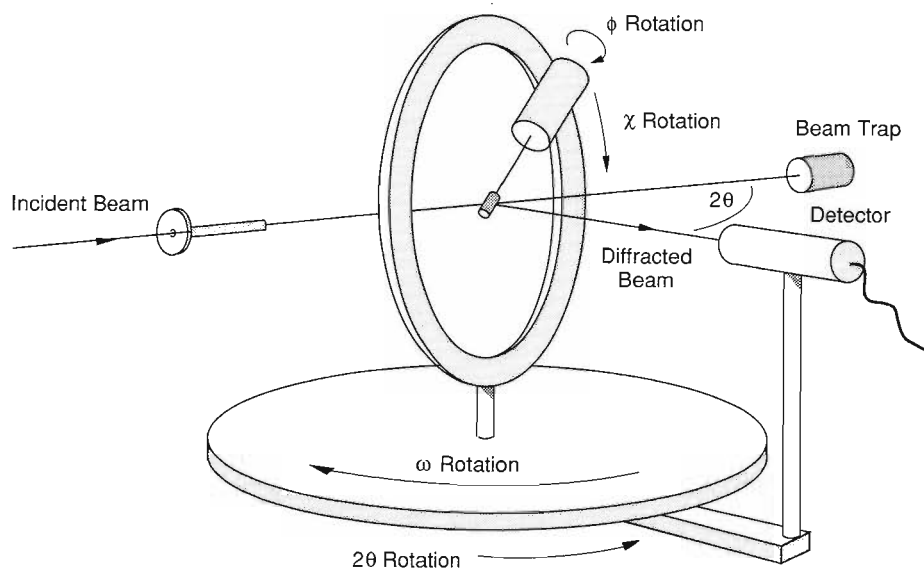

a large number of more weakly scattering light atoms (C, O, N, etc.). The heavy-atom positions can then be applied to Eq. 4 to get an estimate of the phase angles to use in Eq. 2, which produces an approximation to $\rho(\mathbf{r})$. This scattering-density map usually shows enough atom postions to repeat the process and expand the solution to include all the remaining atom positions, thus solving the structure. The heavy-atom method is not so useful in the case of neutron scattering because the scattering lengths of all atoms are the same within an order of magnitude (Fig. 6). The other methods for solving crystal structures have their roots in the properties of the Patterson function and its inverse, the Van Hove equation. The only exception is the time-honored method of just guessing atom positions and using Eq. 4 to judge the accuracy of the guess by comparing the calculated $|F_\mathbf{h}|^2$ with observations.

The Geometry of Diffraction in Reciprocal Space

Let's consider diffraction experiments that use monochromatic beams, that is, those such that all the neutrons or x rays have the same wavelength λ. Then, as noted in the discussion following Eq. 3 of the primer, one can see Bragg peaks only when the planes that produce them are properly oriented with respect to the beam. In a single-crystal diffraction experiment, the crystal is rotated, in an apparatus like that shown in Fig. 7, so as to generate Bragg reflections with various values of $\mathbf{h}$. Then the detector is positioned to measure their $|F_\mathbf{h}|^2$. A powder experiment involves many crystals at once, all randomly oriented. We need to understand the effect of different crystal orientations in both single-crystal and powder experiments.

As we have seen, the $\mathbf{Q_h}$ vectors are perpendicular to sets of planes of the crystal, and the $\mathbf{h}$ vectors correspond to the Miller indices that describe its faces. Thus there is a connection between the physical appearance of a crystal and its reciprocal-space description. Real space and reciprocal space are "hooked" together (see Fig. 3) so that every feature found in real space corresponds to a feature in reciprocal space via Fourier transformation. Therefore when we rotate the crystal, we also rotate its reciprocal lattice.

Figure 8 depicts Bragg scattering in reciprocal space. In terms of the reciprocal-lattice vectors, we can write Bragg's law simply as $|\mathbf{Q}| = |\mathbf{Q_h}|$, using $|\mathbf{Q}| = 4\pi(\sin\theta)/\lambda$ (see Fig. 5a of the Primer) and $|\mathbf{Q_h}| = 2\pi/d_\mathbf{h}$ where $d_\mathbf{h}$ is the spacing of the planes labeled by $\mathbf{h}$. Thus depicting the elastic-scattering triangle (again in Fig. 5a of the primer) in reciprocal space provides a useful geometric construction (here we multiply all the reciprocal-lattice distances by 2π in order to compare them with wave vectors). Since for elastic scattering the initial and final wave vectors are equal in magnitude, or $|\mathbf{k}_i| = |\mathbf{k}_f|$, all the possible $\mathbf{k}_f$'s fall on the surface of a sphere with ra-

φ Rotation

χ Rotation

Beam Trap

Incident Beam

2θ

Detector

Diffracted Beam

ω Rotation

2θ Rotation

## SINGLE-CRYSTAL DIFFRACTOMETER

Fig. 7. A single-crystal diffractometer with three axes for positioning the crystal in the incident beam so that a particular set of planes in the crystal scatters the incident radiation in the plane containing the detector arm. The detector angle can be set at the proper 2θ to observe the reflection (once a preliminary x-ray experiment has determined the orientation and size of the reciprocal unit cell).

dius $2\pi/\lambda$, commonly called the Ewald sphere. As seen in Fig. 8, the Ewald sphere passes through the origin because $\mathbf{Q} = 0$, or equivalently $\mathbf{k}_i = \mathbf{k}_f$, corresponds to elastic scattering in the forward direction. The $\mathbf{Q_h}$ of a reciprocal-lattice point that falls on the Ewald sphere is equal to the $\mathbf{Q}$ at which Bragg scattering occurs; the scattering angle 2θ between $\mathbf{k}_i$ and $\mathbf{k}_f$ is shown on the figure. The construction makes it easy to see the effect of rotating the crystal. When the crystal rotates, the reciprocal lattice rotates with it so that each of its points moves on an arc centered at the origin. As each point passes through the Ewald sphere, diffraction occurs for that $\mathbf{Q_h}$ at the corresponding scattering angle θ. This is the basis for conventional single-crystal diffraction experiments. The intensity of the scattering seen by the detector when $\mathbf{Q_h}$ passes through the Ewald sphere is proportional to $|F_\mathbf{h}|^2$, but it also depends on the angle that the $\mathbf{Q_h}$ arc makes with the sphere surface (the so-called Lorentz correction to the intensities).

Because a powder consists of a multitude of small crystals, the reciprocal-space picture has to be modified from that given for a single crystal. Instead of an array of points, the $\mathbf{Q_h}$ vectors define a set of nested spheres, each one corresponding to the multitude of directions that each $\mathbf{Q_h}$ points for all the crystals that make up the powder (Fig. 9). Then the orientation of the powder sample is immaterial, and the Ewald sphere for the illuminating radiation intersects all the $\mathbf{Q_h}$ spheres with $|\mathbf{Q_h}| < 4\pi/\lambda$. Thus diffraction occurs simultaneously at a variety of angles. The observed intensity again depends on $|F_\mathbf{h}|^2$, and the Lorentz correction depends on the angle at which the Ewald sphere and the $\mathbf{Q_h}$ sphere intersect. In addition, the crystal symmetry may require that related $\mathbf{h}$ vectors have the same length and therefore that their respective spheres exactly coincide. For example, in the cubic crystal structure for salt (NaCl), the $\mathbf{Q_h}$ vectors with $\mathbf{h} = (1\,0\,0), (0\,1\,0), (0\,0\,1), (\bar{1}\,0\,0), (0\,\bar{1}\,0)$, and $(0\,0\,\bar{1})$ all have identical lengths and identical $F_\mathbf{h}^2$ values. The measured intensity at the corresponding angle is proportional to $6|F_{100}|^2$; the factor of 6 is the reflection multiplicity. A powder pattern then contains all the intensity information inherent in the reciprocal

SINGLE-CRYSTAL DIFFRACTION IN RECIPROCAL SPACE

Fig. 8. A reciprocal-space representation of single-crystal diffraction of monochromatic radiation of wavelength λ. A sphere of radius $|k_i| = 2\pi/\lambda$ is drawn through the origin of the reciprocal lattice–the "Ewald sphere." Since diffraction is an elastic process, the wave vectors of the incident and scattered radiation, k_i and k_f, have equal length and can be drawn so that they are radii of the sphere. A few of the points of the reciprocal lattice touch the surface of the Ewald sphere. Note that the Q_h vector for each such point coincides with Q, the third side of the scattering triangle. In other words, $Q=Q_h$ (a version of Bragg's law) and diffraction occurs at the angle 2θ between k_i and k_f. Rotation of the crystal corresponds to rotating the reciprocal lattice, causing other points to touch the Ewald sphere and diffraction to occur at other scattering angles.
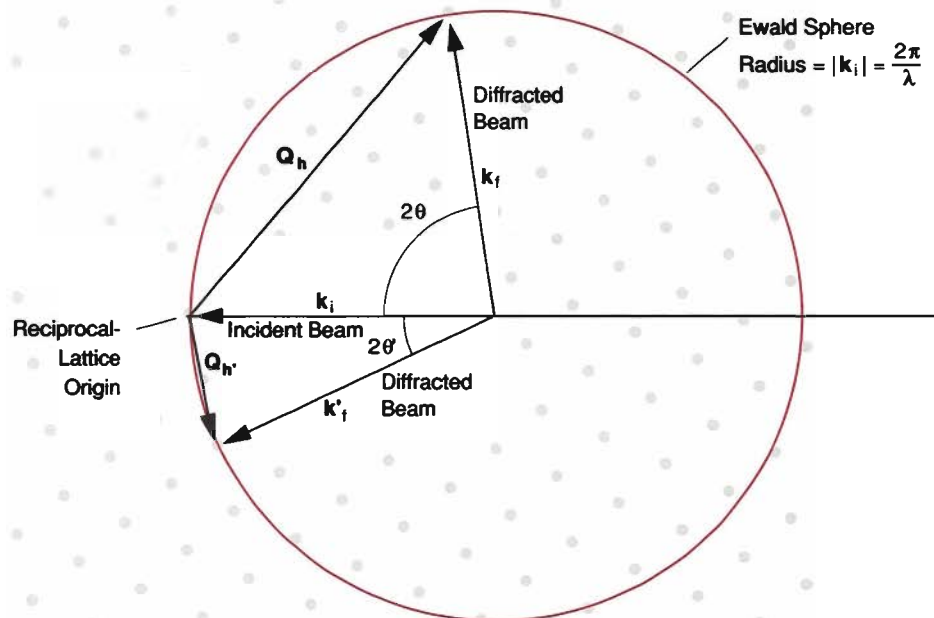
lattice, but all the directions of the vectors are lost along with the phases of the structure factors. This situation was neatly summarized long ago by W. H. Bragg in his 1921 Presidential Address to the Physical Society.

All the spectra of the different planes are thrown together on the same diagram, and must be disentangled. This is not so difficult as it might seem. ... The spectra of the organic substances show how very diversified they are, and illustrate the power of a method of analysis which promises to be of great use, since every crystal has its own characteristic spectrum.

Despite Bragg's optimism about interpretation of a powder pattern, only recently has there been any real progress in powder pattern analysis. The classical use of x-ray powder patterns has been analytical, to enable identification of crystalline phases in an unknown mixture. This is usually achieved by matching the line positions and relative intensities against a compendium of such values obtained by measuring patterns of pure materials. Commercial x-ray powder diffractometers come with software packages that do the matching automatically. The computer file of standard materials maintained by the Joint Commission on Powder Diffraction Standards now contains over 50,000 entries. Our problem, however, is to unravel a powder pattern and extract the crystal structure responsible for the observed intensity distribution.

Crystal Structures from Powder Patterns

For a long time the only way a powder pattern could be interpreted to give the crystal structure was a variation on the methods used for single-crystal diffraction data. The first step consists of identifying the vectors h (or h, k, l), that give rise to the peaks in the pattern thus identifying the crystal lattice and its parameters a, b, and c. This process, known as indexing the pattern, can be complex for low-

(a) Reciprocal-Space Geometry
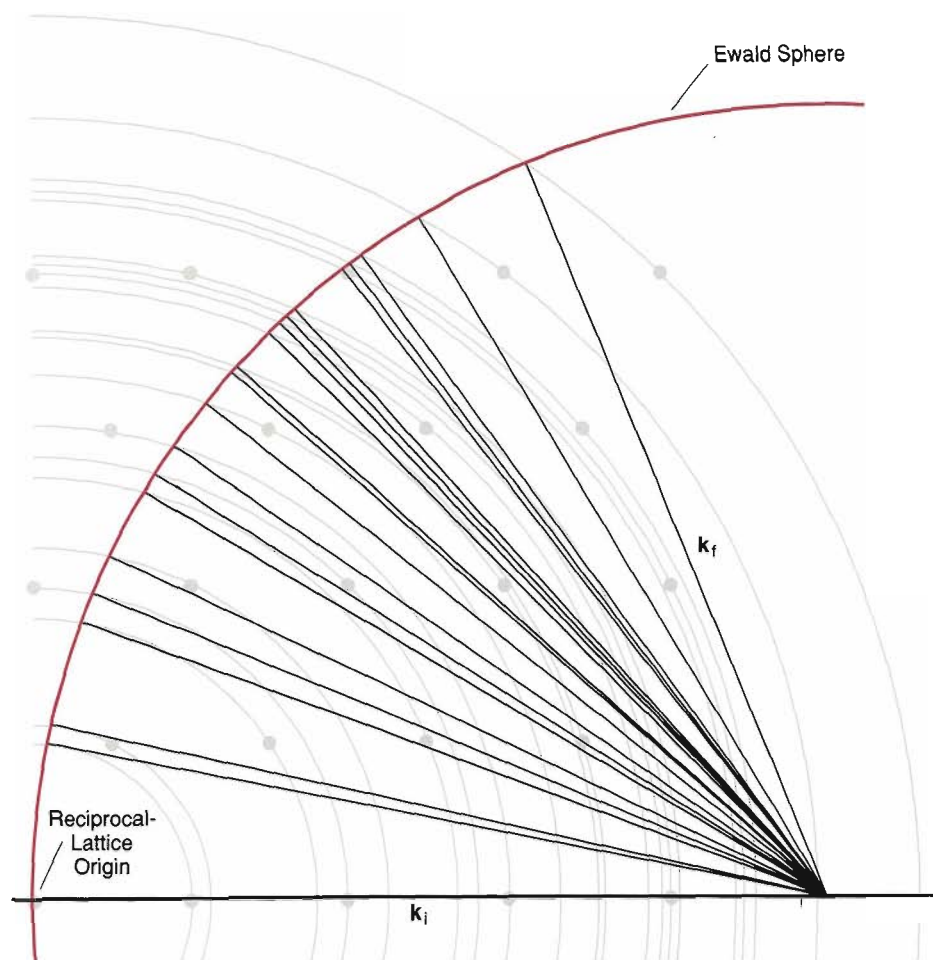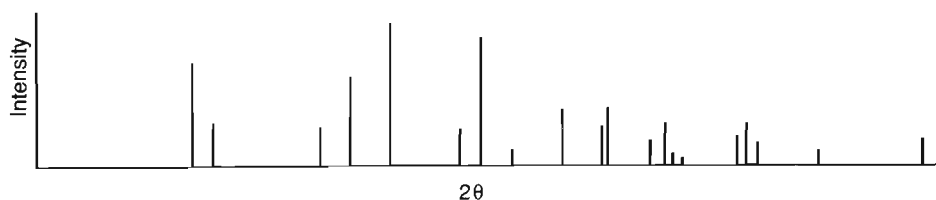
POWDER DIFFRACTION IN RECIPROCAL SPACE

Fig. 9. (a) A reciprocal-space representation of powder diffraction of monochromatic radiation. The reciprocal-lattice points for a powder are smeared out onto the surfaces of a nested set of spheres, all of which intersect the Ewald sphere if $d_h \geq \lambda/2$. Thus diffraction from all planes whose d-spacing is greater than or equal to $\lambda/2$ can be recorded in a single measurement with single-wavelength radiation. (b) The powder diffraction pattern that would be recorded from a crystal having this reciprocal lattice. For clarity, the intensities are shown here but are not plotted on the reciprocal lattice in (a) (whereas they were in Figs. 4 and 5).

(b) Powder-Diffraction Pattern

symmetry crystals but is quite easy for cubic structures. In that case the relationship between **h** and d-spacing gives

$$d^2 = \frac{a^2}{h^2 + k^2 + l^2} \tag{7},$$

where a is the cubic-lattice spacing. One need only number the peaks starting from the origin, skipping those numbers that are not sums of three squares of integers (7, 15, 23, 28, etc.), and then tally up the possible hkl combinations for each peak. Then the intensities of individual peaks are measured and converted to structure factor magnitudes. These could then be used to "solve" the structure (remember this is a puzzle because of the lost complex character of F_h). The main problem with this technique is that only for very simple structures are the peaks in a powder pattern sufficiently separated to allow measurement of individual peak intensities. One can index the pattern of almost any substance and thus find a description of the lattice. However, the peaks are usually so heavily overlapped that extraction of individual peak intensities is impossible, and the magnitudes of most of the individual structure factors are unknown.

About twenty years ago H. M. Rietveld suggested a solution to this problem. He recognized that a mathematical expression could be written to represent the observed intensity I_c at every position $\mathbf{Q}$ in a powder-diffraction pattern:

$$I_c(\mathbf{Q}) = I_b(\mathbf{Q}) + \sum I_h(\mathbf{Q}). \qquad (8)$$

This expression has a contribution from the background and from each of the $\mathbf{Q_h}$ that are in the vicinity of $\mathbf{Q}$ (Fig. 10). Unlike the stick-diagram representation of a pow-

OVERLAPPING PEAKS IN A POWDER DIFFRACTION PATTERN

Fig. 10. A small segment of a time-of-flight powder-diffraction pattern showing the total intensity and contributions to it from background and from several Bragg reflections. Note that four Bragg reflections contribute to the left-most observed peak.
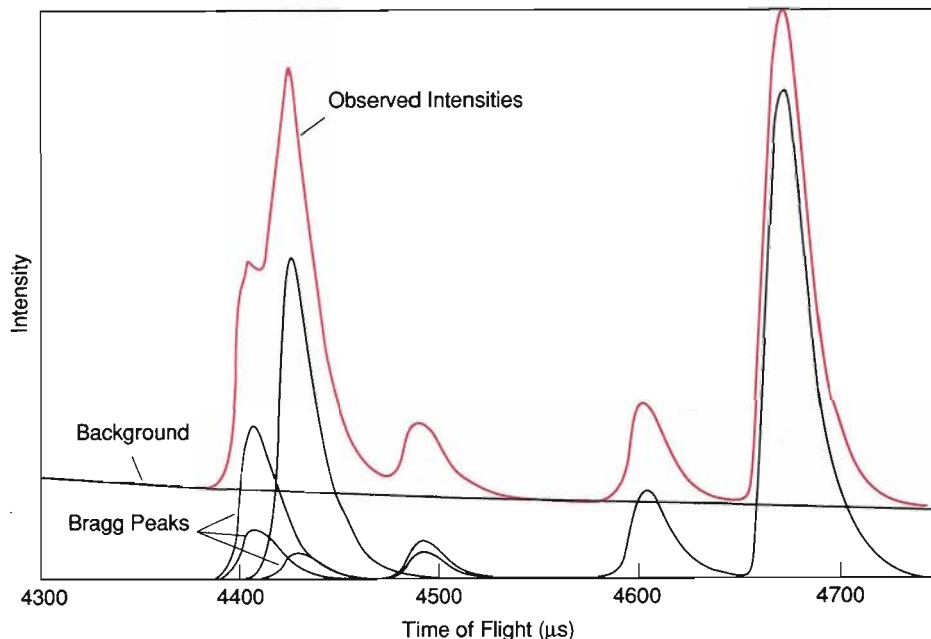
der pattern shown in Fig. 9, a real powder pattern suffers from line broadening, so diffraction from the planes labeled by $\mathbf{h}$ contributes not only at $\mathbf{Q_h}$ but at all nearby $\mathbf{Q}$. The pattern in Fig. 10 also exhibits line anisotropy, which arises from the asymmetry of the spallation-neutron pulse. In the Rietveld method one models the observed pattern by considering the factors that affect both the line shape and its intensity. The adjustable parameters for the model are then refined by a nonlinear least-squares process that is similar to the process very commonly used in single-crystal x-ray structure analysis. The $|F_\mathbf{h}|^2$ parameters obtained from the fit are a reconstruction of the real $|F_\mathbf{h}|^2$; the parameters for line broadening and anisotropy provide information about particle sizes, structural defects, and other phenomena that distort the ideal Bragg pattern.
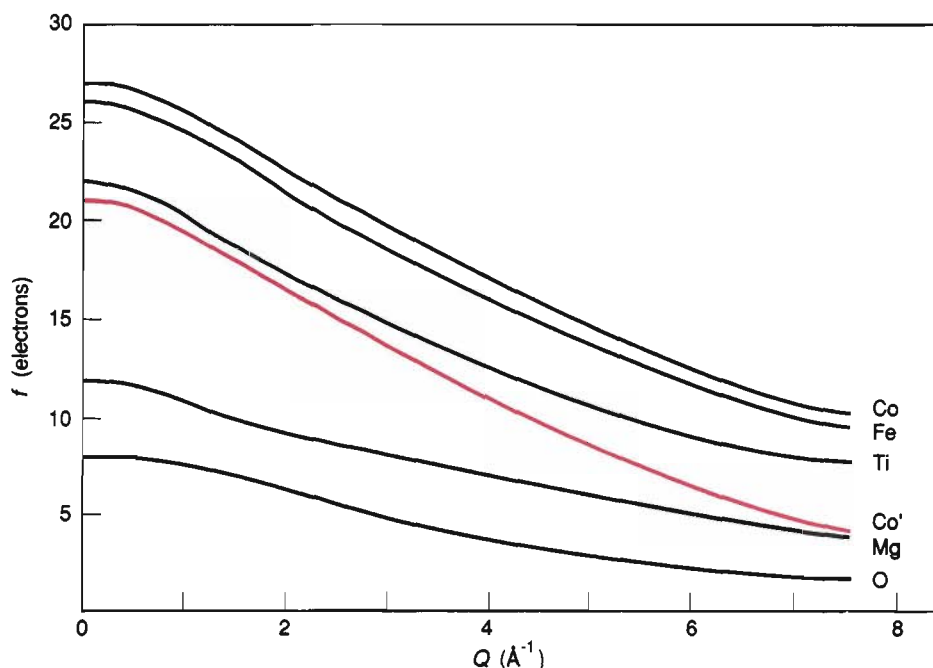
This approach has been so successful that it has led to a renaissance in powder diffraction, and this technique of treating powder-diffraction data is now known as Rietveld refinement.

How Are X-Ray and Neutron Diffraction Complementary?

In generating a model to perform the inverse Fourier transform shown in Eq. 4, we postulate a set of atom positions and assign a scattering factor f_i to each atom, which is the Fourier transform of the scattering density about its position. However, because x rays and neutrons scatter by different mechanisms, the corresponding scattering factors are quite different. Neutrons are scattered primarily by atomic nuclei. Since the nuclear dimensions are roughly 100,000 times smaller than the neutron wavelength, the nuclei act like point scatterers and neutron scattering factors (scattering lengths or b's) are independent of $|\mathbf{Q}|$. Also, nuclear scattering is a combination of "potential" scattering and "resonance" scattering. Potential scattering depends on

the number of nuclear particles and resonance scattering results from neutron absorption by the nucleus. These two factors sometimes add and sometimes subtract to give neutron scattering lengths that vary erratically from one element to another and from one isotope to another (see Fig. 6).

On the other hand, x-ray scattering occurs primarily by interaction with the electrons that surround an atom. Consequently, the strength of the scattering depends on the number of electrons that surround an atom so that the scattering power of an atom increases with atomic number. Thus, x-ray scattering factors are usually ex-

pressed as some multiple of the scattering power of one electron. In addition, the spatial extent of the electron cloud around an atom is roughly the same as the x-ray wavelength, so the x-ray scattering factor falls off with increasing $|\mathbf{Q}|$. The scattered intensity also has a contribution from anomalous dispersion when the x-ray energy is near an absorption edge for the scattering atom. The absorption edge for an inner electron shell of an element is the minimum energy at which an atom can absorb an x ray and consequently eject an electron from that shell. The scattering factor can be strongly modified by this process and acquire both real and imaginary components that are only partially dependent on $|\mathbf{Q}|$. Thus the scattering factors for x rays look like those shown in Fig. 11.

Given these differences we would expect x-ray and neutron powder-diffraction patterns to be very different. Figure 12 shows idealized x-ray and neutron patterns calculated for $MgTiO_3$, the primary constituent of the mineral geikielite. The patterns were generated for essentially identical diffractometer experiments (impossible in real life) but are startlingly different. In fact, the strongest peak in the x-ray pattern (at about 32°) is completely absent in the neutron pattern! The reason for the extreme difference between the two patterns lies in the scattering factors for titanium, magnesium, and oxygen for x-rays and neutrons. The x-ray scattering factors are simply proportional to the atomic number; thus $f_{Ti} > f_{Mg} > f_O$. However, the neutron scattering length of titanium is negative and that of oxygen is only slightly larger than that of magnesium, or $b_O > b_{Mg} > 0 > b_{Ti}$. Therefore the neutron structure factor for each reflection is very different from the x-ray structure factor, and the peak heights in the two powder-diffraction patterns are very different.

The complementarity of x-ray and neutron powder patterns then eliminates one of the most basic problems in crystal-structure analysis. Because the complex nature of the structure factors is lost in any diffraction measurement and the directional

character of reciprocal space also is lost in a powder-diffraction experiment, the Rietveld refinement of a single powder pattern may not yield a unique answer. Clearly, if a crystal-structure model of atom positions, etc., produces calculated patterns that match both a neutron powder pattern and an x-ray powder pattern, that model is more likely to be unique (and correct). To capitalize on this notion we have developed a computer program that will perform a combined x-ray and neutron Rietveld refinement of a crystal structure. The remainder of the article presents some applications of this approach.

## DIFFERENCES BETWEEN X-RAY AND NEUTRON POWDER PATTERNS

Fig. 12. A comparison of simulated x-ray and neutron powder patterns for $MgTiO_3$ at a wavelength of 1.54 Å. The neutron scattering lengths of Mg, Ti, and O are very different from their x-ray scattering factors, so peaks that are prominent in one pattern are small or even invisible in the other. Thus the two patterns give different information about the structure of the crystal.

(a) Neutron Powder Pattern **(b)** X-ray Powder Pattern
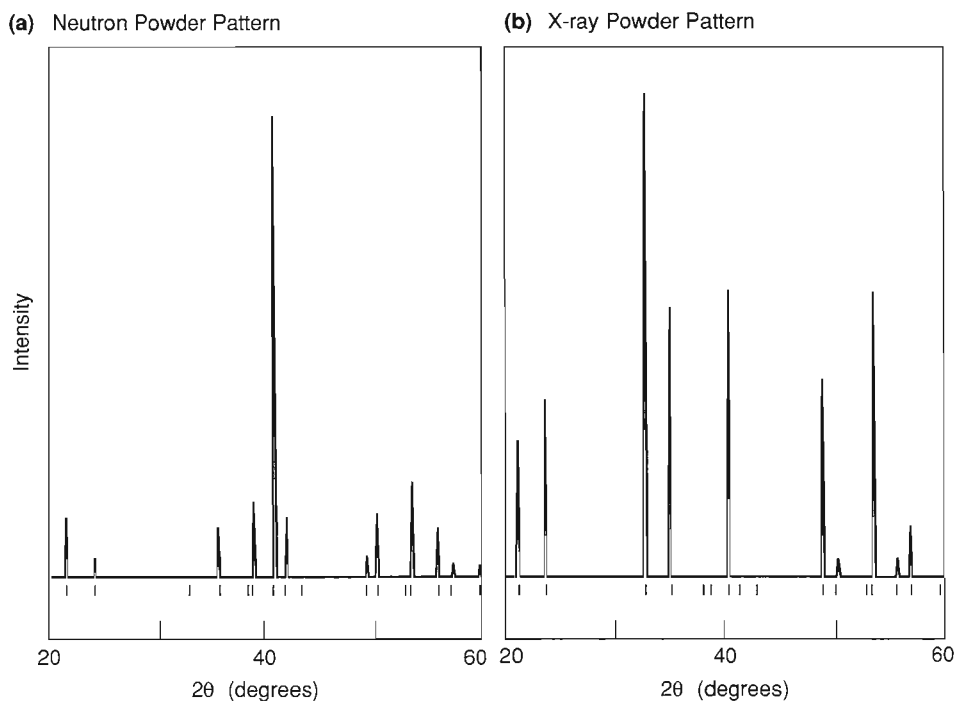
Examples of Combined Rietveld Refinements

One of the most difficult structural problems is to determine the identity of the atoms that occupy a particular site within a crystal structure. Generally an atom is identified by its scattering power relative to the other atoms in the structure. In addition, the distances between it and its nearest neighbors also help in this identification process. The large body of structural work in the literature provides the expected interatomic distances for a particular pair of atoms. The problem becomes much more difficult, however, when more than one kind of atom can occupy a particular site.

The 123 High-Temperature Superconductor. Our first example of an atom-identification problem concerns the high-T_c 123 superconductor $YBa_2Cu_3O_{7-x}$. This material had been investigated at great length by many groups throughout the world, and its structure had been established with little ambiguity within a few months of its discovery by Chu and coworkers at the University of Houston. Almost all structural results came from Rietveld refinements of neutron powder-diffraction data obtained at either reactor or spallation sources, and the atom identities were assigned largely by analogy to other structures as well as by their scattering powers. By unfortunate coincidence the neutron scattering lengths of yttrium and copper are virtually identical, leaving open the possibilities that the assignments of these two atom locations were in fact reversed or that each site was sometimes occupied by yttrium and sometimes by copper. Either case would have considerable impact on any theory proposed to explain the superconductivity. However, the x-ray scattering factors for these two atoms are very different, and by combining some x-ray powder data with the neutron data one can easily resolve this ambiguity.
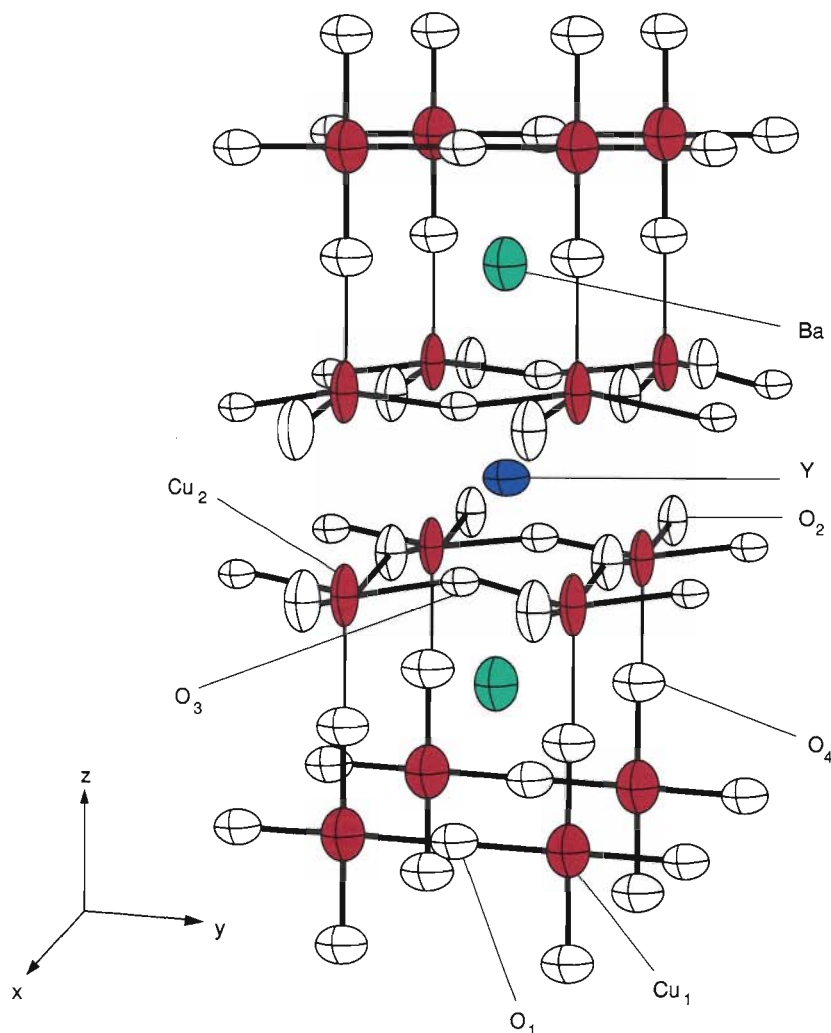
A HIGH-T_c SUPERCONDUCTOR

Fig. 13. A perspective drawing of the structure of $YBa_2Cu_3O_{7-x}$ as deduced from a combined x-ray and neutron Rietveld refinement. The ellipsoids represent the extent of 99% of the atomic thermal motion. The Cu and O atoms are labeled with numbers to distinguish inequivalent locations. Combined x-ray and neutron data proved that there is no interchange between the Cu atoms at these sites and the Y atoms at the site in the middle of the figure.

Here at LANSCE we performed both time-of-flight neutron and x-ray powder-diffraction experiments, collecting six powder patterns for this material The entire set, comprising about 25,000 data points, was subjected to a combined Rietveld refinement involving approximately 120 adjustable parameters. These parameters included the 33 needed to describe the crystal structure of $YBa_2Cu_3O_{7-x}$, namely, atomic positions, fractional occupancies, thermal parameters, and lattice parameters. The rest characterized details of the powder-diffraction patterns and included coefficients for the background, the peak shapes, and intensity correction factors as well as the six scaling factors. The resulting structure, shown in Fig. 13, was dramatically more precise than any of the previous single-measurement results and satisfactorily resolved the metal-site occupancy issue. We found no evidence of any interchange between the metal atoms on their respective sites. Our result had been expected from crystal-chemistry considerations based on comparison of interatomic distances and ionic radii, but this work provided a clear and unambiguous determination.

Vanadium-Doped Iron-Cobalt Alloy. The atom-identification problem in our second example is considerably more difficult. The alloy FeCo is well known as an excellent soft ferromagnet with a high saturation magnetization and low permeability and is of great use commercially. To improve its machinability, a small amount of vanadium (about 2%) is added to the alloy. This alloy is also a well-known example of a second-order β-brass transition: At high temperatures the two metals occupy sites of the body-centered cubic structure at random, but below 720°C the alloy orders so that atoms of the two metals tend to occupy alternate sites (Fig. 14). It had been presumed that in the low-temperature phase the vanadium atoms randomly oc-

cupy either the Fe site or the Co site in the structure, but no definitive determination had been attempted. Because the Fe/Co ordering is incomplete, solution of the problem required determining the fraction of the Fe-rich and the Co-rich sites occupied by Fe, Co, and V. This is not possible with a single-radiation experiment and is extremely difficult with a two-radiation experiment. Conventional x-ray powder data is particularly insensitive to the ordering because Fe and Co have essentially the same x-ray scattering factors and that for V is only about 15% smaller. In fact the β-brass transition for this material is virtually invisible to x rays. We resolved this site-

ORDER-DISORDER TRANSITION IN THE ALLOY FeCo

Fig. 14. A schematic representation of the β-brass transition in FeCo, the iron-cobalt alloy containing equal quantities of Fe and Co. In the high-temperature form on the left, each site is occupied at random by either Fe or Co. The alloy becomes ordered below 720° C to form the structure on the right. Each Fe atom is surrounded by Co atoms, just as the Co atom in the figure is surrounded by Fe atoms.

Fe/Co

Fe

Co

Above 720°C

Below 720°C
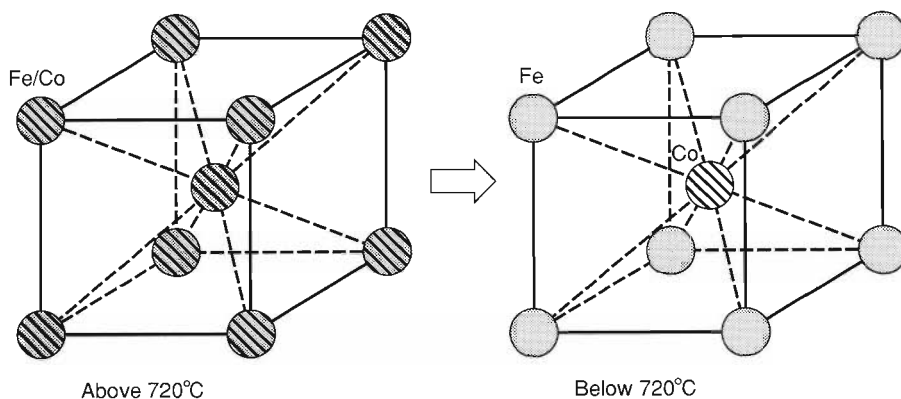
distribution problem by collecting neutron time-of-flight and shortwavelength x-ray data sets that covered a range of $\mathbf{Q}$ sufficient to independently determine the thermal-motion parameters. In addition, we performed synchrotron x-ray experiments at Stanford University. The tunability of synchrotron x-ray radiation allowed us to strongly modify the x-ray scattering factors of the three metals by collecting powder data near each of their respective K absorption edges. The strong anomalous dispersion reduced the x-ray scattering factor for each metal in turn by about 6 to 8 electrons from the dispersionless value given in Fig. 11 and thus provided sufficient contrast between that atom and the others. The entire suite of data, consisting of some 18 powder patterns with a total of about 22,000 data points, was subjected to a combined Rietveld refinement to determine the fractional occupancies for Fe, Co, and V at the two sites. The result clearly showed that the V strongly preferred the Co-rich site over the Fe-rich site and that for this particular sample the Fe/Co ordering was $\sim 80\%$. These results are not obtainable by any other means.

T*-Phase High-Temperature Superconductors. Our final example is an extension of the idea used for the FeCo alloy. The problem is to determine the site preferences of the strontium, lanthanum, and rare-earth ions in the so-called T^*-phase superconductors. These materials have the general formula $La_{1.8-x}R_x Sr_{0.2}CuO_4$, where R is a rare-earth metal, and all have approximately the same structure. They have been synthesized with all the rare earths between Pr and Ho as R. Only those with Sm, Eu, and Gd and $x \simeq 0.9$ form bulk superconductors and then only when annealed at high O_2 pressures. As shown in Fig. 15, one end of the T-phase unit cell resembles the K_2NiF_4-like structure of La_2CuO_{4-x} (called the T-phase), the first high-T_c material to be discovered (by Bednorz and Müller). The other end of the unit cell resembles the structure of the so-called T' superconducting phase, Ce-doped Nd_2CuO_4. Each end has sites for the Sr, La, and R ions; the sites at the T-phase-like end are larger than those at the T'-like end. Consideration of the various atomic radii and the metal-oxygen distances for the two types of sites had led to the assumption that the larger ions (Sr^{2+} and La^{3+}) occupy the larger T-phase sites, whereas the smaller rare-earth ions and the remaining La^{3+} occupy the smaller T'-phase sites. We examined the superconductor $La_{0.9}{}^{152}Sm_{0.9}Sr_{0.2}CuO_4$ with neutrons at LANSCE and with synchrotron radiation at the National Synchrotron Light Source at Brookhaven

National Laboratory. We tuned the synchrotron radiation to absorption edges of La, Sm, and Sr to sufficiently modify their scattering factors. We also collected a fourth x-ray data set at a wavelength far from any edges. We found that the ordering exists but is incomplete: About 10% of the Sm^{3+} ions appear on the larger T site, presumably because the ions are nearly the same size, with a Sm^{3+}/La^{3+} ionic-radius ratio of 0.935. We also examined another superconductor, $La_{0.9}Gd_{0.9}Sr_{0.2}CuO_4$, in which the ions are of more disparate sizes ($Gd^{3+}/La^{3+} = 0.919$). The strong absorption by Gd precluded collection of a neutron powder pattern, but the four x-ray data sets were

T' Phase T Phase T^* Phase
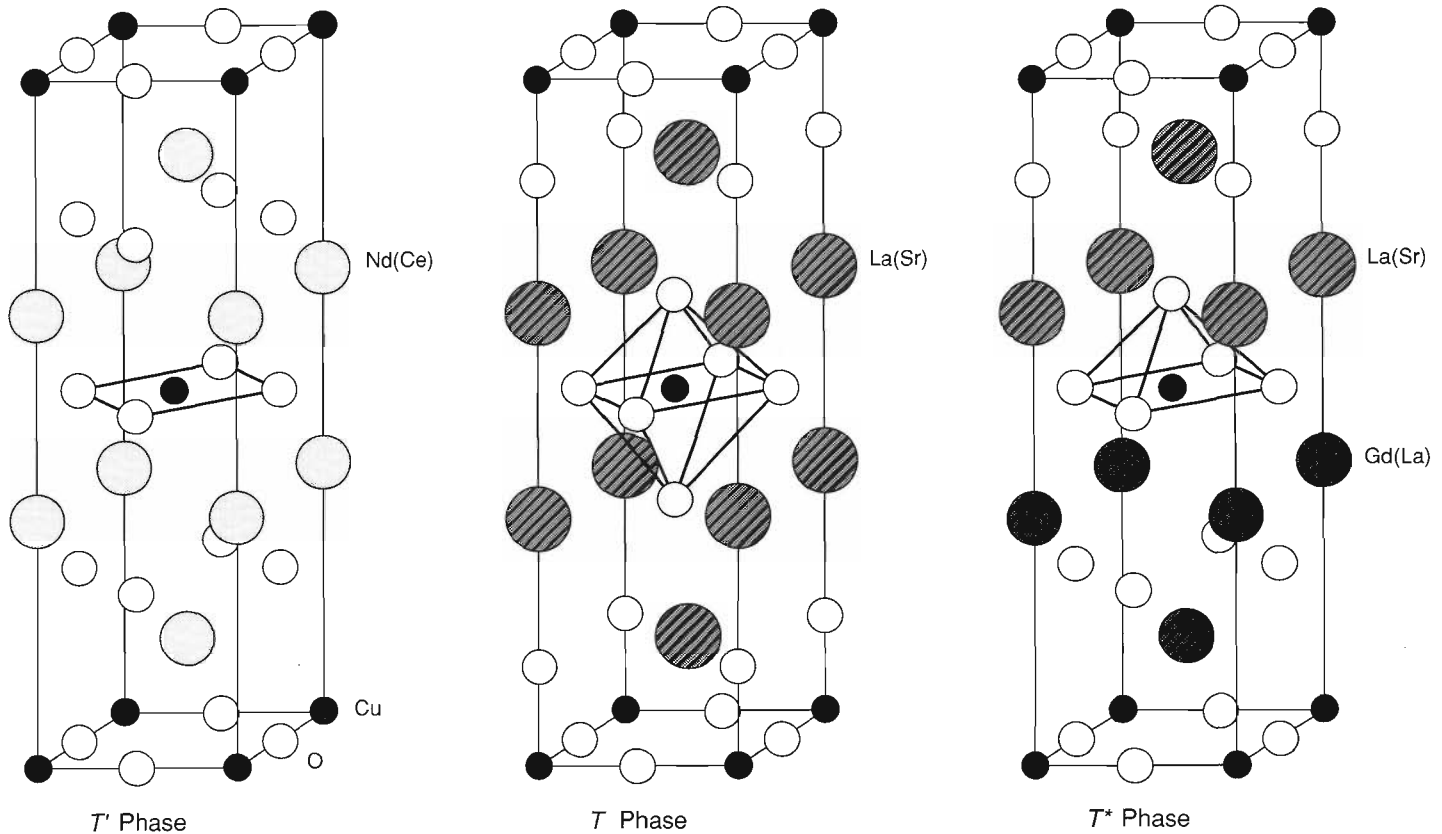
sufficient to unambiguously determine the two site distributions for the three kinds of atoms. In this case the ions were well segregated into the two sites by their size. Since the two materials have similar superconductivity properties, this ordering evidently has little effect on the superconductivity.

Conclusion

As one can see from this discussion, the science of powder diffraction has come a long way from its beginnings as a largely analytical tool. The Rietveld refinement technique has enabled the determination of crystal structures of considerable complexity and in fact was the first to accurately reveal the structures of the superconducting copper oxides. The power of this method can be further enhanced by proper combination of diffraction data from several radiation sources to improve the interatomic contrast and eliminate the ambiguities in powder structure refinements. ∎

T^*-PHASE SUPERCONDUCTOR

Fig. 15. The structure of the T^*-phase superconductor (right) combines those of the T' and T phases, two other structures of M_2CuO_4 (where M can be a lanthanide element or Sr in the compounds of interest for superconductivity). The T' phase (left) has Ce-doped Nd on the M sites (the parentheses around the "Ce" symbol indicate that the amount of Ce is much less than the amount of Nd). The T phase (middle) has larger sites for M, which are occupied by Sr-doped La. The T^* phase has La, Sr, and Gd on the M sites. The larger La and Sr ions occupy the sites in the top half of the unit cell, which are identical to those in the T phase. The smaller Gd (dark gray) ions and the rest of the La occupy the smaller T' sites in the bottom half.

Crystal Symmetry Groups

Symmetry plays an important role in crystallography. The ways in which atoms and molecules are arranged within a unit cell and unit cells repeat within a crystal are governed by symmetry rules. In ordinary life our first perception of symmetry is what is known as *mirror symmetry*. Our bodies have, to a good approximation, mirror symmetry in which our right side is matched by our left as if a mirror passed along the central axis of our bodies. Our hands illustrate this most vividly; so much so that the image is carried over to crystallography when one speaks of a molecule as being either "right"- or "left"- handed. Those of us who live in an old-fashioned duplex will also recognize that such houses are built with mirror symmetry so that the arrangement of the rooms, hallways, and doors are disposed about an imaginary mirror passing through the common wall between the two halves of the house. There are many other examples of this kind of mirror symmetry in ordinary life. We can also see more complex symmetry in the patterns around us. It can be found in wallpaper patterns, floor-tile arrays, cloth designs, flowers, and mineral crystals. The basic mathematics of symmetry also applies to music, dance (particularly folk and square dance), and even the operations needed to solve Rubik's cube.

The rules that govern symmetry are found in the mathematics of group theory. Group theory addresses the way in which a certain collection of mathematical "objects" are related to each other. For example, consider all the positive and negative integers and zero. They can constitute a group because under certain circumstances the relationships between the integers obey the rules of group theory:
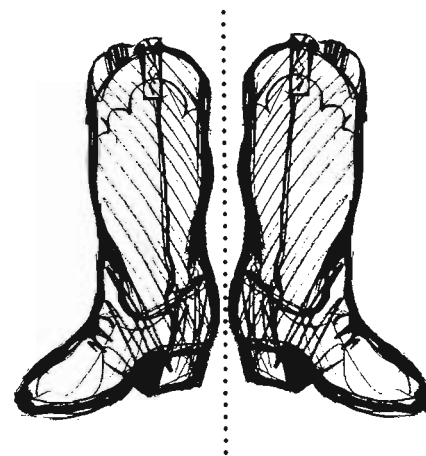
- There must be defined a procedure for combining two elements of the group to form a third. For the integers one can choose the addition operation so that $a + b = c$ is the operation to be performed and a, b, and c are always elements of the group.
- There exists an element of the group, called the *identity element* and denoted I, that combines with any other element to give the second one unchanged. In the case of the integers, the identity element is zero because any integer plus zero gives that integer $(a + 0 = a)$.
- For every element of the group, there exists another element that combines with the first to give the identity element; these are known as *inverse elements*. The negative integers constitute the inverses of the positive integers because their pairwise sums all equal zero, the identity element $(a + (-a) = 0)$.
- Group operations in sequence obey the associative law. For addition of integers this means that $(a + b) + c = a + (b + c)$. Notice that the commutative law, $a + b = b + a$, is not required even though it is true for this particular group.

You might be tempted to say that the positive integers, when related by multiplication $(a \times b = c)$, also constitute a group with the identity element now being one $(a \times 1 = a)$. In fact, the positive integers do not constitute a group under these conditions because, to obey the group-theory rules, the noninteger inverses $(1/a)$ as well as all the rational fractions (b/a) would have to be included. The expanded set of positive rational numbers *is* a group under multiplication, and both it and the integer group already discussed are examples of infinite groups because they each contain an infinite number of elements.

In the case of a symmetry group, an element is the operation needed to produce one object from another. For example, a mirror operation takes an object in one location and produces another of the opposite hand located such that the mirror doing the operation is equidistant between them (Fig. 1). These manipulations are usually called *symmetry operations*. They are combined by applying them to an object se-

THE MIRROR SYMMETRY OPERATION

Fig. 1. A pair of left- and right-"footed" boots illustrates the mirror-plane symmetry operation. The right boot can be positioned identically on the left boot by reflection through a mirror between them and vice versa.

quentially. For example, doing a mirror operation twice on a right-handed object will, with the first operation, move it to the left-handed position, and with the second operation, place it back on its original right-handed position. In fact, applying a mirror operation twice in succession is equivalent to the identity

operation, so that a mirror operation is its own inverse.

The two operations, mirror and identity, obey the four rules of group theory, and thus constitute one of the simplest symmetry groups. A mathematical representation of these operations is

$$m^{-1} = m \text{ and}$$
$$mm^{-1} = mm = I.$$

Further, a "multiplication table" between these two operations can be set up to show the products that any pair of symmetry operations gives in this finite group (Fig. 2).

There are three types of symmetry operations in crystallography. The simplest type is the set of *translation operations* needed to fill a two-dimensional infinite plane or a three-dimensional infinite space. These operations form a group by themselves and have essentially the same characteristics as the example group of integers discussed above. The difference is that the translation group has two or three sets of integers depending on whether a two-dimensional plane or a three-dimensional space is filled. These translation operations make the concept of a unit cell possible, because once the unit cell for a crystal is specified, it takes only the right combination of translation operations to construct the full crystal lattice.

There is also a type of translation operation that relates objects *within* a unit cell so that the same objects are found at coordinates that are half multiples of unit-cell distances along two or three of the axes. These last operations are, for example, responsible for the face- and body-centered lattices found in three dimensions (Fig. 3). The possible combinations of this full set of translations for plane- and space-filling arrays (along with the restrictions on the rotation-symmetry operations that will

be discussed next) gives only five possible plane lattices and fourteen possible space lattices (Fig. 3).

The second type of crystallographic symmetry is *rotation*. For it to be a valid symmetry operation, however, the rotation angle θ must be an integer divisor of 360 degrees, that is, $\theta = 360/n$, where n is an integer. The rotation-symmetry operations will then all be multiples of this rotation angle. For example, if $n = 6$ the rotation angle is 60 degrees and the operations can be represented by the unique set ${}^1C_6$, ${}^2C_6$, ${}^3C_6$ ($= {}^1C_2$), ${}^4C_6$, ${}^5C_6$, and ${}^6C_6$ ($= I$) in which the subscript gives the fraction of

A FINITE SYMMETRY GROUP

Fig. 2. This example of a simple, finite group obeying all the rules of group symmetry consists solely of the identity element, *I*, and the mirror-plane symmetry operation, *m*. The multiplication table shown above for the group gives the products for any pairwise application of the two symmetry operations.

a full circle for each operation (here 1/6) and the superscript gives the multiple of 60 degrees used for the rotation (Fig. 4). Because ${}^6C_6$ is the identity operation, these six rotation operations constitute a group, symbolized by C_6.

If the symmetry is local with no translation component, then the integer n can take on any value from one to infinity. An object that has the extreme case of C_∞ symmetry is a bowling pin, which an infinitesimally small rotation leaves looking the same (ignoring any painted design). However, when the rotation symmetry is part of a plane- or space-filling symmetry with translation operators, only five different rotation angles ($n = 1, 2, 3, 4,$ or 6) can be used. Replication of a unit cell with a rotation symmetry other than these cannot fill a plane surface or three-dimensional space without leaving voids or having overlapping regions. The situation is more complicated in the three-dimensional case because a unit cell may also have different rotation symmetry in different directions. Many different groups result from the various combinations of these rotations.
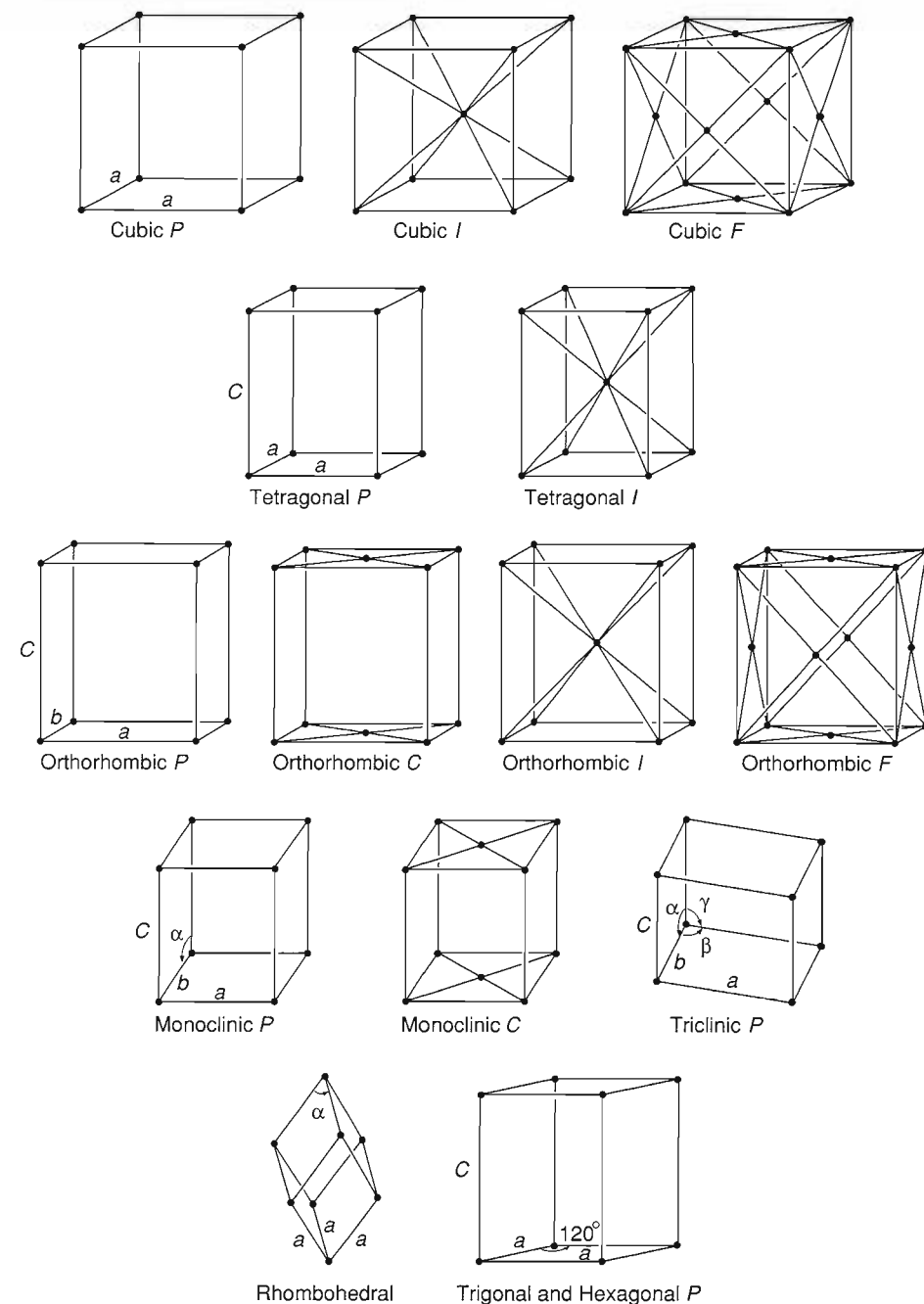
An extension to the concept of rotation symmetry is to include in each rotation operator a translation component (Fig. 5). The resulting objects are helical or screwlike; hence, these operations are called *screw rotations*. These symmetry operations are most prevalent in crystal lattices in which the unit-cell repeat requirement means that the translation operations have the same integer fraction, or some simple multiple, as the rotation operations. For example, the screw rotation 6_1 describes an operation in which the rotation of 60 degrees is accompanied by a translation of 1/6 of the unit cell along the rotation axis. The 6_4 screw rotation has the same 60-degree rotation but this time is accompanied by a translation of 4/6 of the unit cell along the axis. A sufficient number of these is superimposed to give the required unit-cell translation (Fig. 5), and the resulting arrangement is different from that obtained with a 6_1 screw rotation.

The one facet common to the translation, rotation, and screw operations is

THE BRAVAIS SPACE LATTICES

Fig. 3. The fourteen unit cells depicted above represent the only possible ways that space can be filled without gaps or overlaps between cells, that is, consonant with the restrictions of translation and rotation symmetry. The cubic cells at the top all have three orthogonal sides of equal length; the body-centered (*I*) and face-centered cubic cells (*F*) cannot be fully specified without also using translation operations in terms of half-cell distances. The tetragonal and orthorhombic cells also have sides that are mutually orthogonal, but either one side differs in length from the other two sides (tetragonal) or *all* three sides differ in length (orthorhombic). The monoclinic and triclinic cells have three unequal lengths but now either one angle (monoclinic) or all three angles (triclinic) between the sides do not equal 90 degrees. The rhombohedral cell can be thought of as a cubic cell that has been stretched or squeezed along a diagonal: the three sides are equal but the three angles, although equal, are not 90 degrees. The hexagonal cell has two angles of 90 degrees and one of 120 degrees; only two of its three sides are equal.

Cubic *P*   Cubic *I*   Cubic *F*

Tetragonal *P*   Tetragonal *I*

Orthorhombic *P*   Orthorhombic *C*   Orthorhombic *I*   Orthorhombic *F*

Monoclinic *P*   Monoclinic *C*   Triclinic *P*

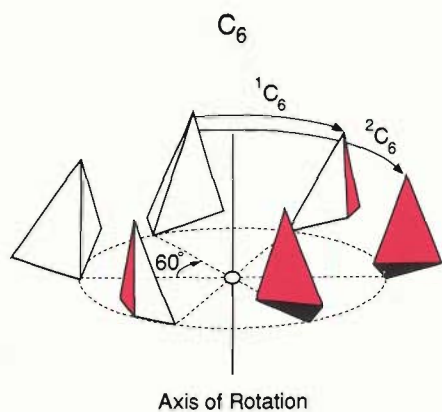Rhombohedral   Trigonal and Hexagonal *P*

that none of these change the handedness of an object, and *changing handedness* is the major feature of the third type of crystallographic symmetry. We have already mentioned the mirror-symmetry operation that relates right- and left-handed objects across a plane. A similar operation is *inversion* (Fig. 6) in which right- and left-handed objects are arranged on opposite sides of a point, called an inversion center. The presence of an inversion center in a crystal is one of the primary classification features for crystal structures: such crystal structures are *centrosymmetric*. An example of the importance of inversion centers is that almost all biologically important molecules (proteins, amino acids, et cetera) do not have a self-contained inversion center and exist

in nature only in one-handed forms. Thus, they always crystallize in noncentrosymmetric crystal structures because the other-handed molecules do not exist.

In analogy to the operations combining rotations with translations to form

screw operations, mirror reflection can be combined with a fractional translation (always one-half of the unit cell) to form a new operation (Fig. 7). This is known as a *glide* operation, and the mirror part of the operation occurs at
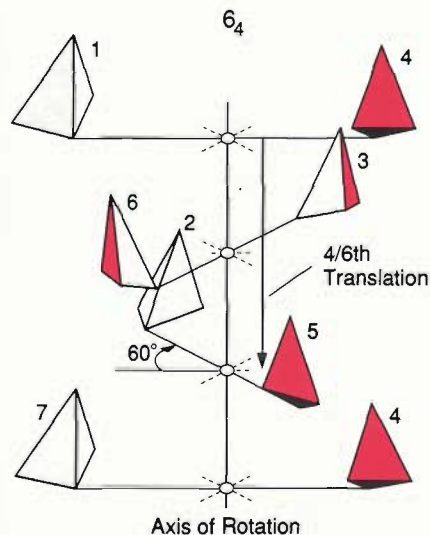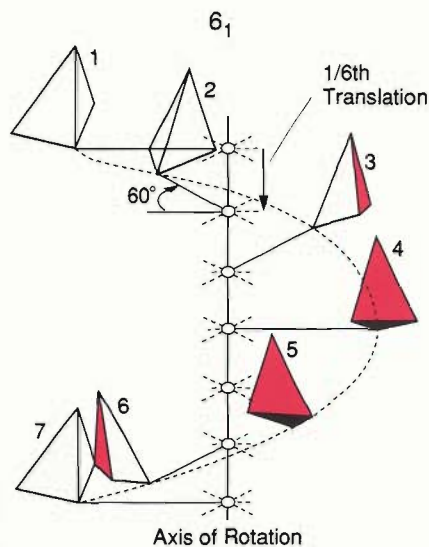
C_6

ROTATION OPERATIONS

Fig. 4. The C_6 rotation symmetry group consists of all the rotations about an axis *a* that carry an object through angles that are multiples of 60 degrees. Two of the operations in the symmetry group, $^1C_6$ and $^2C_6$, are labeled in the figure; $^6C_6$ is the identity operation that carries the object a full 360 degrees back into itself.

a *glide plane*. Just as for the screw operation, glide operations are only found in crystal lattices where the repetition of translation and reflection can extend indefinitely. Similarly, an inversion operation can be combined with a rotation (Fig. 8). Because this operation occurs about a point, however, it is found in both isolated objects and in extended lattices.

When these operations are combined in ways that form two-dimensional planar arrays, only 17 unique plane groups are found. With three dimensions, the combination of operations gives just 92 centrosymmetric and 138 noncentrosymmetric space groups for a total of 230.

An additional type of operation worth considering is one that in a two-dimensional plane would, say, change the color of the object (see the opening figure of the main article). The simplest case is a "black-white" operator,
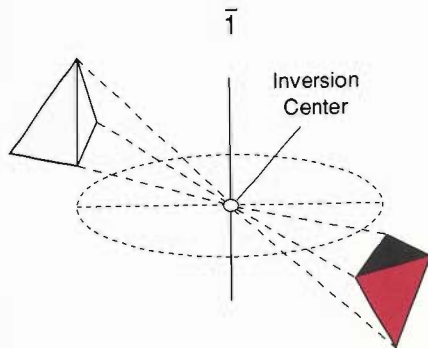
6_1

6_4

and such a color-reversal operator can also be combined with the other operators discussed earlier. An application of this type of operation is to describe the ordering of magnetic moments found in some materials by neutron scattering. Frequently, the moments arrange themselves in an alternating pattern so that every other one is "up" and all the others are "down." The symmetry of these arrangements can be described by including the color-reversal operation, which expands the total number

SCREW ROTATIONS

Fig. 5. a) The 6_1 screw rotation is the application of a 60-degree rotation about a given axis of the unit cell followed by a translation along that axis of one-sixth of the unit-cell distance. This combination of symmetry operations is repeated successively along the full length of the unit cell (in the figure, the tetrahedrons generated by each successive combination of operations are numbered consecutively). Note that the placement of the tetrahedrons in this figure resembles the placement of the tetrahedrons for rotation alone (Fig. 4) except that the circle has been "stretched out" into an arc because of the vertical translation along the axis of rotation. After six rotation-translation operations, the tetrahedron has returned to its original orientation but is translated a full unit along one of the cell's axes. b) The 6_4 screw rotation is the same as the 6_1 screw rotation except the translation is now for four-sixths (two-thirds) of the unit distance. To fill in the whole pattern, the next rotation-translation operation (which ends up one-third of the way into the next unit distance) and successive operations are superimposed on the original unit distance. Note that in the figure the dashed line has been eliminated (because successive operations are superimposed), but the tetrahedrons generated by successive operations are still numbered consecutively. After three of these combined operations, the tetrahedron will have moved an integral number of unit distances (and thus can be pictured at either the bottom or top of the figure) but will have rotated only 180 degrees. In this manner, the tetrahedron ends up on *both* sides of the axis at each point along the way. Once again, after six combined operations the tetrahedron has assumed its original orientation.

of space groups to 1728 in 36 Bravais lattices.

Because there is an intimate relationship between the arrangement of atoms found in real space and the pattern of structure factors in reciprocal space,

$\bar{1}$

Inversion
Center

INVERSION CENTER

Fig. 6. An inversion, denoted $\bar{1}$, is accomplished by "reflecting" everything through a point or "inversion center" between the objects. The three dashed lines drawn between tips on the tetrahedrons and passing through the inversion center illustrate this operation.
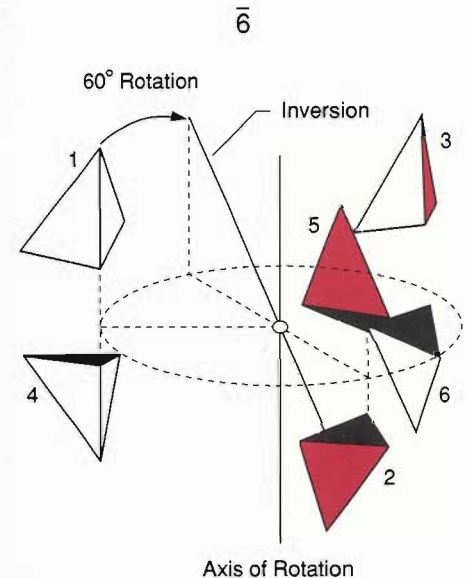
the symmetry of real space must have counterparts in reciprocal space. However, some of the symmetry aspects of reciprocal space may at first glance be surprising. Unlike crystallographic real space, which consists of a multitude of identical unit cells each with its own origin, reciprocal space has just a single origin and an infinite array of reciprocal-lattice points associated with differing and possibly complex numbers (F_h's). Thus, none of the translational aspects of the crystallographic symmetry can show up in the reciprocal lattice other than in the dimensions of the reciprocal lattice itself. However, the rotation, mirror, and inversion symmetries present in the lattice are also present in the pattern of $|F_h|^2$'s on the reciprocal-lattice (that is, in the diffraction pattern). For example, the intensities and locations of the two-dimensional diffraction patterns shown in Figs. 4 and 5 in the main article have the same rotation and mirror symmetries as the two-dimensional patterns of scatterers that generated those patterns.

What of the other possible symmetry elements? A diffraction pattern almost always has a center of inversion—an inversion center is absent only for a noncentrosymmetric crystal containing an atom with a complex scattering factor. Half-cell translations and screw and glide-plane operations are revealed by *systematic extinctions*, that is, certain classes of reciprocal-lattice points with zero intensity. For example, in the diffraction pattern for a face-centered cubic lattice, the only points that have a nonzero intensity are those for which the *hkl* indices are all even (for example, 422) or all odd (for example, 311). Likewise, a glide operation whose glide plane is perpendicular to the *c* crystallographic axis and whose glide direction is parallel to the *a* axis causes the points with *hk*0 indices and odd *h* to have zero intensity (for example, 120, whereas 210 has nonzero intensity). Systematic extinc-

Half
Translation
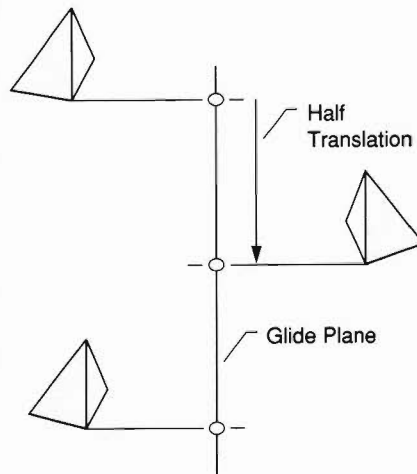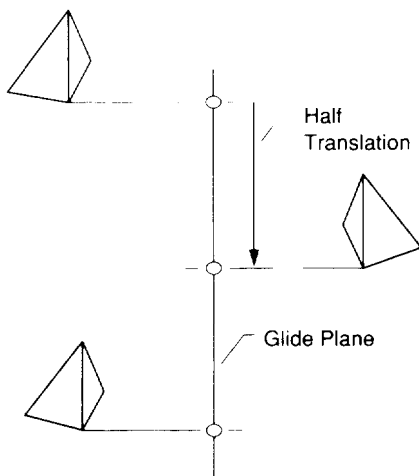
Glide Plane

THE GLIDE OPERATION

Fig. 7. Here, mirror reflection and translation for one-half the unit distance are combined to form a glide operation. Note that the tetrahedron on the right side of the glide plane is the mirror image of the tetrahedrons on the left side; however, each tetrahedron is displaced a half unit from the last one.
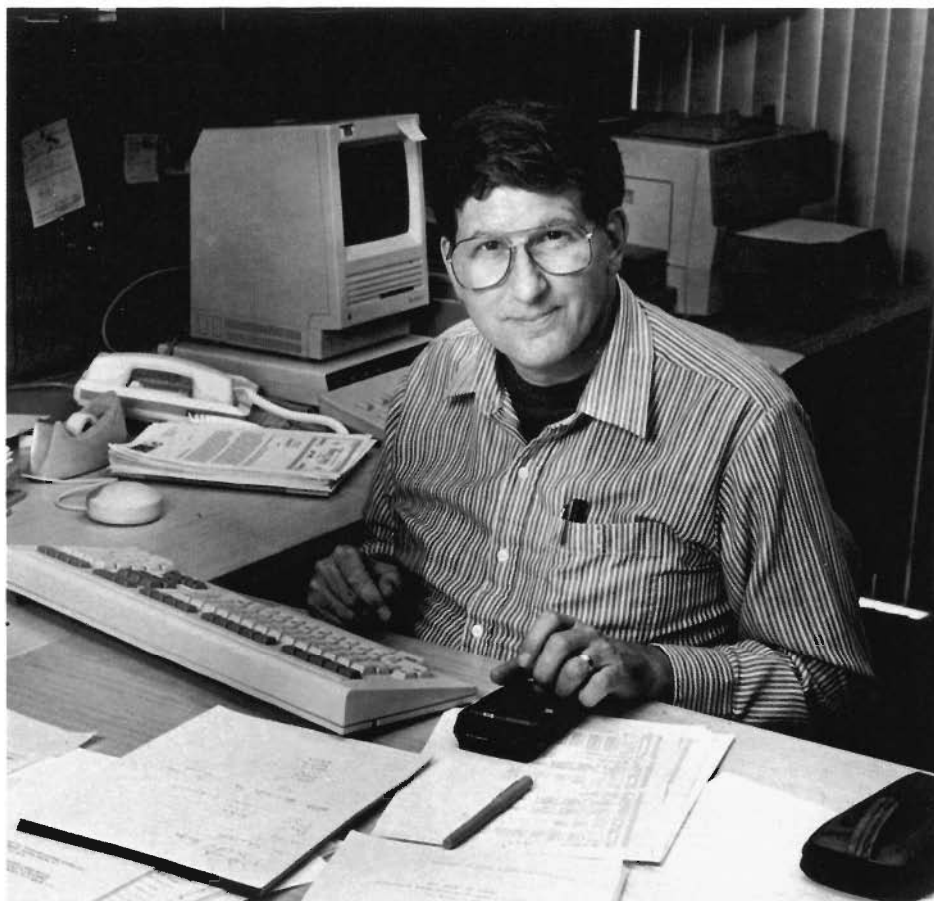
$\bar{6}$

60° Rotation

Inversion

3

5

1

4

6

2

Axis of Rotation
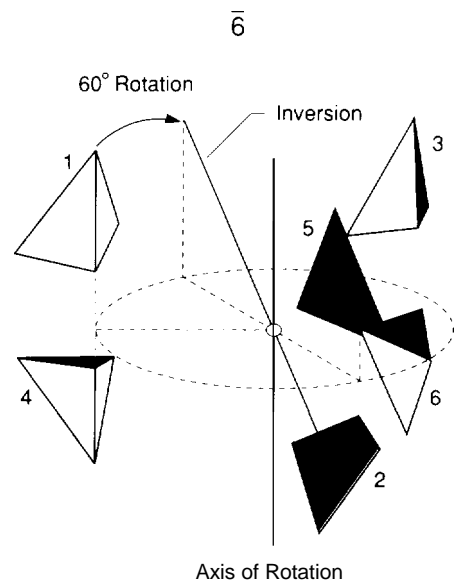
AN INVERSION-ROTATION OPERATION

Fig. 8. The $\bar{6}$ symmetry operation is a combination of a 60-degree rotation followed by an inversion. Note that the three tetrahedrons above the plane are the same as the tetrahedrons in Fig. 4 for rotations of 0, 120, and 240 degrees (that is, I, $^2C_6$, and $^4C_6$). This happens because performing two successive $\bar{6}$ operations is equivalent to performing the $^2C_6$ operation (or two $^1C_6$ operations). Lines showing the first combination of a 60-degree rotation and inversion operation are given on the figure as well as consecutive numbers for the successively generated tetrahedrons.

tions arise because the symmetry operation causes all the atoms to scatter with destructive interference for particular reciprocal-lattice points.

Thus, by examining both the symmetry of a diffraction pattern and the systematic extinctions, a crystallographer can usually identify one or two possible space groups for any crystal. However, some ambiguity may remain because of cases in which pairs of space groups display the same diffraction symmetry and systematic extinctions. ∎

R. A. Young, P. E. Mackie, and R. B. Von Dreele. 1977. Application of the pattern-fitting structure-refinement method to x-ray powder diffractometer patterns. *Journal of Applied Crystallography* 10: 262.

R. B. Von Dreele, J. D. Jorgensen, and C. G. Windsor. 1982. Rietveld refinement with spallation neutron powder diffraction data. *Journal of Applied Crystallography* 15: 581.

Crystallography And Crystal Symmetry

M. J. Buerger. 1963. *Elementary Crystallography*. New York: John Wiley & Sons.

J. E. Brigham, translator. 1972. *The Graphic Works of M. C. Escher*. London: Pan/Ballantine.

N. F. M. Henry and K. Lonsdale, editors. 1969. *Symmetry Groups*. International Tables for X-Ray Crystallography, volume 1. Birmingham: Kynoch Press.

M. Senechal and G. Fleck, editors. 1974. *Patterns of Symmetry*. Amherst, Massachusetts: University of Massachusetts Press.

G. Burns and A. M. Glazer. 1978. *Space Groups for Solid State Scientists*. New York: Academic Press.

B. K. Vainshtein. 1981. *Modern Crystallography. I. Symmetry of Crystals, Methods of Structural Crystallography*. New York: Springer-Verlag.

M. A. Jaswon and M. A. Rose. 1983. *Crystal Symmetry: Theory of Color Symmetry*. New York: Halsted Press.

T. Hahn, editor. 1985. *Space-Group Symmetry*. International Tables for Crystallography, volume A. Boston: Reidel.

Robert B. Von Dreele has been a staff member at Los Alamos since 1986. (He had, however, established a connection with the Laboratory earlier by doing neutron powder diffractometry at the Omega West Reactor and by consulting at what is now LANSCE during the summers of 1980–85.) His current responsibilities include the High Intensity Powder Diffractometer at LANSCE. He received his B.S. in Chemical Engineering and Ph.D. in Chemistry from Cornell University in 1966 and 1971, respectively. Immediately after completing his Ph.D. he joined the faculty of the Department of Chemistry at Arizona State University, where he remained until he joined the Laboratory. In 1971 he was a National Science Foundation Postdoctoral Fellow at Oxford University, where he worked on one of the first applications of the Rietveld refinement technique to neutron powder-diffraction data. A subsequent sabbatical year was spent as the first visitor to the "Neutron Beam Research Unit" (now ISIS) at the Rutherford-Appleton Laboratory in England. There he developed the first computer code capable of Rietveld refinement with time-of-flight powder-diffraction data from a spallation source. He later visited ISIS for seven months as a Fulbright Scholar.

Further Reading

Crystal Structures from X-Ray and Neutron Diffraction

G. H. Stout and L. H. Jensen. 1968. *X-Ray Structure Determination: A Practical Guide*. New York: Macmillan Co.

G. E. Bacon. 1975. *Neutron Diffraction*. Oxford: Clarendon Press.

G. Harburn, C. A. Taylor, and T. R. Welberry. 1975. *Atlas of Optical Transforms*. Ithaca, New York: Cornell University Press

D. L. Bish and J. E. Post, editors. 1989. *Modern Powder Diffraction*. Reviews in Mineralogy, volume 20. Washington, D.C.: Mineralogical Society of America.

H. M. Rietveld. 1969. A profile refinement method for nuclear and magnetic structures. *Journal of Applied Crystallography* 2: 65.

Crystal Symmetry Groups

Symmetry plays an important role in crystallography. The ways in which atoms and molecules are arranged within a unit cell and unit cells repeat within a crystal are governed by symmetry rules. In ordinary life our first perception of symmetry is what is known as *mirror symmetry.* Our bodies have, to a good approximation, mirror symmetry in which our right side is matched by our left as if a mirror passed along the central axis of our bodies. Our hands illustrate this most vividly; so much so that the image is carried over to crystallography when one speaks of a molecule as being either "right"- or "left"- handed. Those of us who live in an old-fashioned duplex will also recognize that such houses are built with mirror symmetry so that the arrangement of the rooms, hallways, and doors are disposed about an imaginary mirror passing through the common wall between the two halves of the house. There are many other examples of this kind of mirror symmetry in ordinary life. We can also see more complex symmetry in the patterns around us. It can be found in wallpaper patterns, floor-tile arrays, cloth designs, flowers, and mineral crystals. The basic mathematics of symmetry also applies to music, dance (particularly folk and square dance), and even the operations needed to solve Rubik's cube.

The rules that govern symmetry are found in the mathematics of group theory. Group theory addresses the way in which a certain collection of mathematical "objects" are related to each other. For example, consider all the positive and negative integers and zero. They can constitute a group because under certain circumstances the relationships
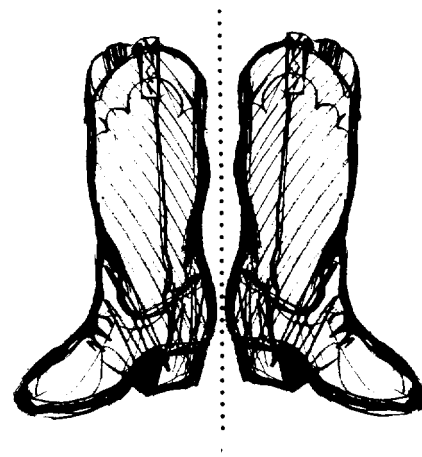
between the integers obey the rules of group theory:

- There must be defined a procedure for combining two elements of the group to form a third. For the integers one can choose the addition operation so that $a + b = c$ *is* the operation to be performed and u, b, and c are always elements of the group.
- There exists an element of the group, called the *identity element* and denoted f, that combines with any other element to give the second one unchanged. In the case of the integers, the identity element is zero because any integer plus zero gives that integer $(a + O = a).$
- For every element of the group, there exists another element that combines with the first to give the identity element; these are known as *inverse elements.* The negative integers constitute the inverses of the positive integers because their pairwise sums all equal zero, the identity element $(a + (-a) = 0).$
- Group operations in sequence obey the associative law. For addition of integers this means that $(a + b) + c = a + (b + c).$ Notice that the commutative law, $a + b = b + a,$ is not required even though it is true for this particular group.

You might be tempted to say that the positive integers, when related by multiplication $(a \times b = c),$ also constitute a group with the identity element now being one $(a \times 1 = a).$ In fact, the positive integers do not constitute a group under these conditions because, to obey the group-theory rules, the noninteger inverses $(1 /a)$ as well as all the rational fractions (b/a) would have to be included. The expanded set of positive

rational numbers *is* a group under multiplication, and both it and the integer group already discussed are examples of infinite groups because they each contain an infinite number of elements.

In the case of a symmetry group, an element is the operation needed to produce one object from another. For example, a mirror operation takes an object in one location and produces another of the opposite hand located such that the mirror doing the operation is equidistant between them (Fig. 1). These manipulations are usually called *symmetry operations.* They are combined by applying them to an object se-

THE MIRROR SYMMETRY OPERATION

Fig. 1. A pair of left- and right-"footed" boots Illustrates the mirror-plane symmetry operation. The right boot can be positioned identically on the left boot by reflection through a mirror between them and vice versa.

quentially. For example, doing a mirror operation twice on a right-handed object will, with the first operation, move it to the left-handed position, and with the second operation, place it back on its original right-handed position. In fact, applying a mirror operation twice in succession is equivalent to the identity

operation, so that a mirror operation is its own inverse.

The two operations, mirror and identity, obey the four rules of group theory, and thus constitute one of the simplest symmetry groups. A mathematical representation of these operations is

$$m^{-1} = m \text{ and}$$
$$mm^{-1} = mm = 1.$$

Further, a "multiplication table" between these two operations can be set up to show the products that any pair of symmetry operations gives in this finite group (Fig. 2).

There are three types of symmetry operations in crystallography. The simplest type is the set of *translation operations* needed to fill a two-dimensional infinite plane or a three-dimensional infinite space. These operations form a group by themselves and have essentially the same characteristics as the example group of integers discussed above. The difference is that the translation group has two or three sets of integers depending on whether a two-dimensional plane or a three-dimensional space is filled. These translation operations make the concept of a unit cell possible, because once the unit cell for a crystal is specified, it takes only the right combination of translation operations to construct the full crystal lattice.

There is also a type of translation operation that relates objects *within* a unit cell so that the same objects are found at coordinates that are half multiples of unit-cell distances along two or three of the axes. These last operations are, for example, responsible for the face- and body-centered lattices found in three dimensions (Fig. 3). The possible combinations of this full set of translations for plane- and space-filling arrays (along with the restrictions on the rotation-symmetry operations that will

be discussed next) gives only five possible plane lattices and fourteen possible space lattices (Fig. 3).

The second type of crystallographic symmetry is *rotation.* For it to be a valid symmetry operation, however, the rotation angle θ must be an integer divisor of 360 degrees, that is, $\theta = 360/n$, where n is an integer. The rotation-symmetry operations will then all be multiples of this rotation angle. For example, if $n = 6$ the rotation angle is 60 degrees and the operations can be represented by the unique set $^1C_6$, $^2C_6$, $^3C_6 (= {}^1C_2)$, $^4C_6$, $^5C_6$, and $^6C_6 (= I)$ in which the subscript gives the fraction of

A FINITE SYMMETRY GROUP

Fig. 2. This example of a simple, finite group obeying all the rules of group symmetry consists solely of the Identity element, 1, and the mirror-plane symmetry operation, *m*. The multiplication table shown above for the group gives the products for any pairwise application of the two symmetry operations.

a full circle for each operation (here 1/6) and the superscript gives the multiple of 60 degrees used for the rotation (Fig. 4). Because $^6C_6$ is the identity operation, these six rotation operations constitute a group, symbolized by C&
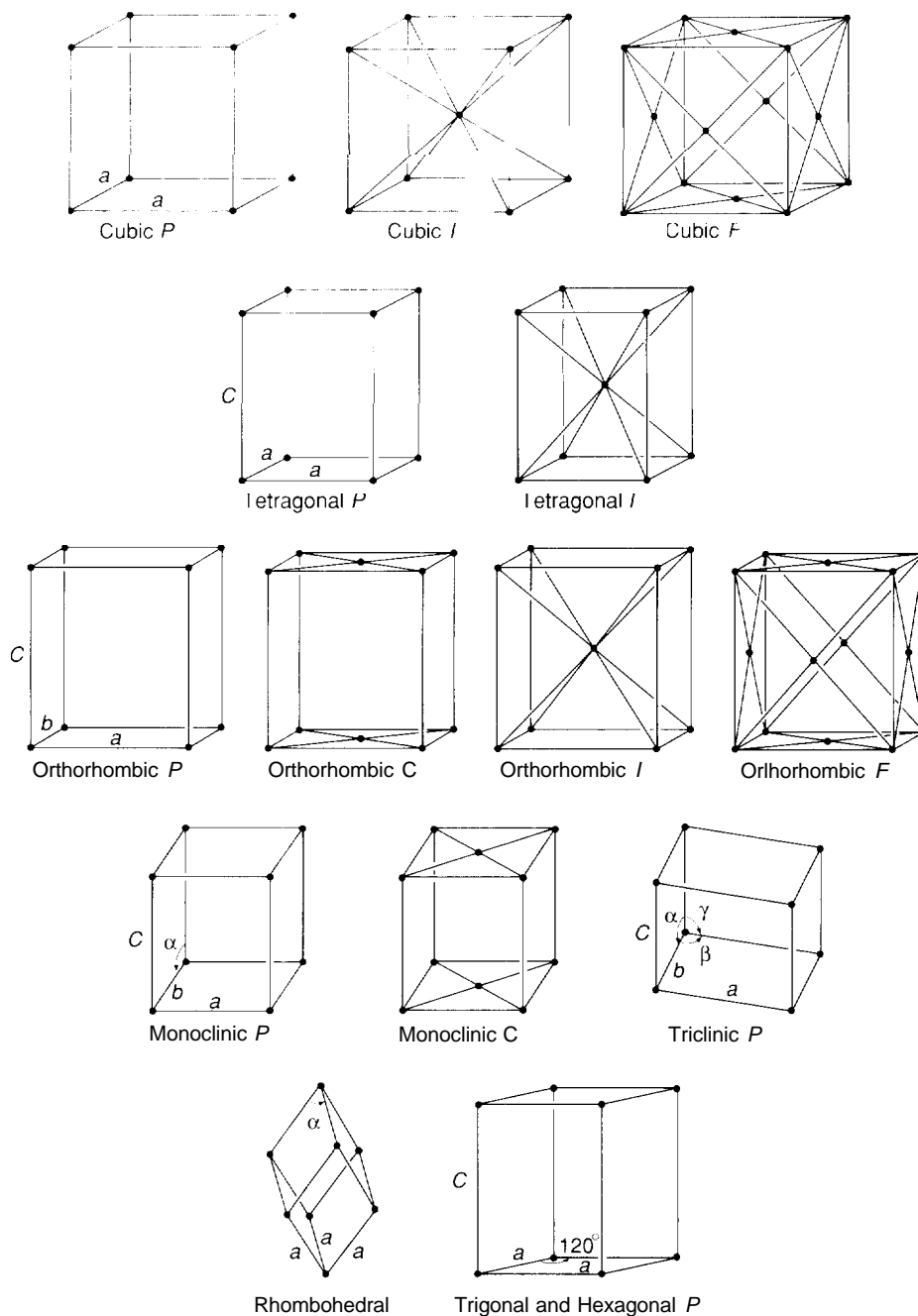
If the symmetry is local with no translation component, then the integer n can take on any value from one to infinity. An object that has the extreme case of C_∞ symmetry is a bowling pin, which an infinitesimally small rotation leaves looking the same (ignoring any painted design). However, when the rotation symmetry is part of a plane- or space-filling symmetry with translation operators, only five different rotation angles *(n = 1,* 2, 3, 4, or 6) can be used. Replication of a unit cell with a rotation symmetry other than these cannot fill a plane surface or three-dimensional space without leaving voids or having overlapping regions. The situation is more complicated in the three-dimensional case because a unit cell may also have different rotation symmetry in different directions. Many different groups result from the various combinations of these rotations.

An extension to the concept of rotation symmetry is to include in each rotation operator a translation component (Fig. 5). The resulting objects are helical or screwlike; hence, these operations are called *screw rotations.* These symmetry operations are most prevalent in crystal lattices in which the unit-cell repeat requirement means that the translation operations have the same integer fraction, or some simple multiple, as the rotation operations. For example, the screw rotation 6_1 describes an operation in which the rotation of 60 degrees is accompanied by a translation of 1/6 of the unit cell along the rotation axis. The 6_4 screw rotation has the same 60-degree rotation but this time is accompanied by a translation of 4/6 of the unit cell along the axis. A sufficient number of these is superimposed to give the required unit-cell translation (Fig. 5), and the resulting arrangement is different from that obtained with a 6_1 screw rotation.

The one facet common to the translation, rotation, and screw operations is

THE BRAVAIS SPACE LATTICES

Fig. 3. The fourteen unit cells depicted above represent the only possible ways that space can be filled without gaps or overlaps between cells, that is, consonant with the restrictions of translation and rotation symmetry. The cubic cells at the top all have three orthogonal sides of equal length; the body-centered (*I*) and face-centered cubic cells (*F*) cannot be fully specified without also using translation operations in terms of half-cell distances. The tetragonal and orthorhombic cells also have sides that are mutually orthogonal, but either one side differs in length from the other two sides (tetragonal) or all three sides differ in length (orthorhombic). The monoclinic and triclinic cells have three unequal lengths but now either one angle (monoclinic) or all three angles (triclinic) between the sides do not equal 90 degrees. The rhombohedral cell can be thought of as a cubic cell that has been stretched or squeezed along a diagonal: the three sides are equal but the three angles, although equal, are not 90 degrees. The hexagonal cell has two angles of 90 degrees and one of 120 degrees; only two of its three sides are equal.
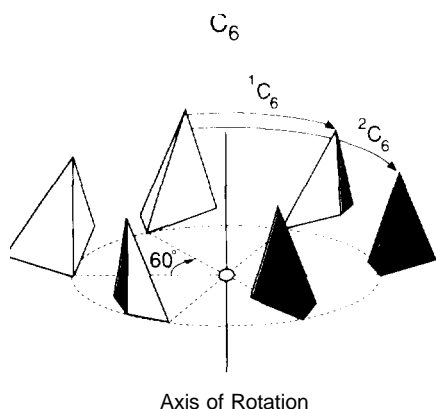

Cubic *P*   Cubic *I*   Cubic *F*

Tetragonal *P*   Tetragonal *I*

Orthorhombic *P*   Orthorhombic C   Orthorhombic *I*   Orlhorhombic *F*

Monoclinic *P*   Monoclinic C   Triclinic *P*

Rhombohedral   Trigonal and Hexagonal *P*

that none of these change the handedness of an object, and *changing handedness is* the major feature of the third type of crystallographic symmetry. We have already mentioned the mirror-symmetry operation that relates right- and left-handed objects across a plane. A similar operation is *inversion* (Fig. 6) in which right- and left-handed objects are arranged on opposite sides of a point, called an inversion center. The presence of an inversion center in a crystal is one of the primary classification features for crystal structures: such crystal structures are *centrosymmetric.* An example of the importance of inversion centers is that almost all biologically important molecules (proteins, amino acids, et cetera) do not have a self-contained inversion center and exist

in nature only in one-handed forms. Thus, they always crystallize in noncentrosymmetric crystal structures because the other-handed molecules do not exist.

In analogy to the operations combining rotations with translations to form

screw operations, mirror reflection can be combined with a fractional translation (always one-half of the unit cell) to form a new operation (Fig. 7). This is known as a *glide* operation, and the mirror part of the operation occurs at
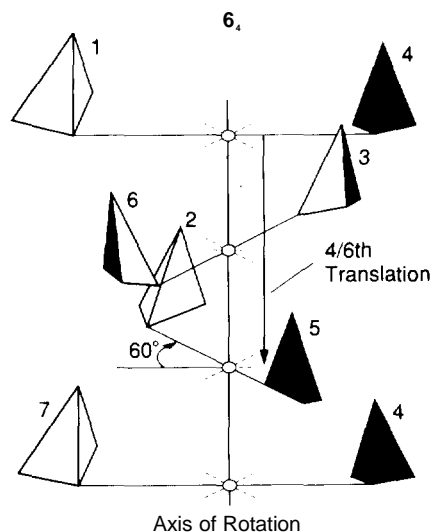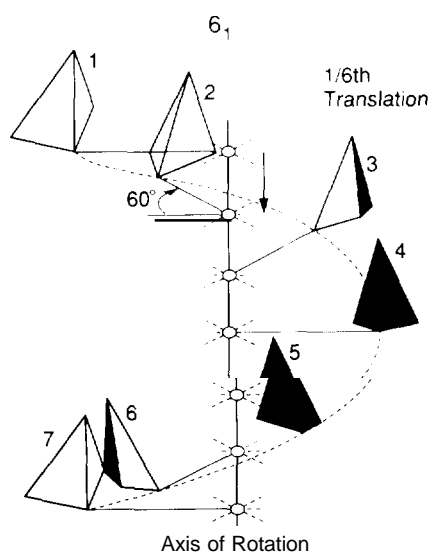
C_6

$^1C_6$

$^2C_6$

60°

Axis of Rotation

ROTATION OPERATIONS

Fig. 4. The C_6 rotation symmetry group consists of all the rotations about an axis a that carry an object through angles that are multiples of 60 degrees. Two of the operations in the symmetry group, $^1C_6$ and $^2C_6$, are labeled in the figure; $^6C_6$ is the identity operation that carries the object a full 360 degrees back into itself.

a *glide plane.* Just as for the screw operation, glide operations are only found in crystal lattices where the repetition of translation and reflection can extend indefinitely. Similarly, an inversion operation can be combined with a rotation (Fig. 8). **Because** this operation occurs about a point, however, it is found in both isolated objects and in extended lattices,

When these operations are combined in ways that form two-dimensional planar arrays, only 17 unique plane groups are found. With three dimensions, the combination of operations gives just 92 centrosymmetric and 138 noncentrosymmetric space groups for a total of 230.

An additional type of operation worth considering is one that in a two-dimensional plane would, say, change the color of the object (see the opening figure of the main article). The simplest case is a "black-white" operator,

6_1

1
2
60°
3
4
5
6
7

1/6th Translation

Axis of Rotation

6_4

1
4
3
6
2
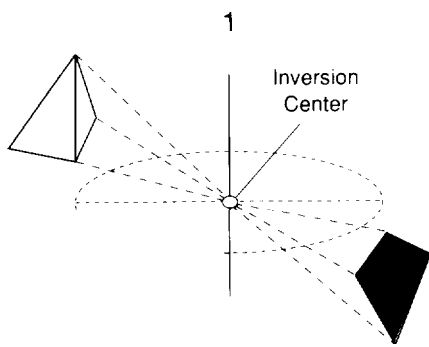5
7
4

4/6th Translation

60°

Axis of Rotation

and such a color-reversal operator can also be combined with the other operators discussed earlier. An application of this type of operation is to describe the ordering of magnetic moments found in some materials by neutron scattering. Frequently, the moments arrange themselves in an alternating pattern so that every other one is "up" and all the others are "down." The symmetry of these arrangements can be described by including the color-reversal operation, which expands the total number

SCREW ROTATIONS

Fig. 5. a) The 61 screw rotation is the application of a 60-degree rotation about a given axis of the unit cell followed by a translation along that axis of one-sixth of the unit-cell distance. This combination of symmetry operations is repeated successively along the full length of the unit cell (in the figure, the tetrahedrons generated by each successive combination of operations are numbered consecutively). Note that the placement of the tetrahedrons in this figure resembles the placement of the tetrahedrons for rotation alone (Fig. 4) except that the circle has been "stretched out" into an arc because of the vertical translation along the axis of rotation. After six rotation-translation operations, the tetrahedron has returned to its original orientation but is translated a full unit along one of the cell's axes. b) The 6_4 screw rotation is the same as the 6_1 screw rotation except the translation is now for four-sixths (two-thirds) of the unit distance. To fill in the whole pattern, the next rotation-translation operation (which ends up one-third of the way into the next unit distance) and successive operations are superimposed on the original unit distance. Note that in the figure the dashed line has been eliminated (because successive operations are superimposed), but the tetrahedrons generated by successive operations are still numbered consecutively. After three of these combined operations, the tetrahedron will have moved an integral number of unit distances (and thus can be pictured at either the bottom or top of the figure) but will have rotated only 180 degrees. In this manner, the tetrahedron ends up on *both* sides of the axis at each point along the way. Once again, after six combined operations the tetrahedron has assumed its original orientation.

of space groups to 1728 in 36 Bravais lattices.

Because there is an intimate relationship between the arrangement of atoms found in real space and the pattern of structure factors in reciprocal space,

1

Inversion
Center

INVERSION CENTER

Fig. 6. An inversion, denoted $\bar{1}$, is accomplished by "reflecting" everything throughs point or "inversion center" between the objects. The three dashed lines drawn between tips on the tetrahedrons and passing through the inversion center illustrate this operation.

the symmetry of real space must have counterparts in reciprocal space. However, some of the symmetry aspects of reciprocal space may at first glance be surprising. Unlike crystallographic real space, which consists of a multitude of identical unit cells each with its own origin, reciprocal space has just a single origin and an infinite array of reciprocal-lattice points associated with differing and possibly complex numbers (F_h's). Thus, none of the translational aspects of the crystallographic symmetry can show up in the reciprocal lattice other than in the dimensions of the reciprocal lattice itself. However, the rotation, mirror, and inversion symmetries present in the lattice are also present in the pattern of $|F_h|^2$'s on the reciprocal-lattice (that is, in the diffraction pattern). For example, the intensities and locations of the two-dimensional diffraction patterns shown in Figs. 4 and 5 in the main article have the same rotation and mirror symmetries as the two-dimensional patterns of scatterers that generated those patterns.

What of the other possible symmetry elements? A diffraction pattern almost always has a center of inversion—an inversion center is absent only for a noncentrosymmetric crystal containing an atom with a complex scattering factor. Half-cell translations and screw and glide-plane operations are revealed by *systematic extinctions,* that is, certain classes of reciprocal-lattice points with zero intensity. For example, in the diffraction pattern for a face-centered cubic lattice, the only points that have a nonzero intensity are those for which the *hkl* indices are all even (for example, 422) or all odd (for example, 31 1). Likewise, a glide operation whose glide plane is perpendicular to the c crystallographic axis and whose glide direction is parallel to the a axis causes the points with *hkO* indices and **odd** *h* to have zero intensity (for example, 120, whereas 210 has nonzero intensity). Systematic extinc-

Half
Translation

Glide Plane

THE GLIDE OPERATION

Fig. 7. Here, mirror reflection and translation for one-half the unit distance are combined to form a glide operation. Note that the tetrahedron on the right side of the glide plane Is the mirror Image of the tetrahedrons on the left aide; however, each tetrahedron is displaced a half unit from the last one.
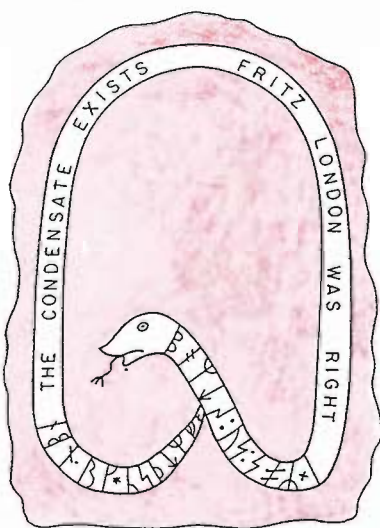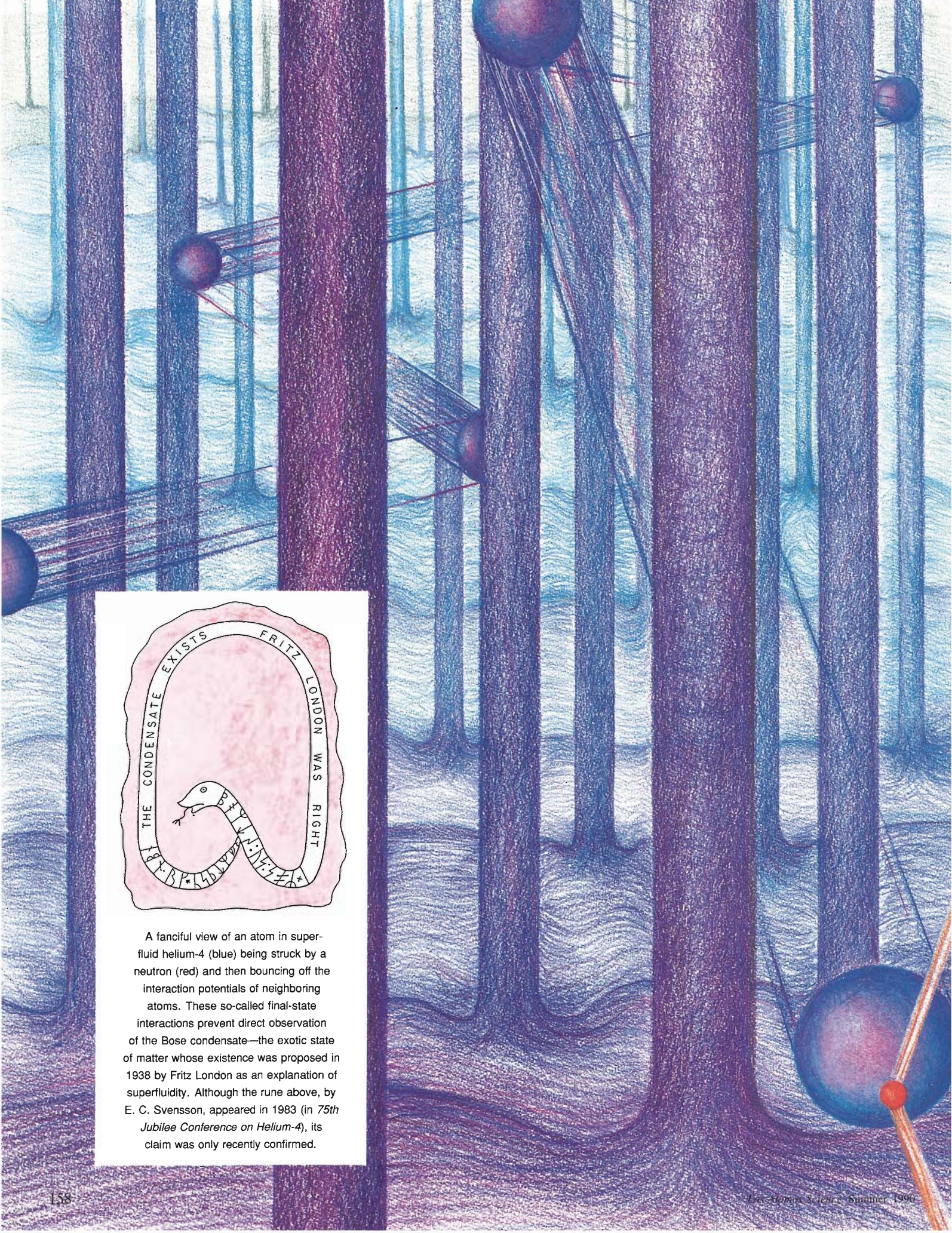
$\bar{6}$

60° Rotation

Inversion

Axis of Rotation

AN INVERSION-ROTATION OPERATION

Fig. 8. The $\bar{6}$ symmetry operation is a combination of s 60-degree rotation followed by an inversion. Note that the three tetrahedrons above the plane are the same as the tetrahedrons in Fig. 4 for rotations of O, 120, and 240 degrees (that is, *I,* $^2C_6$, and $^4C_6$). This happens because performing two successive 6 operations is equivalent to performing the $^2C_6$ operation (or two $^1C_6$ operations). Lines showing the first combination of a 60-degree rotation and inversion operation are given on the figure as well as consecutive numbers for the successively generated tetrahedrons.

tions arise because the symmetry operation causes all the atoms to scatter with destructive interference for particular reciprocal-lattice points.

Thus, by examining both the symmetry of a diffraction pattern and the systematic extinctions, a crystallographer can usually identify one or two possible space groups for any crystal. However, some ambiguity may remain because of cases in which pairs of space groups display the same diffraction symmetry and systematic extinctions. ∎

A fanciful view of an atom in super-fluid helium-4 (blue) being struck by a neutron (red) and then bouncing off the interaction potentials of neighboring atoms. These so-called final-state interactions prevent direct observation of the Bose condensate—the exotic state of matter whose existence was proposed in 1938 by Fritz London as an explanation of superfluidity. Although the rune above, by E. C. Svensson, appeared in 1983 (in *75th Jubilee Conference on Helium-4*), its claim was only recently confirmed.

superfluid helium and neutron scattering
a new chapter in the
CONDENSATE SAGA

by Richard N. Silver

The unusual properties of helium at temperatures near absolute zero have been an endless source of fascination for condensed-matter physicists. Helium is the only atomic system that avoids crystallization and instead remains a fluid to arbitrarily low temperature. Moreover, when a liquid composed of ^{4}He atoms is cooled below a critical temperature T_λ (equal to 2.17 kelvins at atmospheric pressure), it passes from a *normal-fluid* state, so called because its properties are similar to those of other fluids, to a *superfluid* state having dramatically different properties. A normal fluid possesses a finite viscosity, or resistance to shear flow, and therefore current flows dissipate in the absence of a driving force. A normal fluid also has a finite thermal conductivity, or the ability to support temperature gradients. In contrast, a superfluid has a zero viscosity and an infinite thermal conductivity. In a superfluid quantized currents persist indefinitely, and temperature fluctuations propagate like waves. In addition, numerous other properties of superfluids lie outside the realm of common experience. Finally, unlike the behavior of normal fluids, which can usually be described in terms of classical mechanics, the exotic behavior of liquid helium below T_λ requires a quantum-mechanical description. For that reason superfluid helium is called a *quantum fluid*.

In this article we focus on a single question: What quantum features of helium atoms at temperatures below T_λ might explain the transition from normal-fluid to superfluid behavior? We shall eventually arrive at a clear answer, but the route we must follow to find it is circuitous. Along the way we will explore much of the history of and current research on quantum fluids and learn about related research in many areas of modern physics. We will see that the high fluxes of epithermal neutrons available at pulsed neutron sources, such as those at Argonne and Los Alamos National Laboratories, are a powerful tool in addressing our central question.

An important clue to the answer comes from comparing the behavior of two different helium fluids, one composed of ^{3}He atoms and the other of ^{4}He atoms. Both isotopic species have identical interatomic interactions, and the difference in their masses has a negligible effect on their behavior. Under atmospheric pressure both remain fluids to arbitrarily low temperatures. However, ^{4}He undergoes a superfluid transition at 2.17 kelvins, but ^{3}He does not become a superfluid until below 3 millikelvins. The transition temperatures of the two species differ by three orders of magnitude!

The origin of that great difference can be traced to the *spin-statistics relation*, the fundamental principle of quantum mechanics that distinguishes ^{3}He from ^{4}He. *Statistics* refers to symmetry properties of the wave function describing a system of identical (and therefore indistinguishable) particles. The ^{3}He nucleus, composed

of two protons and one neutron, has a spin of $\frac{1}{2}$; that is, it has an intrinsic angular momentum of $\frac{1}{2}\hbar$, where $\hbar$ is Planck's constant divided by 2π. (Electrons also have a spin of $\frac{1}{2}$, but the spins of the two electrons in each helium atom are antiparallel and thus contribute a total spin of zero.) According to the spin-statistics relation, a system of identical particles having half-integer spin must be described by a many-particle wave function that is a completely antisymmetric function of the particle positions. Such a function is zero when two particles have the same position. Therefore, since the probability of finding particles at given positions is the square of the wave function, the spin-statistics relation alone requires particles with half-integer spin to avoid one another even in the absence of any repulsive potential between them. The antisymmetry of the wave function is thus the origin of the famous Pauli exclusion principle, which requires that no two electrons in an atom occupy the same atomic orbital. (Particles with half-integer spin are termed *fermions,* and the spin-statistics relation for fermions is called Fermi-Dirac statistics.)

The converse is true for ^{4}He atoms. The ^{4}He nucleus, composed of two protons and two neutrons, has a spin of 0, and so also does the ^{4}He atom. According to the same spin-statistics relation, a system of identical particles having integer spin must be described by a many-particle wave function that is a completely symmetric function of the particle positions. A completely symmetric function is larger when two particles occupy the same position. Hence, the spin-statistics relation requires integer-spin particles, such as spin-0 ^{4}He atoms and spin-1 photons, to be attracted to one another even in the absence of any attractive potential between them. (Particles with integer spin are termed *bosons,* and the spin-statistics relation for bosons is called Bose-Einstein statistics.)

Although the spin-statistics relation applies to the *microscopic* behavior of a system of identical particles, its effects are visible at the human, or *macroscopic*, level of perception. For example, if all the electrons in an atom could occupy the lowest atomic orbital, then the universe as we know it would collapse. Fortunately electrons are fermions, and such a collapse is forbidden by the Pauli exclusion principle. An outstanding macroscopic effect for bosons is the phenomenon of lasing. A laser can produce an intense beam of coherent light because the photons emitted by a large population of excited atoms are allowed, and indeed prefer, to enter the same quantum-mechanical state.

In 1938 Fritz London proposed that the superfluidity observed in liquid ^{4}He earlier that year was just such a macroscopic manifestation of the symmetry requirement for the wave function of a system of identical bosons, an insight that preceded the invention of lasers by twenty-seven years. He reasoned by analogy with the theoretically predicted behavior of a non-interacting gas of spin-0 particles obeying Bose-Einstein statistics. His argument is given in terms of the momentum wave function of the system, which is simply the Fourier transform of the position wave function mentioned above. (Position and momentum are expressed mathematically in quantum mechanics as Fourier conjugate variables that obey the Heisenberg uncertainty principle, $\Delta p \Delta x \leq \hbar$, where Δp and Δx are the uncertainties in momentum and position.

Consider a system of identical atoms inside a box of finite size. The probability that an atom has momentum of magnitude p in, say, the x direction is termed the *momentum distribution, $n(p)$*. For all systems of atoms obeying classical mechanics, $n(p)$ is given by a Maxwell-Boltzmann distribution, which is a Gaussian function of width $\Delta p = \sqrt{mk_BT}$ in each direction. Here m is the atomic mass, k_B is Boltzmann's constant, and T is temperature. This classical-mechanical momentum distribution is independent of any interactions between the atoms and yields an average kinetic, or thermal, energy ($\propto (\Delta p)^2/2m$) per atom of $\frac{3}{2}k_BT$. Thus, all liquids that obey classical mechanics must crystallize as the temperature is lowered because the potential energy gained by the localization of atoms at lattice sites overcomes the kinetic energy due to thermal motion. That ^{4}He and ^{3}He require a quantum description is already
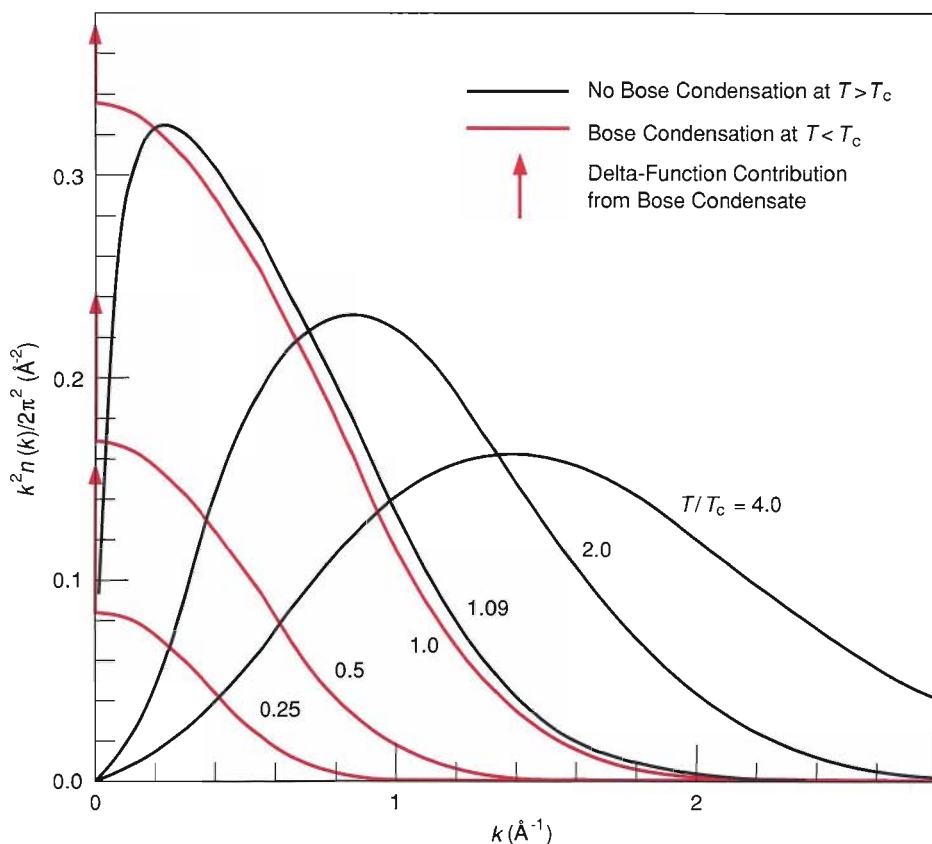
Fig. 1. Shown here are plots, at various temperatures above and below the Bose-condensation temperature T_c, of $k^2 n(k)/2\pi^2$ versus k, where $n(k)$ is the momentum distribution of an ideal Bose gas of density ρ. As the temperature decreases toward T_c, $k^2 n(k)$ deviates more and more from the classical prediction of a (Gaussian) Maxwell-Boltzmann momentum distribution. At temperatures below T_c, a nonzero fraction $n_0(T)$ of the bosons occupies the zero-momentum state, and $n(k)$ includes a delta-function contribution equal to $n_0(T)\rho(2\pi)^3\delta(k)$. Thus $k^2 n(k)$ increases as k approaches 0 and exhibits a positive discontinuity at $k = 0$.

evident from the fact that they remain liquid all the way down to absolute zero.

In quantum mechanics the momenta of atoms in a finite-sized box are quantized in integer multiples of $\hbar L^{-1}$, where L is a dimension of the box. This quantization is a consequence of the requirement that the wave function of an atom must be zero at the boundaries of the box and therefore must have an integral number of nodes. Thus the momentum distribution for the system is discontinuous (unlike the continuous Maxwell-Boltzman distribution of classical systems) and becomes continuous only for a box of infinite size.

Now consider the momentum distribution for a non-interacting (ideal) gas of spin-0 atoms. At high temperature the atoms are thermally excited, and the probability of an atom being in any particular momentum state is inversely proportional to the size of the system (L^3) and proportional to a Maxwell-Boltzmann distribution. However, as the temperature is reduced, the preference of bosons for occupying the same momentum state causes deviations from the Maxwell-Boltzmann distribution. As the temperature is further reduced below a critical *Bose-condensation temperature*, a significant fraction of the atoms begin to occupy the lowest (or zero) momentum state. The fraction, n_0, is called the Bose-condensate fraction, and its value is independent of the size of the system. In the momentum distribution n_0 shows up as a delta function, of weight n_0, at $p = 0$. The width Δp of the momentum distribution for the remainder of the atoms is on the order of $\sqrt{mk_BT}$ and goes to zero at zero temperature. On the other hand, n_0 approaches one as the temperature approaches zero; that is, the entire system becomes a Bose condensate. Figure 1 shows plots of $k^2 n(k)$ versus k for a system of non-interacting bosons at various temperatures. (In this article momentum $\mathbf{p}$ and the wave vector $\mathbf{k} \equiv \mathbf{p}/\hbar$ are used interchangeably. The natural unit for wave-vector magnitudes is the inverse angstrom, $\mathring{A}^{-1}$.) Figure 2 is a plot of the Bose-condensate fraction n_0 versus temperature.

London reasoned that the superfluidity observed in ^{4}He was a macroscopic consequence of the microscopic Bose condensation of ^{4}He atoms into the zero-
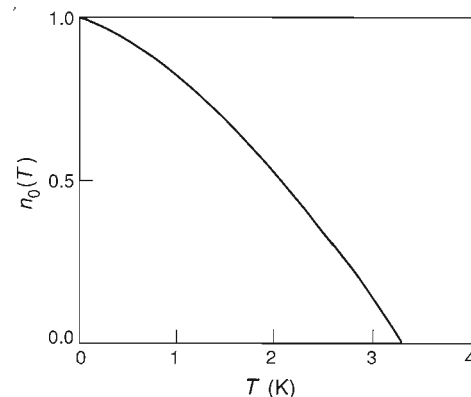
Fig. 2. The Bose-condensate fraction in an ideal Bose gas with a density equal to that of liquid ^{4}He increases monotonically from 0 at its Bose-condensation temperatue (3.3 kelvins) to 1 at absolute zero.

DEPARTURE OF LIQUID ^{4}He FROM IDEAL-GAS BEHAVIOR

Fig. 3. (a) The potential $V(r)$ for the interaction between atoms in liquid ^{4}He is steeply repulsive (positive) at interatomic distances below about 2.6 angstroms. At larger distances van der Waals forces cause the potential to be weakly attractive (negative). Such an interaction potential, which differs enormously from that of an ideal gas, leads to spatial correlations between the atoms in liquid ^{4}He. (b) Shown here is the pair-correlation function $g(r)$ for superfluid ^{4}He determined by neutron diffraction. It is proportional to the probability distribution for finding two atoms of superfluid ^{4}He a distance r apart. Comparison of $g(r)$ and $V(r)$ reveals that the atoms in the superfluid tend to stay in the attractive well of the potential, outside its steeply repulsive core.

(a) ^{4}He-^{4}He Interaction Potential

(b) Superfluid ^{4}He Pair-Correlation Function
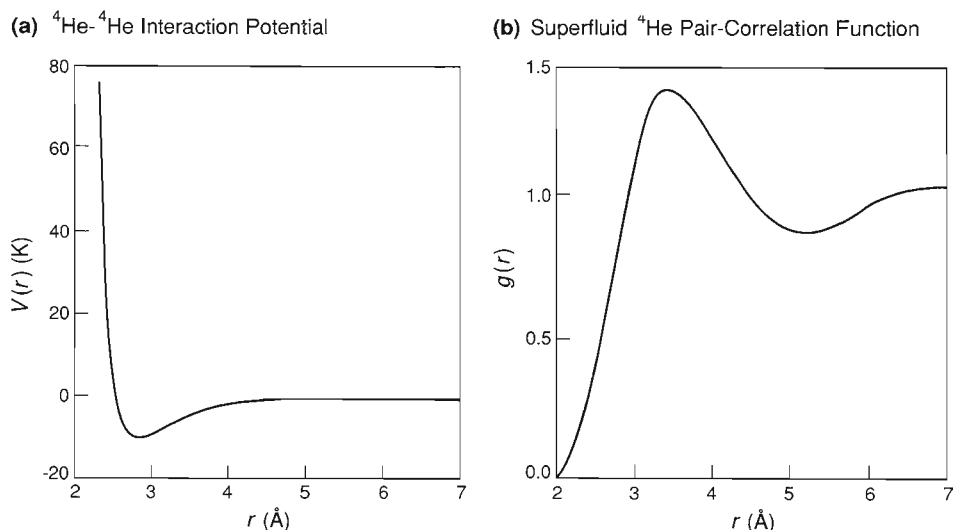
momentum state. He pointed out that the Bose-condensation temperature of a system of non-interacting atoms having the same mass and density as ^{4}He is 3.3 kelvins and is thus remarkably close to the observed ^{4}He superfluid-transition temperature of 2.17 kelvins. London's hypothesis also suggests why liquid ^{3}He behaves so differently from liquid ^{4}He at low temperature: The spin-statistics relation for fermions forbids Bose condensation of ^{3}He atoms. London's analogy between ^{4}He and a system of non-interacting bosons is imperfect because the atoms in liquid helium interact strongly. Such interatomic interactions have a significant effect on the momentum distribution of a quantum system. As shown in Fig. 3a, the interaction potential between ^{4}He atoms is strongly repulsive at interatomic distances less than 2.5 angstroms and weakly (van der Waals) attractive at larger distances. Therefore the atoms tend to stay a minimum of 2.5 angstroms apart. Indeed, as shown in Fig. 3b, the pair-correlation function, $g(r)$, for liquid ^{4}He (which is proportional to the probability distribution for finding two ^{4}He atoms a distance r apart) has a maximum at about 3.5 angstroms. Now, since liquid helium is a quantum system obeying the Heisenberg uncertainty principle, the correlation in the positions of the the atoms must result in a spread in the probability of their occupying any of the momentum states. In particular, the tendency of ^{4}He atoms to stay at least 2.5 angstroms apart results in an expected width of the momentum distribution of $\Delta k = (2\pi/2.5)$ Å^{-1}, or about 2 Å^{-1}. Thus, even at absolute zero interacting atoms have a finite kinetic energy of $(\hbar\Delta k)^2/2m$, which is termed the *zero-point energy*. The uncertainty in the atom's momentum increases its kinetic energy above the classical value at any temperature.

For most atomic systems the zero-point energy is too small to prevent crystallization at low temperatures. The only exceptions are ^{4}He and ^{3}He. At pressures below tens of atmospheres, helium atoms tend to sit in the shallow potential well created by the weakly attractive van der Waals force, but their comparatively low masses result in zero-point kinetic energies that are higher than the van der Waals potential energy. Only at very high densities are the atoms close enough that their potential energy due to the steeply repulsive part of the potential exceeds the zero-point energy and produces crystallization. Thus the low-temperature phase diagram for ^{4}He (Fig. 4) shows crystallization above 25 atmospheres. At lower pressures ^{4}He remains a liquid down to absolute zero. Moreover, along the so-called λ line ^{4}He undergoes the phase transition from the normal to the superfluid state.

The really daring aspect of London's hypothesis was to propose that, despite the strong interactions between ^{4}He atoms, which tend to broaden the momentum distribution, the superfluid should contain a non-negligible fraction of atoms in the zero-momentum state. The remaining atoms should have a broad momentum distribution
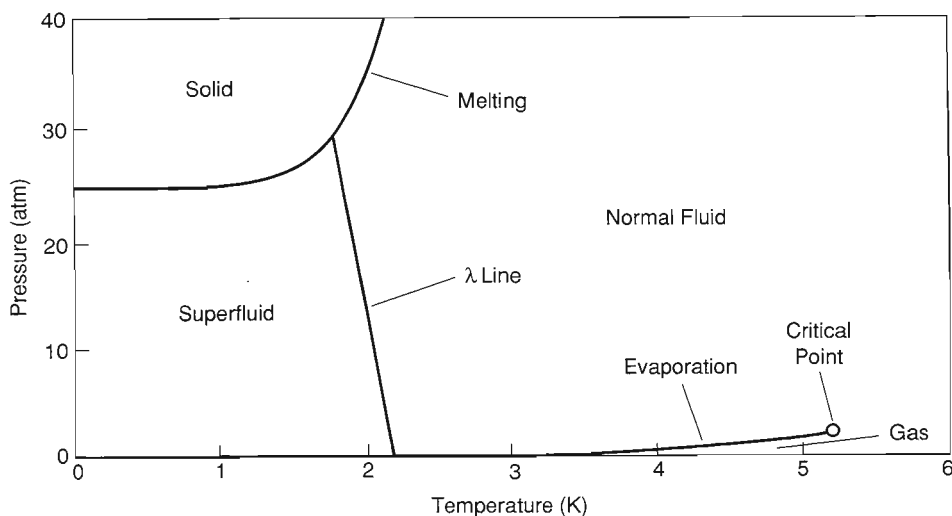
PHASE DIAGRAM OF ^{4}He

Fig. 4. At pressures below 25 atmospheres, the zero-point energy of liquid ^{4}He is sufficiently large to prevent its solidification, even at a temperature of absolute zero. The liquid phase is separated by the λ line into normal-fluid and superfluid phases. At atmospheric pressure the superfluid-transition temperature T_λ is 2.17 kelvins, which is not much different from the Bose-condensation temperature of an ideal Bose gas of the same density (3.3 kelvins). The critical point shown in the phase diagram gives the pressure and temperature at which the liquid and gas phases of ^{4}He form one phase.

that remains wide (about 2 Å^{-1}) even at zero temperature because of zero-point motion. (Superfluid flow is attributed to Bose condensation into a state other that the zero-momentum state. Coherent interactions among the ^{4}He atoms stabilize the flow against scattering processes, which tend to dissipate flow in normal fluids.)

London's hypothesis began what has been called the *condensate saga*—the story of numerous attempts by theorists to calculate the Bose-condensate fraction and by experimentalists to measure it. The condensate fraction has been elusive, and the saga has had many twists and turns. Both the existence and the size of the Bose-condensate fraction in superfluid ^{4}He have remained very controversial subjects. Most of the successful theories of superfluid helium (developed by, among others, Lev Landau, L. Tisza, and London) do not invoke a many-atom wave function at a microscopic level and especially not a Bose condensate. Rather, most of the remarkable properties of superfluid helium can be explained by the *two-fluid* model in which a macroscopic wave function extending throughout the sample corresponds to the superfluid component of the fluid, and elementary excitations out of the superfluid state, the so-called *phonons* and *rotons,* correspond to the normal component of the fluid. At zero temperature all the atoms are condensed into the macroscopic wave function, and helium consists entirely of superfluid. As the temperature is raised above zero, the number of thermally excited phonons and rotons increases, and they act as a normal fluid component. At T_λ, where the number of thermally excited elementary excitations equals the number of atoms in the system, the entire system becomes a normal fluid. This two-fluid description can account successfully for many experiments on superfluid ^{4}He. Thus establishing a theoretical connection between the two-fluid model and the many-atom wave function is unnecessary to successful prediction of most macroscopic experiments. However, a goal of microscopic (atomic level) theory has been to predict the parameters of the two-fluid model, now obtained by fitting the model to macroscopic experiments. Until we achieve such a connection between the microscopic and macroscopic theories, we cannot determine n_0 from macroscopic experiments.

Nevertheless, many theorists over the years have spent an enormous effort on the microscopic theory of superfluid ^{4}He. One goal has been to calculate n_0, and indeed the entire momentum distribution, from first principles (*ab initio*) and the measurable interatomic potential. In addition to scientists' gut-level "need-to-know," many other factors motivate this enterprise. Helium is arguably the simplest among a wide variety of strongly interacting many-body systems currently under intense study by condensed-matter physicists. Those systems relate to such fashionable topics as high-temperature superconductors, heavy-fermion metals, the quantized Hall effect, and so on. Countless theoretical methods developed for and tested on helium have

subsequently provided the key to understanding more complex systems. Also, superfluidity in helium shares many similarities with superconductivity in metals, so that work on helium can be directly related to technologically important questions. Finally, the microscopic theory of helium, one of the most strongly interacting many-body systems, presents a fundamental challenge.

Progress in *ab initio* many-body calculations of the properties of liquid helium has been dramatic, especially since the advent of supercomputers in the last decade. The names of the methods employed are characteristically arcane (variational wave function, Green's function Monte Carlo, and path-integral Monte Carlo, for example) and reflect their great diversity. Nevertheless, the results of all these methods have converged on the conclusion that the Bose-condensate fraction should exist in superfluid ^{4}He but not in the normal fluid and that its value should be about 10 percent at zero temperature. Hence, measurements of n_0 provide a fundamental test of the predictive power of modern condensed-matter theory.

A seminal suggestion made by P. Hohenberg and P. Platzman in 1966 still remains the best hope for a direct measurement of momentum distributions in liquid helium. It involves neutron-scattering experiments at momentum and energy transfers sufficiently high that the struck atoms acquire kinetic energies much larger than the binding energies in the liquid. Under such conditions (hopefully!) the *impulse approximation*, which approximates the scattering from the many-atom system as the sum of scatterings from individual free atoms, is valid, and the observed scattering cross section is a direct measure of the momentum distribution in the sample. Analogous experiments are of interest in all of modern physics because momentum distributions are measurable properties of all many-particle wave functions, provided the energy and momentum transfers are high enough. X-ray Compton scattering at energies on the order of tens of keV can measure electron momentum distributions in atoms and solids, electron scattering at a few GeV can measure nucleon momentum distributions in nuclei, and electron scattering at hundreds of GeV can measure quark momentum distributions in nucleons. Scattering at very high energy transfers relative to the binding energies of a system is called *deep inelastic scattering*.

The suggestion of Hohenberg and Platzman initiated an effort that has lasted more than twenty years and has involved over one hundred experimentalists at reactor and pulsed neutron sources all over the world. The effort has spawned improvements in spectrometers and advances in data-analysis procedures. The attempted momentum-distribution measurements have used neutrons with energies between 10 and 1000 meV. Increases in neutron energy have been made in order to come closer to the regime in which the impulse approximation is valid. The goal of inferring the value of n_0 from the data inevitably involves accounting for instrumental broadening, statistical and background uncertainties, and corrections to the impulse approximation necessitated by the finite neutron energies available. Although the condensate saga through 1987 included many (often conflicting) indirect determinations of n_0, it did not include any direct observation of the delta function in the momentum distribution of superfluid ^{4}He predicted by the London theory. Direct experimental evidence for a Bose condensate in superfluid helium was weak.

This article presents the latest episode in the condensate saga—a recent breakthrough in confirming the existence and size of the Bose-condensate fraction. New experiments using the high epithermal flux of a pulsed neutron source are in excellent agreement with sophisticated new *ab initio* calculations of momentum distributions, provided that the prediction of the impulse approximation is broadened by final-state interactions according to a first-principles theory developed by the author. Experiment and theory on the momentum distributions of ^{4}He have converged, and both are consistent with a condensate fraction of 9.2 percent in the superfluid at zero temperature. The story of the breakthrough is accompanied by a discussion of its implications for the study of other condensed-matter systems.

Neutron Scattering and the Impulse Approximation

As mentioned above, the best hope for measuring momentum distributions in helium is neutron scattering at high momentum and energy tranfers. To understand why, let's review what we learn from neutron-scattering experiments (see also "Neutron Scattering—A Primer" in this issue). The double differential scattering cross section, or the scattering per unit solid angle Ω and per unit energy transfer $\hbar\omega$, is defined as

$$\frac{d^2\sigma}{d\omega d\Omega} \equiv \frac{\sigma_{total}^n}{4\pi} \frac{k_i}{k_f} S(Q,\omega), \tag{1}$$

where σ_{total}^n is the total neutron-scattering cross section of a single helium atom, $\mathbf{k}_i$ and $\mathbf{k}_f$ are the initial and final neutron wave vectors, the energy transfer $\hbar\omega = (\hbar^2/2m_n)(k_i^2 - k_f^2)$, m_n is the neutron mass, $\hbar Q \equiv \hbar|\mathbf{k}_i - \mathbf{k}_f|$ is the momentum transfer, and $S(Q,\omega)$ is the *dynamic structure factor* (also sometimes termed the *neutron scattering law*).

As discussed in the primer, $S(Q,\omega)$ is the Fourier frequency and spatial transform of the time-dependent pair-correlation function of the liquid. Measurement of $S(Q,\omega)$ provides a rich variety of information on the properties of quantum fluids. By the uncertainty principle experiments at Q values smaller than or comparable to the inverse of the interatomic spacing (that is, at $Q \leq$ a few Å^{-1}) are sensitive to the collective behavior of helium atoms. For example, *diffraction* experiments, which involve elastic scattering ($k_i = k_f$), yield the *static structure factor* $S(Q) \equiv \int d(\hbar\omega)S(Q,\omega)$. The Fourier transform of $S(Q)$ for ^{4}He yields the pair-correlation function shown in Fig. 3b. And experiments involving *inelastic* scattering ($k_i \neq k_f$) at low Q determine the spectrum of elementary excitations (phonons and rotons) in ^{4}He.

On the other hand, inelastic-scattering experiments at Q values much larger than a few Å^{-1} probe individual atoms rather than collective behavior. That fact led to Hohenberg and Platzman's suggestion for using such experiments to measure the momentum distributions in helium. They assumed, first, that if the energy transfer is large compared to typical binding energies in the condensed phase (which are on the order of meV), then the initial binding energy of the atoms could be ignored, and second, that if the energy transfer is very large compared to the potential energy of the scattered atoms in the condensed phase, then a struck atom in its final state will be negligibly affected by the surrounding atoms and may be considered to be a free particle. Those assumptions imply that neutron scattering at very high energy and momentum transfers may be approximated as scattering from a collection of free helium atoms with initial momenta $\mathbf{p}_i$ distributed according to the momentum distribution $n(\mathbf{p}_i)$ and with final momenta $\mathbf{p}_f = \mathbf{p}_i + \hbar Q$ (Fig. 5). The approximation that the neutrons scatter from a collection of free atoms is termed the impulse approximation (IA). The dynamic structure factor in the impulse approximation is given by

$$S_{IA}(Q,\omega) = \frac{1}{\rho} \int \frac{d^3\mathbf{p}_i}{(2\pi\hbar)^3} n(\mathbf{p}_i)\delta(\hbar\omega - E_{p_f} + E_{p_i}), \tag{2}$$

where ρ is the density of the liquid helium, $E_p = p^2/2m$, m is the mass of the helium atom, and the delta function is an expression of energy conservation. Note that $S_{IA}(Q,\omega)$ is normalized so that its integral over $\hbar\omega$ is unity at large Q.

The important feature of the impulse approximation is that it provides the desired simple relation between the neutron scattering law for helium and its momentum distribution. But how do we know whether an experiment has been performed at conditions for which the impulse approximation is valid? Equation 2 implies that, at a given value of Q, a plot of the observed $S(Q,\omega)$ versus ω should have a single peak that is symmetric about the recoil energy of an atom at rest, $\hbar\omega_{recoil} = E_{\hbar Q}$.
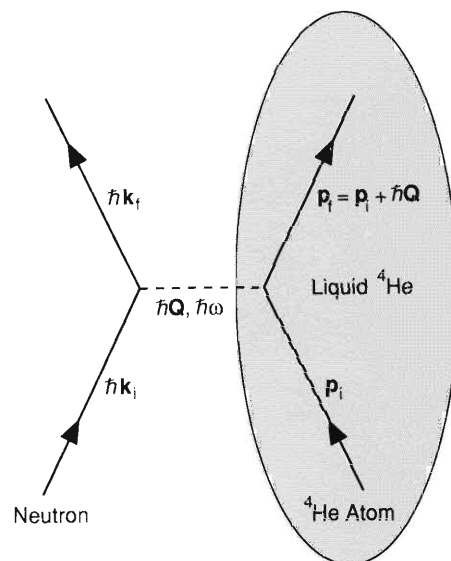
KINEMATICS FOR IMPULSE APPROXIMATION

Fig. 5. According to the impulse approximation, neutron scattering from liquid ^{4}He can be regarded, at sufficiently high momentum and energy transfers, as scattering from free atoms. That situation is depicted here schematically. Arrows represent the momenta of a neutron and a helium atom before and after scattering. The dashed line represents the momentum $\hbar Q$ and energy $\hbar\omega$ transferred to the helium atom during the scattering. Momentum conservation demands that $\mathbf{p}_f = \mathbf{p}_i + \hbar Q$, where $\mathbf{p}_i$ and $\mathbf{p}_f$ are, respectively, the initial and final momenta of the helium atom. Because the binding energy of the helium atoms is ignored in the impulse approximation, the difference between the energies of a helium atom before and after scattering is simply equal to the difference in its kinetic energies before and after scattering, $(1/2m)(p_f^2 - p_i^2)$. Furthermore, energy conservation requires that $(1/2m)(p_f^2 - p_i^2) = \hbar\omega$.

Moreover, since $\hbar\omega - E_{p_f} + E_{p_i} = \hbar\omega - \hbar\omega_{\text{recoil}} - \mathbf{Q} \cdot \mathbf{p}_i/m$, the peak width should be proportional to Q times the width of the momentum distribution.

In the presence of a Bose condensate, the initial momentum distribution should have the form

$$n(\mathbf{p}_i) = n_0\rho(2\pi\hbar)^3\delta^{(3)}(\mathbf{p}_i) + n^*(\mathbf{p}_i), \tag{3}$$

where $\delta^{(3)}(\mathbf{p}_i)$ is a three-dimensional delta function and $n^*(\mathbf{p}_i)$ is a smooth momentum-distribution function for the remainder of the atoms. Combining Eqs. 2 and 3 yields a unique form for the dynamic structure factor in the impulse approximation:

$$S_{\text{IA}}(Q,\omega) = n_0\delta(\hbar\omega - \hbar^2Q^2/2m) + S_{\text{IA}}^*(Q,\omega), \tag{4}$$

where $S_{\text{IA}}^*(Q,\omega)$ is the contribution to the scattering law from $n^*(\mathbf{p}_i)$. A Bose-condensate peak in the momentum distribution should show up as a delta function in the observed $S(Q,\omega)$ at $\hbar\omega = E_{\hbar Q}$, and n_0 should equal the fraction of the integral of the observed $S(Q,\omega)$ over $\hbar\omega$ contributed by the delta function. It is this simple prediction that has motivated the condensate saga, that is, the many attempts to observe the condensate by neutron-scattering experiments.

More generally, the impulse approximation, which assumes the helium atoms scatter neutrons as if they are free particles, predicts that the scattering is no longer a function of ω and Q separately. Rather, the energy-conservation delta function in Eq. 2 forces a relationship between ω, Q, and $k_\parallel$, the component of an atom's initial wave vector parallel to the direction of the momentum transfer: $k_\parallel = (m/\hbar^2Q)(\hbar\omega - \hbar^2Q^2/2m)$. To express this fact we can introduce a new variable Y, first suggested by Geoffrey West:

$$Y \equiv \frac{m}{\hbar^2Q}\left(\hbar\omega - \frac{\hbar^2Q^2}{2m}\right). \tag{5}$$

Then we rewrite $S(Q,\omega)$ in terms of a function that depends on Y rather than ω:

$$J(Y,Q) \equiv \hbar^2QS(Q,\omega)/m. \tag{6}$$

Eqations 5 and 6 are essentially redefinitions (since Y is just a dummy variable and has not been given a physical interpretation) and are therefore valid whether or not the impulse approximation is valid. The quantity $J(Y,Q)$ is termed the *Compton profile* and was used to plot the results of seminal experiments by A. H. Compton and J. DuMond in the 1920s that measured the electron momentum distributions in atoms and metals by x-ray scattering at keV energies. Compton and Dumond plotted their results in terms of $p_\parallel$, but, because the impulse approximation was valid in their experiments, $p_\parallel$ is identically equal to $\hbar Y$.
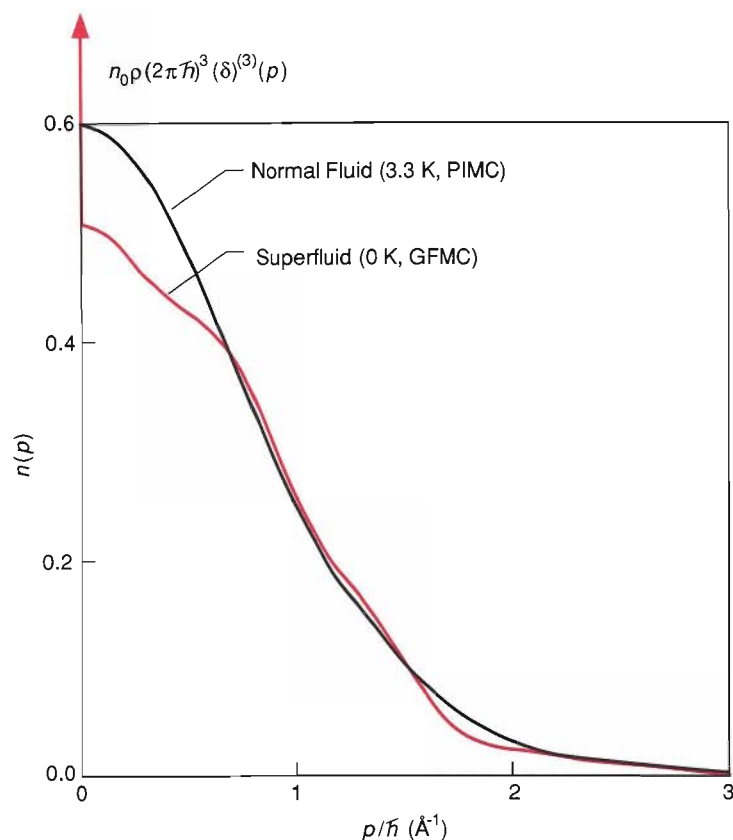
The advantage of expressing scattering laws as Compton profiles is that, in the impulse approximation, $J(Y,Q)$ depends only on Y and not on Q, a phenomenon we refer to as Y-*scaling*. In particular the Compton profile for liquid helium in the impulse approximation is given by

$$J_{\text{IA}}(Y) = n_0\delta(Y) + \frac{1}{4\pi^2\rho}\int_{|Y|}^{\infty} dk\, kn(k). \tag{7}$$

Note that the momentum distribution in Eq. 7 depends only on the magnitude of $\mathbf{k}$ because the liquid is isotropic. If the impulse approximation is valid, then it should be possible to extract n_0 and $n(k)$ from neutron scattering experiments by direct inversion of Eq. 7.

We will now examine whether conditions for the impulse approximation, which implies Y-scaling, have been met experimentally. We will also consider the possibility, first suggested by West, that Y-scaling is independent of the validity of the im-

(a) Momentum Distributions

$$n_0 \rho (2\pi \hbar)^3 (\delta)^{(3)}(p)$$

0.6 — Normal Fluid (3.3 K, PIMC)

— Superfluid (0 K, GFMC)

0.4

$n(p)$

0.2

0.0

$p/\hbar$ (Å^{-1})

(b) Compton Profiles

$$n_0 \delta(Y)$$

Superfluid (0 K, GFMC) — Normal Fluid (3.3 K, PIMC)

$J_{IA}(Y)$

Y (Å^{-1})
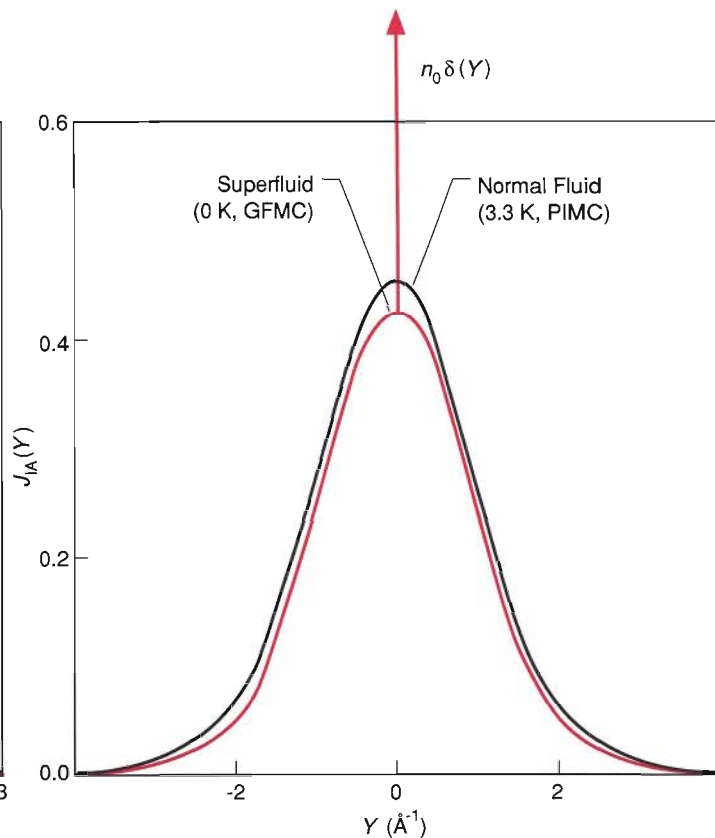
pulse approximation. In other words, even when conditions for the impulse approximation are not met, the Compton profile may still depend only on Y. Scaling phenomena are of intense interest throughout modern physics, for they occur whenever the number of variables involved in a measurement on a physical system exceeds the number of relevant length or energy scales for the system. Thus, observation of scaling is indicative of a fundamental simplicity of the physical system being studied. In the case of high-energy neutron scattering from helium, the number of relevant variables might, in principle, be on the order of 10^{23}, the number of atoms in a typical sample. Observation of Y-scaling would indicate that the number of relevant variables has been reduced to one. In the impulse approximation the relevant variable is $k_{\parallel}$, and it determines the scattering through the momentum distribution. Later when we discuss corrections to the impulse approximation due to final-state effects, we will present a theory in which the scattering law obeys Y-scaling but Y does not equal $k_{\parallel}$ and has an alternative interpretation.

Theoretical Predictions

To advance from qualitative arguments to quantitative predictions for neutron-scattering experiments, we need to consider the theoretical predictions for momentum distributions that are used as input to the impulse approximation. A wide variety of many-body calculational methods have been developed and applied to calculate the momentum distributions in helium. Remarkably, the most straightforward method, which involves perturbative expansion in the potential about the noninteracting ground state using Feynman-diagram (field-theory) methods, works poorly for helium. The reason is that the He-He potential (Fig. 3a) is singular (infinite) at short distances. Thus, an infinite-order resummation of the perturbative expansion is required in order to obtain nonsingular answers. (We return to this point in the sidebar "How Final-State Effects Were *Really* Calculated.") However, several

PREDICTION OF BOSE CONDENSATE BY *AB INITIO* THEORY

Fig. 6. (a) An *ab initio* calculation of the momentum distribution $n(p)$ in superfluid helium at absolute zero (red curve) exhibits a delta-function spike at $p = 0$, which is interpreted as the signature of a Bose-condensate fraction in the superfluid of about 9.24 percent. As expected, an *ab initio* calculation of $n(p)$ in normal-fluid helium at 3.3 kelvins lacks a delta-function spike at $p = 0$. The acronyms GFMC and PIMC refer to calculational methods described in the text. (b) The theoretical Compton profiles $J_{IA}(Y)$ shown here were calculated by using the theoretical momentum distributions as input to the impulse approximation. The integral over Y of each Compton profile is unity. The fraction of the integral of the 0.32-kelvin $J_{IA}(Y)$ contributed by its delta-function spike is the Bose-condensate fraction at that temperature.

methods have been developed that avoid the pathologies of perturbative expansions for strong potentials.

One approach is to invoke a quantum-mechanical variational principle that allows one to calculate the momentum distribution by minimizing the energy of trial wave functions for the ground state of ^{4}He. The most successful of such variational methods is the *hypernetted chain* (HNC) approximation. Another approach, termed the *Green's function Monte Carlo* (GFMC) method, is to use a stochastic Monte Carlo algorithm to solve the many-body Schrödinger equation for the ground-state wave function. A third approach, termed the *path-integral Monte Carlo* (PIMC) method, is to invoke the path-integral formulations of quantum mechanics and statistical mechanics originally proposed by Feynman to solve for the momentum distribution as a sum over classical paths in imaginary time. The path-integral method is especially applicable to nonzero temperatures, whereas the variational and Green's function methods, which calculate the ground-state wave function, yield the momentum distribution at zero temperature only.

Figure 6a shows the momentum distribution predicted by the GFMC method for superfluid ^{4}He at absolute zero and the momentum distribution predicted by the PIMC method for normal-fluid ^{4}He at 3.3 kelvins. The GFMC method yields a delta function in the zero-temperature momentum distribution at $p = 0$ corresponding to a Bose-condensate fraction of 9.2 percent, much less than the non-interacting value of 100 percent. (The HNC method also yields a delta function at $p = 0$ and the same value for n_0.) The PIMC calculation, which has been carried out only for $T > 1$ kelvin, shows a condensate fraction that tends toward the GFMC and HNC value at the lowest temperature calculated and tends toward zero as the temperature approaches T_λ. At $T \geq T_\lambda$ the PIMC method yields a momentum distribution that is smooth and approximately Gaussian, as seen in Fig. 6a for $T = 3.3$ kelvins. The widths of all these momentum distributions are roughly equal to the $\Delta k \approx 2$ Å^{-1} that we estimated heuristically from examination of the He-He potential and the pair-correlation function for liquid ^{4}He.

Using the GFMC and the PIMC momentum distributions of Fig. 6a to calculate the Compton profiles in the impulse approximation yields the theoretical predictions for $J_{IA}(Y)$ shown in Fig. 6b. Note that a delta-function peak at $Y = 0$ is predicted for the superfluid at temperatures below T_λ, whereas a smooth $J_{IA}(Y)$ is predicted for the normal fluid at temperatures above T_λ.

Experimental Results

The critical experimental issue for momentum-distribution measurements is to achieve conditions at which the impulse approximation may be valid, namely, high values of Q and ω. To do so requires neutrons with relatively high energies, on the order of hundreds of meV. Because the spectrum of neutrons from a reactor is a Maxwell-Boltzmann distribution that peaks at 23 meV, the neutron flux decreases exponentially with further increases in energy. Therefore, reactor experiments are generally limited to Q values less than 12 Å^{-1}. Many experiments on ^{4}He have been done at reactor sources, and the most carefully analyzed data lie in the Q range between 4 Å^{-1} and 7 Å^{-1}. However, in this Q range the deviations from the impulse approximation are large.

More recently, experiments have also been done at pulsed neutron sources. In order to maintain short pulse widths for time-of-flight experiments, pulsed neutron sources have an undermoderated neutron specturm; that is, a relatively large fraction of the neutrons fail to reach a thermal distribution before they exit the moderator. Thus, the flux of high-energy neutrons decreases only inversely with increasing energy, so higher energy and momentum transfers are achieved. In 1986 a team headed by Paul Sokol of The Pennsylvania State University organized an effort to build a
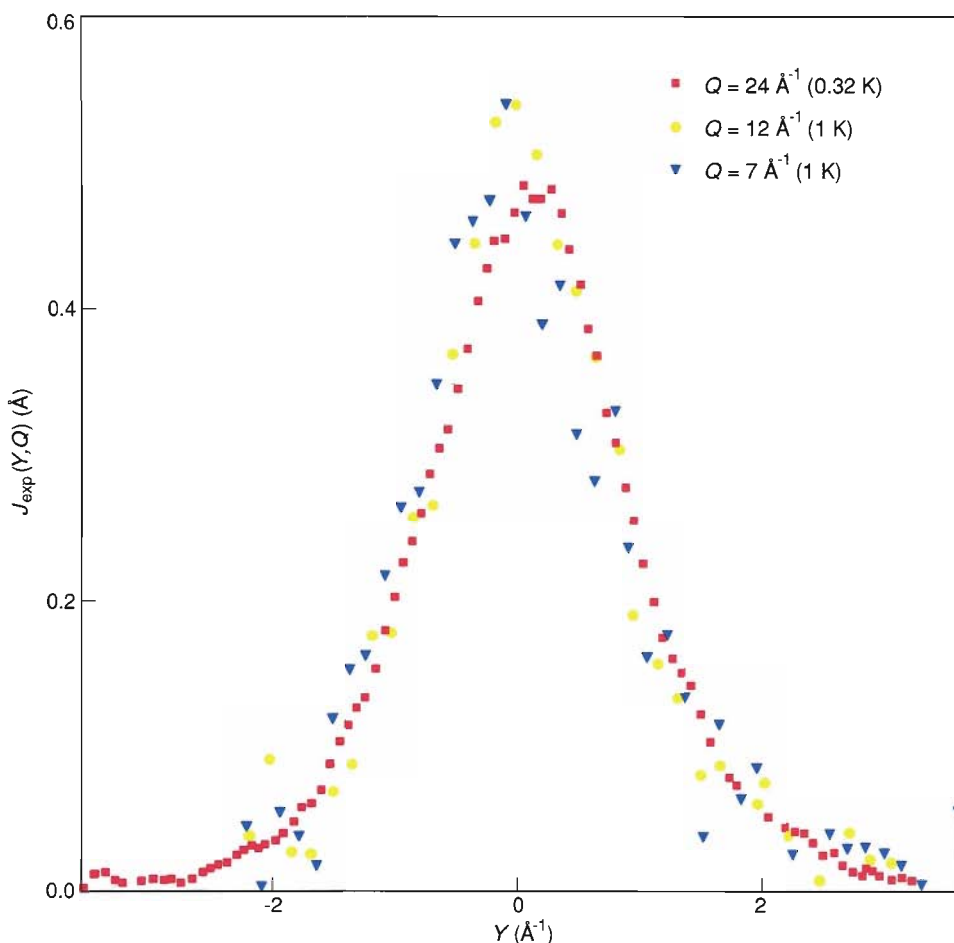
**Y-SCALING OF MEASURED COMPTON PROFILES FOR SUPERFLUID ^{4}He**

Fig. 7. Each of the measured Compton profiles, $J_{exp}(Y, Q)$, for superfluid ^{4}He shown here is broadened by a Gaussian instrumental resolution function with a full width at half maximum of about 0.6 Å^{-1}. The 7-Å^{-1} and the 12-Å^{-1} data were taken at a reactor, and the 24-Å^{-1} data were taken at a pulsed neutron source. Note that each data set lies approximately on the same curve irrespective of the Q value at which it was obtained. In other words, the measured Compton profiles exhibit the Y-scaling predicted by the impulse approximation. Also note that small deviations from the impulse approximation, in the form of an asymmetry in the peak shape and a leftward shift of the peak center, are visible in the 7- and 12-Å^{-1} data.

chopper spectrometer optimized for momentum-distribution studies. The spectrometer, known as Phoenix, was built at the Intense Pulsed Neutron Source at Argonne National Laboratory. Phoenix has provided data at Q values up to 24 Å^{-1}. The pulsed-source experiments thus more closely approach conditions at which the impulse approximation might be valid. They also permit much greater control of the instrumental resolution function than do reactor experiments.

Figure 7 shows experimental data for the Compton profile of superfluid ^{4}He. (The 7-Å^{-1} and 12-Å^{-1} data were obtained at a reactor source and the 23-Å^{-1} data were obtained at a pulsed neutron source. Each data set is broadened by an approximately Gaussian instrumental resolution function with a full width at half maximum, ΔY_{FWHM}, of about 0.6 Å^{-1}.) At first glance all the data sets look symmetric and centered at $Y = 0$. Moreover, the widths and shapes of the $J_{exp}(Y, Q)$ appear to be independent of Q, in agreement with our expectations for Y-scaling of the Compton profile. The width not only fits our heuristic estimate of 2 Å^{-1}, which was based on a consideration of zero-point energies, but also is comparable to the prediction of *ab initio* theory (see Fig. 6b).

However, a more careful examination reveals some discrepancies with this simple picture. Note that the 7-Å^{-1} data are slightly asymmetric and the peak center is shifted to the left of $Y = 0$. The asymmetry is still present but smaller in the 12-Å^{-1} data and is smallest in the 23-Å^{-1} data. Additional reactor data at $Q \leq 12 \text{ Å}^{-1}$ show that the width of the peak does not remain constant but rather oscillates about an average value as Q increases. These observations call into question the validity of the impulse approximation for the reactor Q range. Moreover, no reactor experiment has shown a well-resolved delta-function peak due to a Bose condensate.

Heretofore the above discrepancies have been minimized by data-analysis procedures that include symmetrizing the data and averaging over several Q. The data-
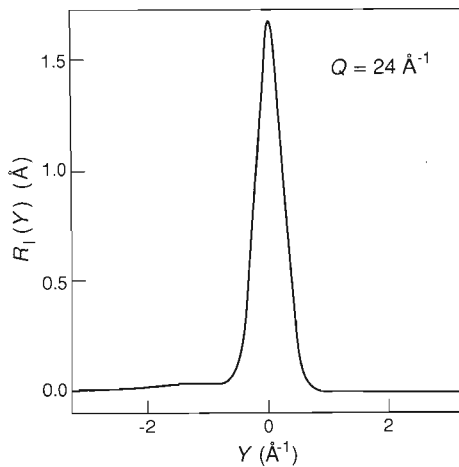
INSTRUMENTAL RESOLUTION FUNCTION FOR PHOENIX

Fig. 8. The chopper spectrometer known as Phoenix is located at the Intense Pulsed Neutron Source at Argonne National Laboratory. Its instrumental resolution function, $R_I(Y)$, at Q = 24 Å^{-1} was calculated by a Monte Carlo simulation and checked by experiments on samples whose scattering is well documented. Theoretically predicted Compton profiles must be convolved with $R_I(Y)$ before being compared with Phoenix data.

analysis procedures have assumed that some additional unknown mechanism broadens the impulse approximation and thereby broadens the delta function that would be produced by the Bose condensate. To determine the Bose-condensate fraction, the data are fit to a model for the momentum distribution in which n_0 is the only free parameter, and data at $T > T_\lambda$ are used to fix the momentum distribution for the remainder of the atoms. Some authors claimed that these procedures lead to a value for the condensate fraction of about 10 percent. The rune in the opening illustration of this article was drawn in 1983 to celebrate those claims and the seventy-fifth jubilee of the discovery of liquid helium. However, the data-analysis procedures leading to the claims had a serious conceptual error. The most credible value for n_0 obtained by model-fitting procedures is between 4 and 5 percent, in serious disagreement with *ab initio* calculations of momentum distributions. That is where the condensate saga stood in 1986 when we entered the picture.

As mentioned earlier, three new elements have converged to resolve the question of the momentum distributions in ^{4}He. First, instruments at pulsed neutron sources have yielded more accurate data at higher Q values. Second, highly accurate many-body calculations of momentum distributions have become feasible on supercomputers. And third, a new theory for the broadening of the impulse approximation due to final-state effects has been developed by this author. The new theory of final-state effects eliminates model fitting and permits a direct comparison between *ab initio* theory and experiment.
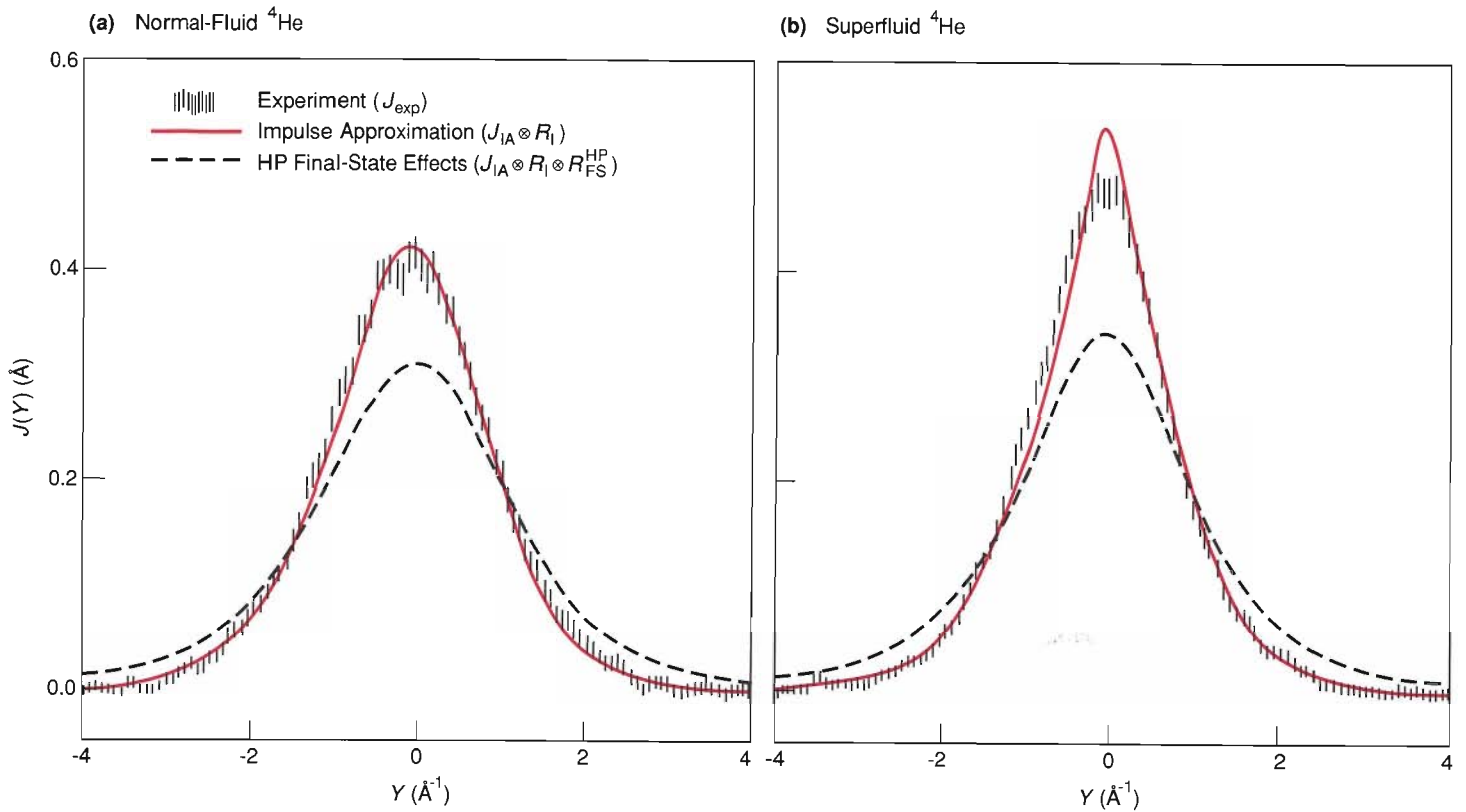
Before discussing final-state effects, we need to show more clearly the discrepancies between theory and experiment. The best data for comparison with theory is the pulsed-source data of Sosnick, Snow, Sokol, and Silver obtained at Q = 24 Å^{-1} with the Phoenix instrument. To compare theory with the measured Compton profile, we must first determine the instrumental resolution function of the spectrometer, $R_I(Y)$. We used a Monte Carlo simulation of the Phoenix spectrometer to calculate $R_I(Y)$; the result (Fig. 8) was confirmed by experiments on samples whose scattering is well known. We then convolve $R_I(Y)$ with the theoretical Compton profile, $J(Y, Q)$, to obtain a prediction for $J_{exp}(Y, Q)$, the measured Compton profile:

$$J_{exp}(Y, Q) = R_I \otimes J \equiv \int_{-\infty}^{\infty} dY' R_I(Y - Y', Q) J(Y', Q), \qquad (8)$$

where the symbol $\otimes$ denotes *convolution*.

Figure 9a compares theory and experiment for the normal fluid at 3.3 kelvins. The theoretical prediction for $J_{exp}(Y, Q)$ was calculated by using the PIMC momentum distribution for the normal liquid as input to the impulse approximation and convolving the result with $R_I(Y)$ according to Eq. 8. Despite the absence of adjustable parameters, the agreement between *ab initio* theory and experiment is excellent.

A similar comparison between theory and experiment for the superfluid at 0.32 kelvin is shown in Fig. 9b. The fact that the superfluid data are more sharply peaked around Y = 0 than the normal-fluid data suggests changes in the momentum distribution that might be associated with the presence of a Bose condensate. The theoretically predicted $J_{exp}(Y, Q)$ for the superfluid was calculated by using the GFMC momentum distribution as input to the impulse approximation and convolving the result with $R_I(Y)$. Although the predicted width of $J(Y)$ agrees with experiment, the data are much less sharply peaked in the region around Y = 0 than the impulse-approximation prediction. If we now fit the data using the data-analysis procedures previously applied to reactor data, then we might conclude that the Bose-condensate fraction in the superfluid is much smaller than the theoretical value of 9.2 percent. Instead we believe that the discrepancies apparent in Fig. 9b between *ab initio* theory and experiment require corrections to the impulse approximation according to a new theory of final-state effects.

(a) Normal-Fluid ^{4}He   **(b)** Superfluid ^{4}He

Legend:
- ‖ Experiment (J_{exp})
- Impulse Approximation ($J_{IA} \otimes R_I$)
- HP Final-State Effects ($J_{IA} \otimes R_I \otimes R_{FS}^{HP}$)

Axis labels: $J(Y)$ (Å), Y (Å^{-1})

Final-State Effects

The impulse approximation, the basis for all the *ab initio* calculations discussed so far, assumes that the incident neutrons scatter from free helium atoms. In reality, however, after an atom is struck by a neutron, it interacts with neighboring atoms in the liquid (see opening illustration). Those interactions, which result in a broadening of the impulse approximation, are called *final-state effects*. Although the last two decades have produced many calculations of final-state effects, the various theories have been controversial and many conflicting results have been published. As a prelude to our recent work on final-state effects, we shall review only those theories that directly contribute to our current understanding.

The earliest and simplest theory of final-state effects was put forward in 1966 by Hohenberg and Platzman. Here we give the heuristic argument leading to their result for final-state broadening. After a helium atom with initial momentum p_i is struck by a neutron, it collides with neighboring atoms at a rate $1/\tau = \rho\sigma_{total}(Q)\hbar Q/m$, where τ is the average time between collisions, $\sigma_{total}(Q)$ is the total cross section for scattering of a helium atom with momentum $\hbar Q$ from other helium atoms, and we assume $\hbar Q \gg p_i$. The Heisenberg uncertainty principle implies that the energy of the recoiling atom should have an uncertainty of $\hbar/\tau$. To account for that uncertainty, we alter Eq. 2 by adding to the final neutron energy E_{p_f} an imaginary part called a *self energy*, $\Sigma_{p_f} = -i\hbar/2\tau$. The energy-conserving delta function in Eq. 2, $\delta(\hbar\omega - E_{p_i} + E_{p_f}) \equiv \delta(\Delta E) = (1/2\pi\hbar)\int_{-\infty}^{\infty} dt \exp(it\Delta E/\hbar)$, is replaced by a new delta function:

$$\delta(\Delta E - \Sigma_{p_f}) = \frac{1}{2\pi\hbar} \int_{-\infty}^{\infty} dt \exp\left(\frac{it\Delta E}{\hbar} - \frac{i|t|\Sigma_{p_f}}{\hbar}\right). \qquad (9)$$

With this replacement it is straightforward to derive that $J(Y,Q)$ is a convolution of the impulse approximation with a broadening function due to final-state effects, $R_{FS}(Y,Q)$; that is,

$$J(Y,Q) = R_{FS} \otimes J_{IA}. \qquad (10)$$

EXPERIMENTAL COMPTON PROFILES COMPARED WITH THEORY

Fig. 9. The experimental Compton profiles, $J_{exp}(Y,Q)$, for (a) normal-fluid ^{4}He and (b) superfluid ^{4}He are based on 24-Å^{-1} Phoenix data obtained at 3.5 kelvins and 0.32 kelvin, respectively. Shown for comparison (solid curves) are the Compton profiles calculated by convolving the predictions of *ab initio* theory and the impulse approximation (Fig. 6b) with the Phoenix instrumental resolution function (Fig. 8). Note that the impulse approximation agrees well with experiment only in the case of normal-fluid ^{4}He. Also shown (dashed curves) are the Compton profiles calculated by convolving the solid curves with Hohenberg and Platzman's final-state-broadening function, $R_{FS}^{HP}(Y)$. Note that the dashed curves deviate substantially from the data.
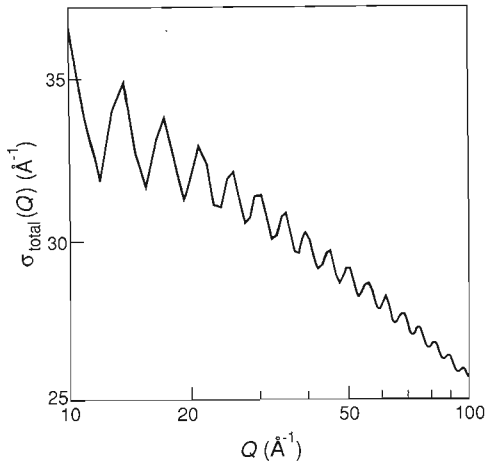
TOTAL ^{4}He-^{4}He SCATTERING CROSS SECTION

Fig. 10. The small oscillations, or hard-sphere glories, in this plot of $\sigma_{\text{total}}(Q)$ versus Q are an interference effect due to the symmetry requirement on the two-boson wave function. However, apart from those oscillations the cross section decreases linearly with ln Q. As a result, the neutron scattering law also has a logarithmic dependence on Q at large Q and thus approaches the impulse approximation (which is independent of Q) only very slowly (see Fig. 13).

The final-state broadening function should have a Lorentzian form:

$$R_{\text{FS}}(Y,Q) = \frac{1}{\pi}\frac{\Gamma}{Y^2 + \Gamma^2},$$

where

$$\Gamma = \frac{1}{2}\rho\sigma_{\text{total}}(Q). \tag{11}$$

As we will see later, many features of this early theory coincide with our current model of final-state broadening. For example, both models predict that the full width at half maximum of the broadening has the form

$$\Delta Y_{\text{FWHM}} \approx \rho\sigma_{\text{total}}(Q). \tag{12}$$

Moreover, if the potential is infinitely steep (*hard core*) and thus $\sigma_{\text{total}}(Q)$ is independent of Q, the final-state-broadened $J(Y,Q)$ of the Hohenberg and Platzman model obeys Y-scaling even if the impulse approximation is not applicable, a result in accord with the original suggestion of West and a feature of our model also. Unfortunately, nature has not been so kind: We see in Fig. 10 that, instead of being constant, $\sigma_{\text{total}}(Q)$ is a logarithmic function of Q (apart from small *glory* oscillations, which are an interference effect due to the Bose-Einstein spin-statistics relation). The logarithmic dependence of $\sigma_{\text{total}}(Q)$ reflects the steepness of the He-He potential at short distances (see Fig. 3a). It also implies that even at very high Q the impulse approximation cannot be applied directly. Instead Y-scaling is approximately true (the deviations vary as $\ln Q$), and corrections for final-state effects must be made.

Although the Hohenberg and Platzman model shares many features with our present model, it is not completely correct for two reasons. First, the exact *kinetic-energy sum rule* (ω^2) on $S(Q,\omega)$ requires that, at high Q,

$$\int_{-\infty}^{\infty} dY \, Y^2 R_{\text{FS}}(Y,Q) = 0. \tag{13}$$

That is, final-state effects should not affect the second moment, or Gaussian width, of $J(Y,Q)$. The form for $R_{\text{FS}}(Y,Q)$ given in Eq. 11 does not satisfy the kinetic-energy sum rule and, in fact, yields an infinite value for the integral in Eq. 13. Second, the broadening predicted by Eq. 11 is much larger than is observed experimentally. For example, Fig. 9a shows that convolution of the Hohenberg-Platzman broadening function with the PIMC impulse-approximation prediction for the normal fluid yields a $J(Y)$ that is in serious disagreement with the data.

The additional qualitative physics required to complete the theory for final-state effects was first proposed by Gersch and Rodriguez in 1973, but their results were ignored in more than twenty subsequent theoretical papers on final-state effects as well as in the many papers analyzing the reactor experiments. In 1987 this author independently developed a theory embodying the same qualitative physics but implying new many-body techniques (see "How Final-State Effects Were *Really* Calculated"). The new theory predicted exactly the results obtained soon after from the new pulsed-source experiments.

Here we present a heuristic description of the new theory. As discussed earlier, the atoms in liquid ^{4}He are not uniformly distributed in space; rather they are distributed according to the pair-correlation function shown in Fig. 4. At Q values of many Å^{-1}, the motion of an atom recoiling from a neutron collision can be described heuristically by a classical trajectory. Figure 3b shows that initially an atom is likely to be in the attractive part of the potential and that, after being struck by a neutron, the atom travels for some distance before it begins to collide with the

steeply repulsive cores of the potentials of neighboring atoms. The collision rate, $1/\tau$, depends on the recoil distance $x = \hbar Q/m$: $1/\tau(x) \approx \rho g(x)\sigma_{total}(Q)\hbar Q/m$. Since $g(x)$ is zero at small x, there will be no scattering at short recoil times. Also, since $g(x)$ approaches unity at large x, the scattering rate should approach the Hohenberg and Platzman prediction at long recoil times. With this reasoning we deduce that the self energy to be used in Eq. 9 depends on the recoil distance; that is, $\Sigma_{p_f}(x) = -i\hbar/2\tau(x)$. The result for the final-state broadening then becomes

$$R_{FS}(Y,Q) \approx \frac{1}{2\pi} \int_{-\infty}^{\infty} dx \, \exp[iYx - |x|g(|x|)\Gamma], \tag{14}$$

where again $\Gamma = \frac{1}{2}\rho\sigma_{total}(Q)$. The Hohenberg and Platzman prediction, Eq. 11, can be obtained from Eq. 14 in the limit of a structureless fluid ($g(r) \rightarrow 1$). However, unlike Eq. 11, Eq. 14 satisfies the kinetic-energy sum rule because, in fact, $g(0) = 0$. Note also that the broadening given by Eq. 14 has the same FWHM as the broadening of Hohenberg and Platzmann, but it is negative at large $|Y|$ in order to satisfy the sum rule. A somewhat more accurate field-theoretic version of the new theory is discussed in the sidebar.

In the new theory the scaling variable, Y, acquires a new physical interpretation as the variable conjugate to the distance traveled by a recoiling atom. In the limit of a hard-core potential (that is, $\sigma_{total}(Q) = 2\pi r_0^2$, where r_0 is the hard-core radius and the factor of 2 accounts for forward diffractive scattering), final-state effects become a geometric problem that depends only on r_0 and the relative atomic positions given by $g(r)$. In quantum mechanics the position of an atom, in this case the recoil distance x, is not a variable distinct from its momentum. Rather, they are conjugate variables in the sense of a Fourier transform. Hence, introducing the relative positions of the atoms introduces no new variables into the problem, and Y-scaling will continue to hold even though the impulse approximation does not. Remember, all this assumes σ_{total} is independent of Q. If we now include the Q dependence of $\sigma_{total}(Q)$, we have indeed introduced a new variable and Y-scaling breaks down. However, since in practice $\sigma_{total}(Q)$ varies approximately as $\ln Q$, the corrections to Y-scaling in our theory are logarithmic, that is, slowly varying with Q.

Figure 11 shows the final-state broadening function predicted by the somewhat more sophisticated version of our theory presented in the sidebar. To make experimental predictions we must convolve $J_{IA}(Y)$ with the $R_{FS}(Y,Q)$ of Fig. 11 and with the $R_1(Y,Q)$ of Fig. 8. The results for the normal fluid and the superfluid are shown, together with the pulsed-source data, in Figs. 12a and 12b, respectively. The agreement between theory and experiment is now excellent for both the normal fluid and the superfluid! Experiment and theory both converge to a Bose-condensate fraction in the superfluid of 9.2 percent. We emphasize that the many-body calculations of momentum distributions and the theory for final-state effects were completed before the pulsed-source experiments were performed, so that in this case *ab initio* theory accurately predicted experiment. Further experiments by Sokol and collaborators show excellent agreement over the entire quantum-liquid region of the phase diagram of ^{4}He and at a variety of Q.

Implications of the Breakthrough

The convergence of theoretical and experimental routes to the momentum distributions of ^{4}He has finally confirmed London's fifty-year-old hypothesis connecting superfluidity with the existence of a Bose condensate—and may thus be considered a triumph for modern condensed-matter physics. The success of the momentum-distribution calculations provides confidence in the new supercomputer calculational methods for many-body quantum systems. The new theory of final-state effects points
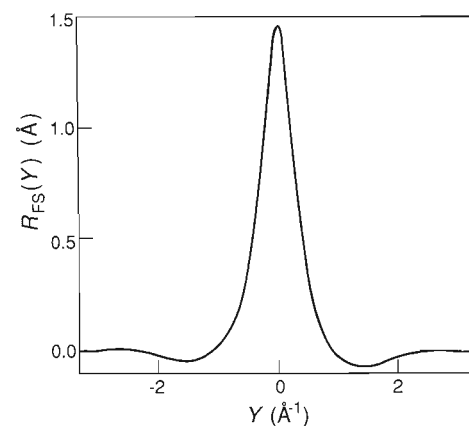
BROADENING FUNCTION PREDICTED BY NEW FSE THEORY

Fig. 11. Shown here is the broadening function, $R_{FS}(Y)$, predicted by the author's new theory of final-state effects (FSE). The full width at half maximum of $R_{FS}(Y)$, like that of Hohenberg and Platzman's final-state-broadening function, is given approximately by $\rho\sigma_{total}(Q)$. However, $\int_{-\infty}^{\infty} dY \, Y_{FS}^2(Y) = 0$, and thus, unlike Hohenberg and Platzman's final-state-broadening function, $R_{FS}(Y)$ satisfies the exact kinetic-energy sum rule on the neutron scattering law.

(a) Normal-Fluid ^{4}He

(b) Superfluid ^{4}He
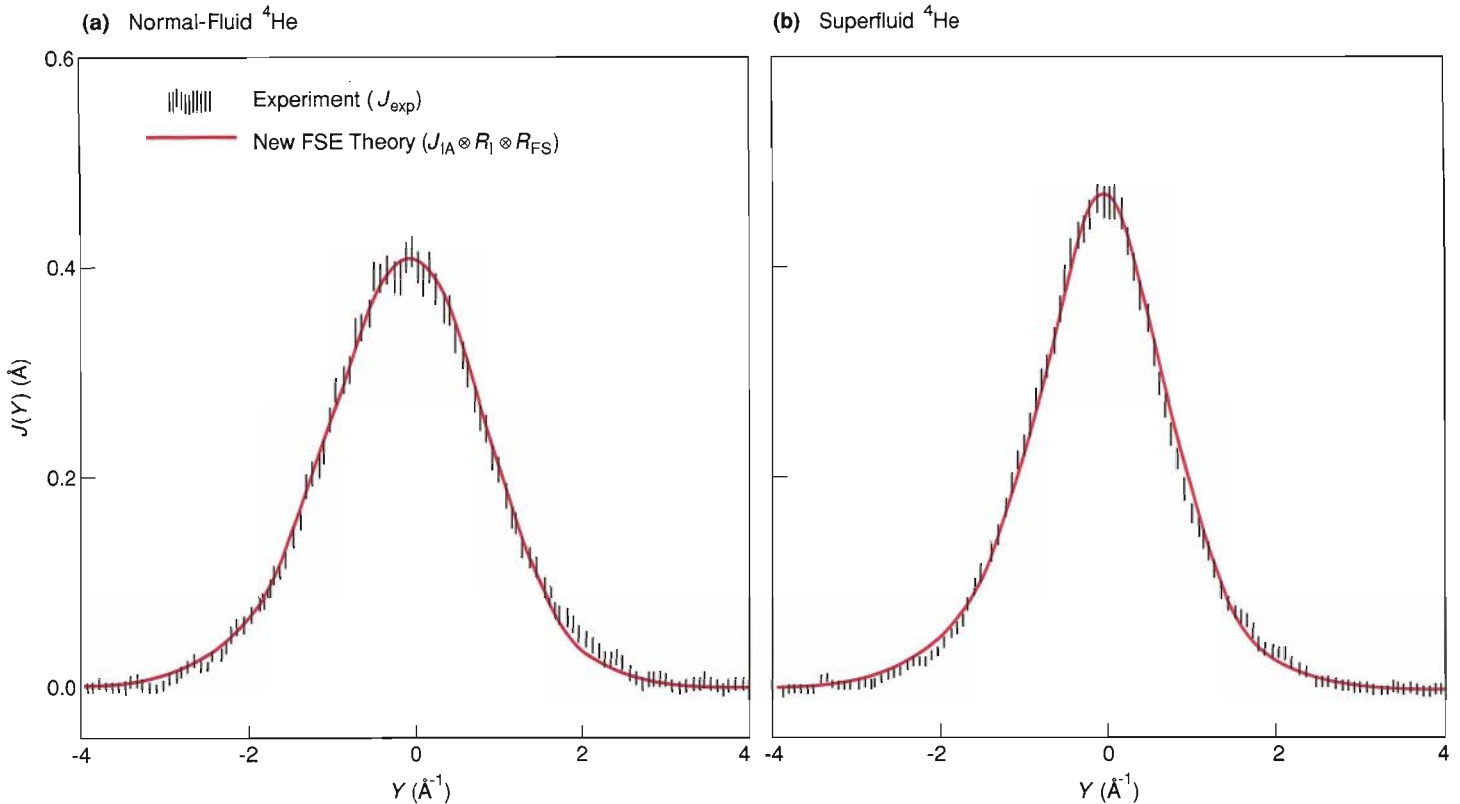
CONVERGENCE OF THEORY AND EXPERIMENT

Fig. 12. Convolution of the impulse approximation (Fig. 6b) with the broadening function given by the author's final-state-effects theory (Fig. 11) yields predicted Compton profiles that agree well with the data for both (a) normal-fluid ^{4}He and (b) superfluid ^{4}He. The agreement for normal-fluid ^{4}He arises because the sum rule satisfied by the new final-state broadening function ensures that final-state effects do not change the width of the normal fluid's essentially Gaussian momentum distribution. In the case of superfluid ^{4}He, both theory and experiment yield a Bose-condensate fraction of about 9.24 percent.

toward novel perturbative methods for calculating the dynamical response of strongly correlated systems (see sidebar). And the success of the experiments with the Phoenix spectrometer demonstrates the utility of high fluxes of epithermal neutrons in measuring quantities of fundamental scientific importance. However, the convergence does not mean that the condensate saga is finally at an end. Like many research breakthroughs, this one has important implications for future research.

If the *ab initio* calculations of momentum distributions in ^{4}He are correct, the data are consistent with no other theory for final-state effects except that of the author. Thus we believe that the new theory, or improved versions thereof, can be used with confidence to interpret and predict future experiments on momentum distributions in condensed-matter systems. The fact that the new theory predicts much smaller final-state effects than those predicted by Hohenberg and Platzman is encouraging. However, the theory also tells us that final-state broadening washes out sharp structure in the momentum distribution and that, because $\Delta Y \propto \sigma_{\text{total}}(Q)$ (see Eq. 12), the broadening decreases only slowly with increasing Q. Figure 13 shows theoretical predictions for $J(Y, Q)$ for increasing values of Q. Note that the broadening in the region near $Y = 0$ decreases, but only slowly, as Q increases from 30 to 270 Å^{-1}. Thus the theory predicts that the Bose-condensate fraction will not produce a sharp peak in $J(Y)$ in any feasible neutron-scattering experiment. This prediction should be an important test of the new theory.

To analyze reactor experiments, the theory will have to be augmented to account for the additional deviations from the impulse approximation that are observed at Q values less than 12 Å^{-1}. Among those deviations are the oscillations in the width of $J(Y, Q)$ with Q and the asymmetry about $Y = 0$. Both effects may be due to collective behavior in the condensed phase.

Perhaps the most important momentum-distribution experiment to attempt is observation of the Fermi-surface discontinuity in ^{3}He (that is, the discontinuity in $n(p)$ at $p = p_F$). Figure 14 shows the momentum distribution of a gas of non-interacting fermions with the same density as ^{3}He. Also shown are the momentum distribution for ^{3}He predicted by using HNC calculations as input to the impulse approxima-

tion and the "apparent" momentum distribution that would be inferred by analyzing final-state-broadened data as if the impulse approximation did not require corrections for final-state effects. In the absence of final-state effects, the discontinuity in $n(p)$ would cause a sharp change in slope in $J_{IA}(Y, Q)$ at $Y = p_F/\hbar$. Unfortunately, the new theory of final-state effects predicts that such a change in slope will not be directly apparent in the data. Nevertheless, experiments can distinguish the HNC theory from other theories that do not have a Fermi-surface discontinuity (such as the pairing theory of Lhuillier and Bouchaud). The HNC theory would predict a different Gaussian width for $J(Y, Q)$ than would other theories. Experiments on ^{3}He are particularly difficult because ^{3}He is a strong absorber of neutrons (and is therefore the primary component of thermal-neutron detectors!). Nevertheless, we hope that theory and experiment will resolve the question of the Fermi-surface discontinuity in ^{3}He in less time than the more than twenty years required to confirm the existence and size of the Bose condensate in ^{4}He.

At the beginning of this article, we mentioned that ^{3}He becomes a superfluid at 3 millikelvins. Does that mean a Bose condensate forms in ^{3}He despite the spin-statistics relation? The answer is no. Superfluidity in ^{3}He is caused by the formation of Cooper pairs of ^{3}He atoms and is thus somewhat analogous to superconductivity in metals, which is caused by formation of Cooper pairs of electrons. The difference is that the Cooper pairs of ^{3}He atoms are uncharged and form a relative p-wave bound state, whereas the Cooper pairs of electrons in metals are charged and usually form a relative s-wave bound state.

The present experimental and theoretical techniques for determining momentum
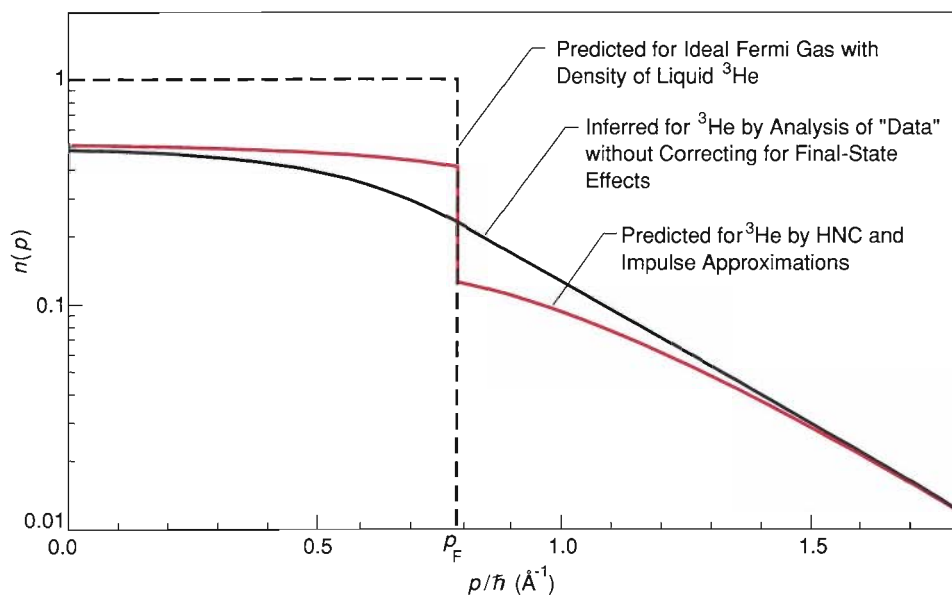
distributions in helium can also be applied to questions concerning the pressure and temperature dependence of the condensate fraction, the non-existence of a Bose condensate in solid ^{4}He, the non-existence of a Fermi surface in solid ^{3}He, the behavior of $n(\mathbf{p})$ at high $|\mathbf{p}|$ (which many-body theories predict is exponential rather than Gaussian even in the normal quantum liquid), the larger n_0 predicted for ^{4}He in a porous medium (which is expected to behave like a low-density Bose system), the complex momentum distributions expected for mixtures of ^{3}He and ^{4}He, the predicted absence of a Bose condensate in the two-dimensional ^{4}He systems produced by physisorption of ^{4}He on surfaces, and so on. All such experiments will benefit from the epithermal neutrons provided by pulsed neutron sources or by hot sources at reactors because they all require measurements at high Q values. Although conditions suitable for applying the impulse approximation may never be reached, the final-state corrections will be understood.
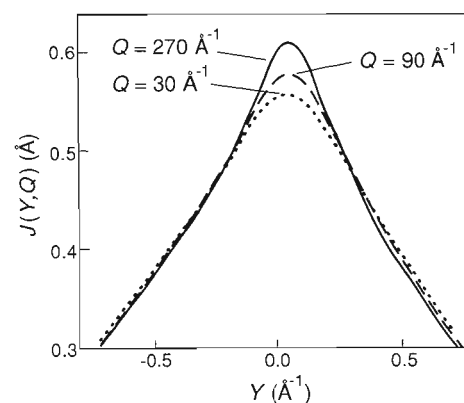
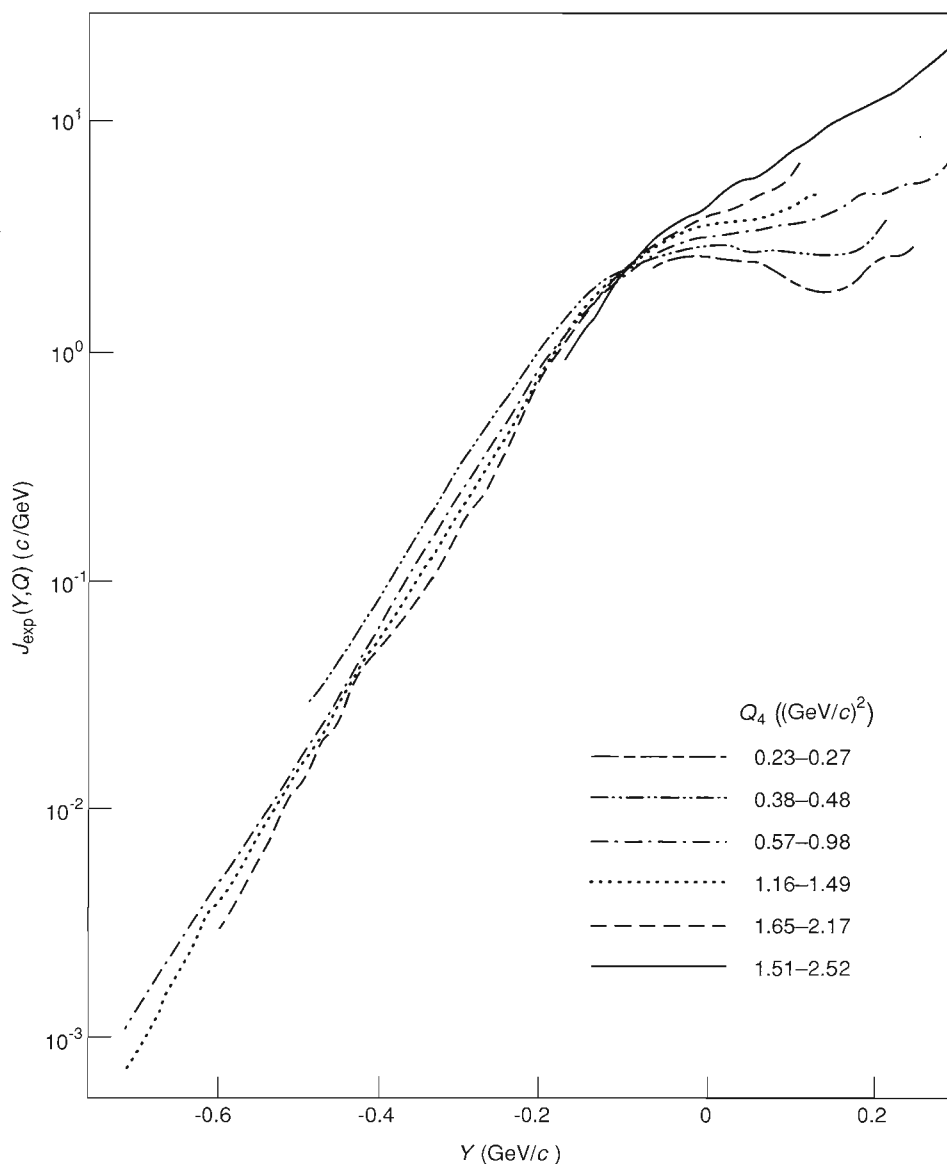## CAN THE BOSE CONDENSATE IN ^{4}He BE OBSERVED DIRECTLY?

Fig. 13. The author's theory of final-state effects predicts that the Compton profile of ^{4}He at absolute zero sharpens only slowly with increasing Q near $Y = 0$, the region relevant to the Bose condensate. That prediction implies that the Bose condensate will not produce a distinct peak in the Compton profile at any experimentally feasible Q.

CAN THE FERMI SURFACE IN ^{3}He BE OBSERVED DIRECTLY?

Fig. 14. The Pauli exclusion principle implies that, at absolute zero, the momentum distribution of an ideal Fermi gas is a step function with a discontinuity of 1 at the Fermi surface, that is, at the momentum of the highest filled momentum state, p_F. Shown here, as a dashed curve, is the absolute-zero momentum distribution of an ideal Fermi gas with the same density as ^{3}He, for which $p_F/\hbar = 0.789$ $Å^{-1}$. Also shown, as a red curve, is the absolute-zero momentum distribution predicted for real (interacting) ^{3}He by the HNC approximation. The discontinuity remains at 0.789 $Å^{-1}$, but its magnitude is reduced. Finally, the black curve is the "apparent" momentum distribution, which is inferred as follows. First, the impulse approximation prediction (at 30 $Å^{-1}$) obtained by using the red curve as input is convolved with a final-state broadening function appropriate to ^{3}He. The resulting final-state-broadened "data" are then analyzed by assuming the validity of the impulse approximation but not correcting for final-state effects. Note that the momentum distribution so inferred exhibits no discontinuity at p_F because, like the delta function at $k = 0$ in the momentum distribution of superfluid ^{4}He, it has been washed out by final-state effects.

Y-SCALING IN ELECTRON-NUCLEON SCATTERING

Fig. 15. Shown here are Compton profiles based on data for the quasi-elastic scattering of electrons from the ^{12}C nucleus at various relativistic four-momentum transfers. The Y variable has been generalized to include relativistic kinematics. The scattering follows Y-scaling for $Y < -0.1$ GeV/c, as evidenced by the fact that all the profiles fall on the same curve in that region. Y-scaling breaks down at greater Y values because of excitation of internal degrees of freedom of the nucleons making up the nucleus, such as the Δ(1238 MeV) excited state. If final-state effects are ignored, the impulse approximation suggests that the nucleon momentum distribution decreases exponentially with p over nearly four decades. A theoretical explanation for this almost universal behavior is the subject of current research. Determining whether the momentum distribution of helium quantum fluids behave similarly at large p would prove interesting, but the poorer signal-to-noise ratio in neutron-scattering experiments has so far prevented such an experiment.

Q_4 $((GeV/c)^2)$

– – – – –	0.23–0.27
–··–··–··	0.38–0.48
–·–·–·–··	0.57–0.98
············	1.16–1.49
– – – – –	1.65–2.17
———————	1.51–2.52

Y (GeV/c)

Going beyond helium, we are now ready to expand our knowledge of momentum distributions to a wide variety of many-body quantum systems. For example, quasi-elastic electron-nucleus scattering (QENS) at GeV energies is aimed at measuring the momentum distributions of nucleons in nuclei. Even though the energy scale characteristic of QENS differs by ten orders of magnitude from the energy scale of neutron scattering from helium, the two types of experiments share many common elements, including Y-scaling phenomena, the importance of final-state effects, and the methods for calculating properties of the many-particle wave functions. Figure 15 shows $J(Y, Q)$ for electron scattering from the ^{12}C nucleus. For $Y \leq 0$ the Compton profile exhibits a relativistic analogue of Y-scaling over nearly four orders of magnitude in $J(Y, Q)$. For $Y \geq 0$ excitation of internal degrees of freedom of the nucleons, such as the Δ(1238 MeV) resonance, destroys the Y-scaling. Note also the nearly exponential dependence of $J(Y, Q)$ on $|Y|$. We do not know whether the exponential dependence is a property of the underlying momentum distribution or is a manifestation of final-state effects. That question has been as important in nuclear physics as the existence of a Bose condensate in helium. Thus we plan to make the extension of our final-state theory to nuclei a high priority. ■

How final-state effects were *really* calculated

T he derivation of final-state broadening presented in the main text was phys-
ically intuitive but, like all heuristic arguments, involved a sleight of hand:
The classical-trajectory concept was not derived from first principles. In prac-
tice, the theory of final-state effects is a very difficult many-body problem. Conven-
tional perturbative expansion about the non-interacting ground state, a technique so
successful in calculating the properties of weakly interacting systems, is not a use-
ful approach here because helium atoms interact at short distances through a steeply
repulsive potential. However, alternatives to perturbation methods, such as the varia-
tional and Monte Carlo methods, are capable of handling strongly interacting systems
and thus have been most successful in calculating ground-state properties of helium,
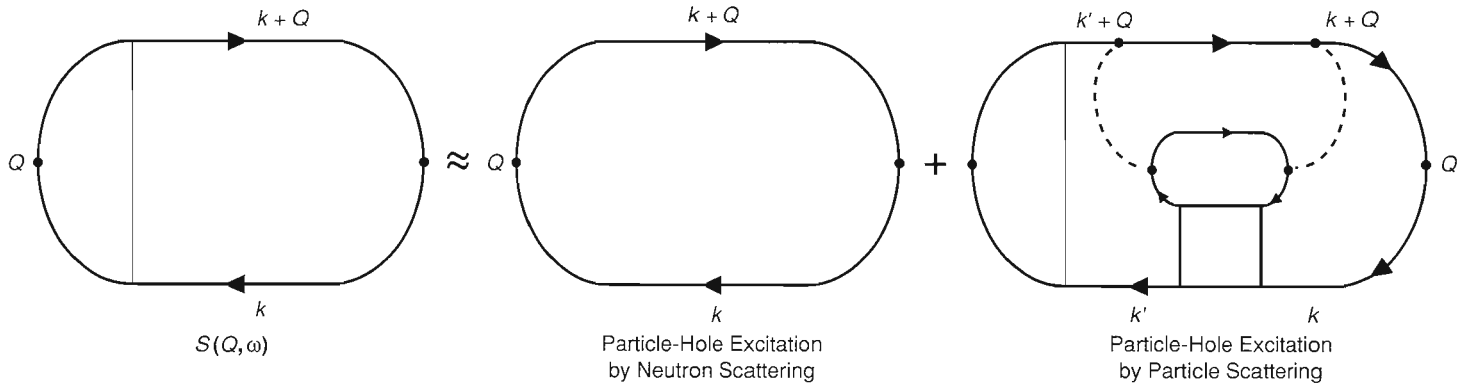including the momentum distribution $n(p)$ and the pair-correlation function $g(r)$.

In order to test the ground-state results against neutron-scattering experiments,
we need to calculate the dynamical response of the system to neutron scattering.
Since we have an obvious interest in not repeating the considerable work involved
in generating the ground-state results, we want to calculate the response by applying
perturbation theory to the variational and Monte Carlo results for the ground state.
However, conventional perturbation theory is again out of the question because the
dynamical response also involves helium-helium interactions.

Before we present our solution to this problem, let's outline the starting point.
We assume that neutron scattering at momentum transfer $\hbar Q$ introduces, at time zero,
a fluctuation about the ground state in the density of atoms with wave vector Q. By
calculating the amplitude of that density fluctuation at a later time t and taking its
Fourier transform, we can determine $S(Q,\omega)$, the observed scattering law. (Note
that ω is conjugate to t.) The density fluctuation is equal to a summation over all
so-called particle-hole excitations about the ground state, that is, over all processes
that add to the ground state an atom with wave vector $k + Q$ and remove from the
ground state an atom with wave vector k.

In the impulse approximation we assume that the particle-hole excitations prop-
agate freely without interacting with other atoms. Final-state effects, on the other
hand, are due to interaction of the excitations with other atoms. Scattering of a par-
ticle and a hole creates more particle-hole excitations about the ground state. Al-
though in principle an infinity of multiple scatterings of a particle-hole pair can oc-
cur, the correlations in the ground-state wave function imply that only single addi-
tional particle-hole excitations need be considered. In effect, the correlations screen
the steeply repulsive core interaction at short distances, rendering that interaction fi-
nite. After all, to minimize their energy in the ground state, the atoms tend to sit in
the attractive part of the potential, far away from its steeply repulsive core. Thus the
effective final-state interactions can be characterized by a small parameter, and per-
turbation theory can be used for systematic, controlled calculations.

The divergent terms in the perturbative expansion of $S(Q,\omega)$ involve all pro-
cesses that transform a $(k + Q, k)$ particle-hole pair to a $(k' + Q, k')$ pair. To obtain
finite results, those divergent terms must be explicitly resummed to all orders in the
perturbation expansion. In practice, the summation is accomplished by defining a
"projection superoperator," which acts in the Hilbert space of $(k + Q, k)$ particle-hole
excitations about the ground state much as ordinary operators act in the Hilbert space

Neutron Scattering Law ≈ Impulse Approximation + New Final-State-Effects Theory

$S(Q, \omega)$

Particle-Hole Excitation
by Neutron Scattering

Particle-Hole Excitation
by Particle Scattering

NEW THEORY OF FINAL-STATE EFFECTS

The author approximates the neutron scattering law for helium as the sum of the impulse approximation and one additional scattering that accounts for final-state effects. Shown here are Feynman diagrams for that approximation. The Feynman diagram for the neutron scattering law represents the propagation of a particle-hole excitation that removes a particle of wave vector k from the ground state and adds to the ground state a particle of wave vector $k + Q$. Arrows denote the direction of momentum flow. Arrows pointing right denote particle lines; arrows pointing left denote hole lines. Only the particle lines carry high momentum. The hatched area denotes the exact result for $S(Q, \omega)$ including all scatterings of particles and holes. The Feynman diagram for the impulse approximation indicates that both particles and holes propagate without scattering. The Feynman diagram for the final-state effects indicates that each particle scatters from another atom and creates a new particle-hole excitation. (Further scatterings are possible but not included in the approximation.) The shaded square is the two-particle density matrix describing the correlations between the two holes in the ground state created by the two particle-hole excitations. The hole-hole correlations are related by sum rules to the pair-correlation function of the ground state. The dashed lines represent the two-particle t-matrix that describes particle scattering. Because the hatched area appears in the Feynman diagrams for both the neutron scattering law and the final-state effects, the scatterings that transform a $(k + Q, k)$ excitation to a $(k' + Q, k')$ excitation must be calculated self-consistently.

of quantum-mechanical states. The neutron scattering law then equals the expectation value of the projection superoperator, and calculations analogous to ordinary perturbation theory can be carried out in the superoperator Hilbert space. The effective interaction is the two-atom scattering matrix multiplied by a ground-state correlation function, which acts to screen the short-distance pathologies of the potential. Additional restrictions on the important scattering processes are obtained by noting that all k entering a two-particle density matrix must be characteristic of the ground-state wave function, as given by the momentum distribution, and that Q is much larger than those characteristic values.

After the above procedure is implemented, the neutron scattering law can be expressed as the sum of the impulse approximation and one additional scattering process. The accompanying figure shows the Feynman diagrams for the components of the sum. In the Feynman diagram for the one additional scattering process, the dashed line represents the t-matrix describing the scattering of two particles and the square represents the two-particle density matrix for the ground state. The latter matrix is a generalization of the correlation functions, such as $g(r)$ and $n(p)$, that characterize the ground-state wave function.

If we approximate the density matrix in terms of $g(r)$ and $n(p)$ in a way that satisfies sum rules and, since Q is large, use a semiclassical approximation for the t-matrix, then the final "Dyson" equation can be solved analytically. The result for the final-state broadening, $R(Y, Q)$, is given by
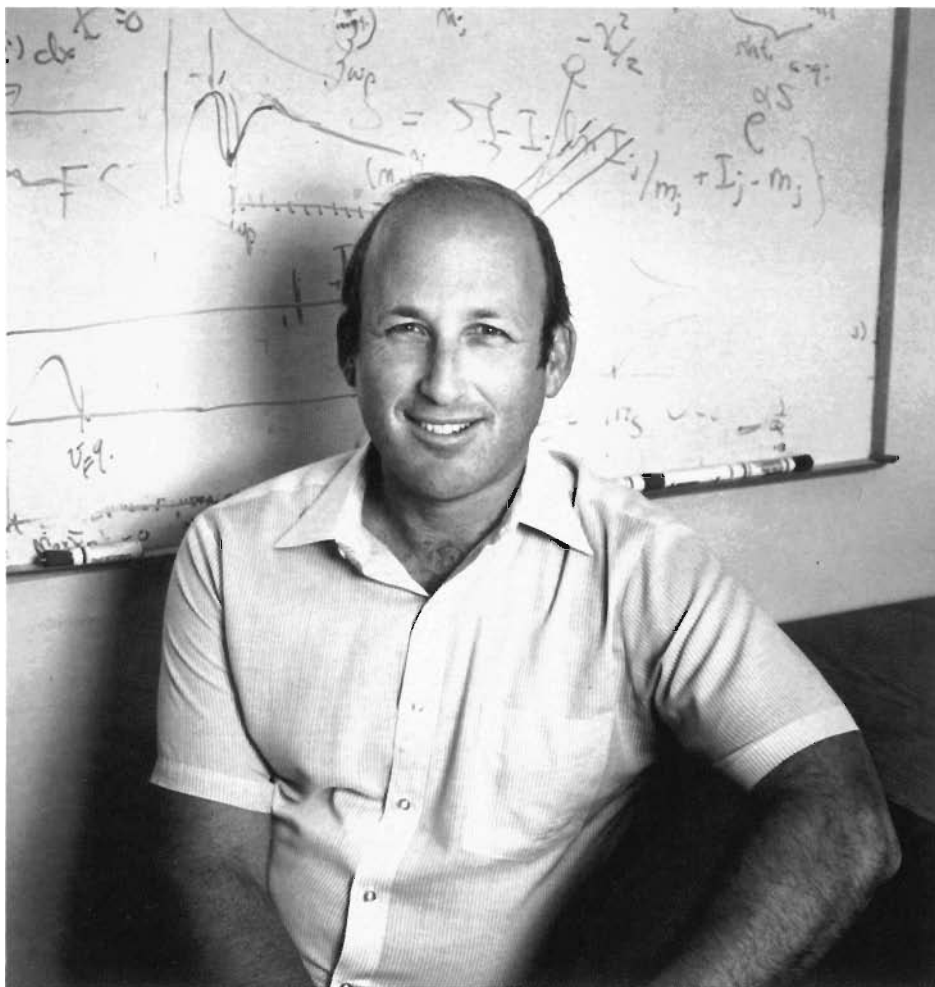
$$R(Y, Q) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dx \, \exp\left[iYx - \int_0^{|x|} \Gamma(x')dx' \right],$$

where

$$\Gamma(x) = \rho\pi \int_0^{\infty} db^2 \left(e^{2i\delta_b} - 1 \right) g\left(\sqrt{x^2 + b^2} \right).$$

The phase shift δ_b is the semiclassical value for scattering at impact parameter b.

The above expression for $R(Y, Q)$, which is somewhat more complicated than Eq. 14 in the main text, is the expression we have plotted in Fig. 11 of the main text and used in comparing theory with experiment. It is essentially the same as the familiar Wentzel-Kramers-Brillouin (WKB) classical-trajectory approximation taught in elementary quantum mechanics except that the potential, $V(x)$, is replaced by an "optical potential," $\hbar^2 Q\Gamma(x)/m$ that accounts for all repeated scatterings from the same helium atom. The quantity $\hbar^2 Q\Gamma(\infty)/m$ is simply the forward scattering t-matrix for the scattering of two helium atoms. The approach taken here is required for helium, a strong scatterer, but it is satisfying that the result reduces to the WKB approximation in the limit of a weak scatterer. ∎

Richard N. Silver received a B.S. in physics (with honors) from the California Institute of Technology in 1966 and a Ph.D. in theoretical physics with a minor in applied mathematics from the same institution in 1971. Then followed a one-year postdoctoral fellowship at Brown University in particle physics and a two-year IBM Postdoctoral Fellowship in Applied Physics at the California Institute of Technology. He joined the Laboratory's Laser Theory Group in 1974, became a founding member of the Condensed Matter and Statistical Physics Group, T-11, in 1977, served as Leader of the Neutron Scattering Group from 1979 to 1986, and since 1986 has held a joint appointment in T-11 and the Los Alamos Neutron Scattering Center. He has organized several conferences and served as a member of several national committees, including the LAMPF Program Advisory Committee and the IPNS Program Advisory Committee. He served as chairman of the Laboratory's Postdoctoral Committee from 1986 to 1989. The results of his research are documented in more than eighty publications on particle physics, semiconductor physics, optical properties of solids, many-body theory, neutron scattering, spectrometer design, maximum entropy and Bayesian methods of statistical inference, and quantum Monte Carlo.

Further Reading

David R. Tilley and John Tilley. 1986. *Superfluidity and Superconductivity*, second edition. Bristol, England: Adam Hilger Ltd. A general introduction to the properties of liquid helium.

Henry R. Glyde and Eric C. Svensson. 1987. Solid and liquid helium. In *Neutron Scattering B*, edited by D. L. Price and K. Skold, 303. Methods of Experimental Physics, volume 23. San Diego: Academic Press, Inc.

Richard N. Silver and Paul E. Sokol, editors. 1989. *Momentum Distributions*. New York: Plenum Publishing Co. A survey of work on momentum distributions in quantum fluids, condensed matter, and nuclei.

R. N. Silver and P. E. Sokol. To be published. Bose condensate in superfluid ^{4}He and momentum distributions by deep inelastic scattering. In *Recent Progress in Many-Body Theories II*, edited by Y. Avishai. New York: Plenum Publishing Co.

Acknowledgments

How final-state effects were *really* calculated

The derivation of final-state broadening presented in the main text was physically intuitive but, like all heuristic arguments, involved a sleight of hand: The classical-trajectory concept was not derived from first principles. In practice. the theory of final-state effects is a very difficult many-body problem. Conventional perturbative expansion about the non-interacting ground state. a technique so successful in calculating the properties of' weakly interacting systems, is not a useful approach here because helium atoms interact at short distances through a steeply repulsive potential. However. alternatives to perturbation methods, such as the variational and Monte Carlo methods. are capable of handling strongly interacting systems and thus have been most successful in calculating ground-state properties of helium, including the momentum distribution $n(p)$ and the pair-correlation function $g(r)$.
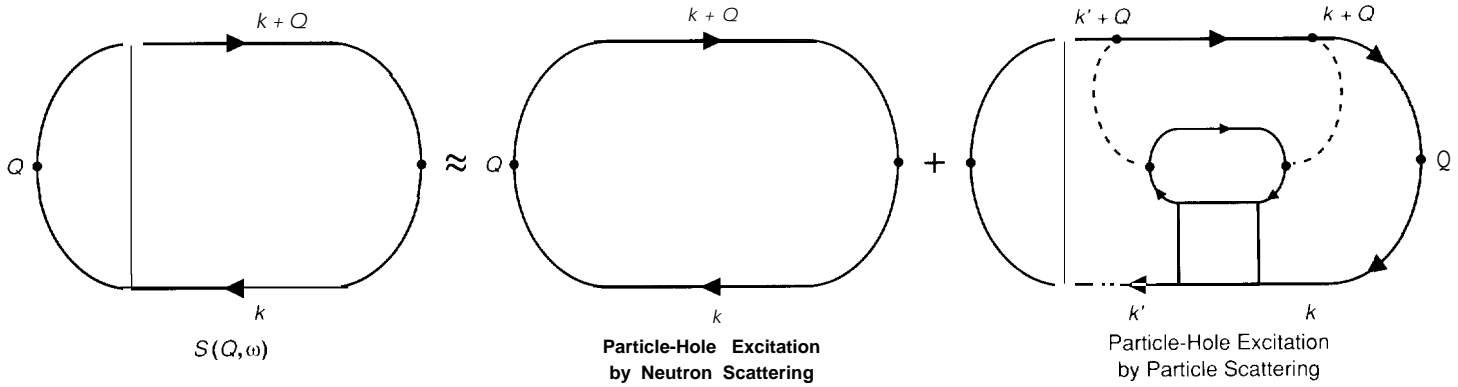
In order to test the ground-state results against neutron-scattering experiments, we need to calculate the dynamical response of the system to neutron scattering. Since we have an obvious interest in not repeating the considerable work involved in generating the ground-state results, we want to calculate the response by applying perturbation theory to the variational and Monte Carlo results for the ground state. However, conventional perturbation theory is again out of the question because the dynamical response also involves helium-helium interactions.

Before we present our solution to this problem, let's outline the starting point. We assume that neutron scattering at momentum transfer $h Q$ introduces, at time zero, a fluctuation about the ground state in the density of atoms with wave vector Q. By calculating the amplitude of that density fluctuation at a later time t and taking its Fourier transform, we can determine $S(Q, w)$, the observed scattering law. (Note that w is conjugate to t.) The density fluctuation is equal to a summation over all so-called particle-hole excitations about the ground state, that is, over all processes that add to the ground state an atom with wave vector $k + Q$ and remove from the ground state an atom with wave vector k.

In the impulse approximation we assume that the particle-hole excitations propagate freely without interacting with other atoms. Final-state effects, on the other hand, are due to interaction of the excitations with other atoms. Scattering of a particle and a hole creates more particle-hole excitations about the ground state. Although in principle an infinity of multiple scattering of a particle-hole pair can occur, the correlations in the ground-state wave function imply that only single additional particle-hole excitations need be considered. In effect, the correlations screen the steeply repulsive core interaction at short distances, rendering that interaction finite. After all, to minimize their energy in the ground state, the atoms tend to sit in the attractive part of the potential, far away from its steeply repulsive core. Thus the effective final-state interactions can be characterized by a small parameter. and perturbation theory can be used for systematic, controlled calculations.

The divergent terms in the perturbative expansion of $S(Q, w)$ involve all processes that transform a $(k + Q, k)$ particle-hole pair to a $(k' + Q, k')$ pair. To obtain finite results, those divergent terms must be explicitly resummed to all orders in the perturbation expansion. In practice. the summation is accomplished by defining a "projection superoperator." which acts in the Hilbert space of $(k + Q, k)$ particle-hole excitations about the ground state much as ordinary operators act in the Hilbert space

Neutron Scattering Law = Impulse Approximation + New Final-State-Effects Theory

$S(Q, \omega)$

Particle-Hole Excitation by Neutron Scattering

Particle-Hole Excitation by Particle Scattering

NEW THEORY OF FINAL-STATE EFFECTS

The author approximates the neutron scattering law for helium as the sum of the impulse approximation and one additional scattering that accounts for final-state effects. Shown here are Feynman diagrams for that approximation. The Feynman diagram for the neutron scattering law represents the propagation of a particle-hole excitation that removes a particle of wave vector k from the ground state and adds to the ground state a particle of wave vector $k + Q$. Arrows denote the direction of momentum flow. Arrows pointing right denote particle lines; arrows pointing left denote hole lines. Only the particle lines carry high momentum. The hatched area denotes the exact result for $S(Q, w)$ including all scattering of particles and holes. The Feynman diagram for the impulse approximation indicates that both particles and holes propagate without scattering. The Feynman diagram. for the final-state effects indicates that each particle scatters from another atom and creates a new particle-hole excitation. (Further scattering are possible but not included in the approximation.) The shaded square is the two-particle density matrix describing the correlations between the two holes in the ground state created by the two particle-hole excitations. The hole-hole correlations are related by sum rules to the pair-correlation function of the ground state. The dashed lines represent the two-particle t-matrix that describes particle scattering. Because the hatched area appears in the Feynman diagrams for both the neutron scattering law and the final-state effects, the scattering that transform a $(k + Q, k)$ excitation to a $(k' + Q, k')$ excitation must be calculated self-consistently.

of quantum-mechanical states. The neutron scattering law then equals the expectation value of the projection superoperator, and calculations analogous to ordinary perturbation theory can be carried out in the superoperator Hilbert space. The effective interaction is the two-atom scattering matrix multiplied by a ground-state correlation function, which acts to screen the short-distance pathologies of the potential. Additional restrictions on the important scattering processes are obtained by noting that all k entering a two-particle density matrix must be characteristic of the ground-state wave function, as given by the momentum distribution, and that Q is much larger than those characteristic values.

After the above procedure is implemented, the neutron scattering law can be expressed as the sum of the impulse approximation and one additional scattering process. The accompanying figure shows the Feynman diagrams for the components of the sum. In the Feynman diagram for the one additional scattering process, the dashed line represents the t-matrix describing the scattering of two particles and the square represents the two-particle density matrix for the ground state. The latter matrix is a generalization of the correlation functions, such as g(r) and n(p), that characterize the ground-state wave function.

If we approximate the density matrix in terms of $g(r)$ and $n(p)$ in a way that satisfies sum rules and, since Q is large, use a semiclassical approximation for the t-matrix, then the final "Dyson" equation can be solved analytically. The result for the final-state broadening, $R(Y, Q)$, is given by

$$R(Y, Q) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dx \, \exp \left[iYx - \int_0^{|x|} \Gamma(x')dx' \right],$$

where

$$\Gamma(x) = \rho \pi \int_0^{\infty} db^2 \left(e^{2i\delta_b} - 1 \right) g \left(\sqrt{x^2 + b^2} \right).$$

The phase shift δ_b is the semiclassical value for scattering at impact parameter b.

The above expression for $R(Y, Q)$, which is somewhat more complicated than Eq. 14 in the main text, is the expression we have plotted in Fig. 11 of the main text and used in comparing theory with experiment. It is essentially the same as the familiar Wentzel-Kramers-Brillouin (WKB) classical-trajectory approximation taught in elementary quantum mechanics except that the potential, V (x), is replaced by an "optical potential," $\hbar^2 Q \Gamma(x)/m$ that accounts for all repeated scatterings from the same helium atom. The quantity $\hbar^2 Q \Gamma(\infty)/m$ is simply the forward scattering t-matrix for the scattering of two helium atoms. The approach taken here is required for helium, a strong scatterer, but it is satisfying that the result reduces to the WKB approximation in the limit of a weak scatterer. ∎

Bayesian Inductive Inference Maximum Entropy & Neutron Scattering

by *Devinder Singh Sivia*

*T*he result of this experiment was inconclusive, so we had to use statistics. Such oft-heard statements reflect the "cookbook" approach to statistics that we are taught as undergraduates. Not satisfied with the maze of seemingly ad hoc statistical tests, many of us become inclined to avoid the subject as much as possible.

Fortunately, statistics does not have to be like that! A more logical and unified approach
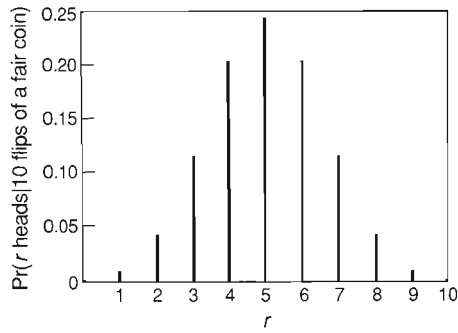
DIRECT PROBABILITIES

Fig. 1. What is the probability of getting r heads in 10 flips of a fair coin, Pr(r heads|10 flips of a fair coin)? Deductive logic tells us that the probability in question is given by $N_{r\ heads}/N$, where N is the number of all possible sequences of heads and tails that the 10 flips can generate and $N_{r\ heads}$ is the number of those sequences that contain r heads (in any order). To obtain a numerical answer, note that $N = 2^{10}$ and that $N_{r\ heads} = 10!/(10 - r)!\,r!$. Thus Pr($r$ heads|10 flips of a fair coin) = $[10!/(10 - r)!\,r!]/2^{10}$.

to the whole subject is provided by the probability formulations of Bayes and Laplace. Bayes' ideas (published in 1763) were used very successfully by Laplace (1812) but were then allegedly discredited and largely forgotten until they were rediscovered by Jeffreys (1939). In more recent times they have been expounded by Jaynes and others. Here we present an introductory glimpse of the Bayesian approach. We then illustrate how Bayesian ideas, and developments such as the maximum entropy method, are affecting data analysis and thoughts on instrument design at the Manuel Lujan, Jr. Neutron Scattering Center (LANSCE).

Everyday games of chance are governed by deductive logic. For example, if we are told that a fair coin is flipped ten times, we can deduce accurately the chances that all ten flips produced heads, or that nine produced heads and one produced tails, ..., or that all ten flips produced tails (Fig. 1). Turning to neutron scattering, let's suppose we know the scattering law for a particular sample and the geometry of the diffractometer, the efficiencies of the detectors, and so on. Then we can predict the chances of observing a certain number of neutron counts in any given detector. These examples are in the realm of deductive logic, or pure mathematics: Given the rules of a "game," we can predict the chances of various outcomes.

Most scientists, however, are concerned with the more difficult inverse problem. Given that a coin of unknown origin was tossed ten times and the result was seven heads, was it a fair coin or a weighted one? Further, what is the best estimate of the bias-weighting of the coin and what is the confidence in the prediction? If we are now given more data on the coin, how should we incorporate the new information and how do our prediction and confidence level change? This type of problem is in the realm of inductive logic, plausible reasoning, or inference: Having seen the outcome of several "moves" in a game, we want to infer the rules governing that game. Returning to neutron scattering, let's suppose we have recorded so many neutron counts in various detectors and wish to infer the scattering law for the sample. Like all problems in inductive logic, this problem has no clear-cut answer. The most we can hope to do is make the "best" inference based on both the experimental evidence and any prior knowledge we have at hand, reserving the right to revise our position if new information comes to light. Around 500 B.C. Herodotus said much the same thing: "A decision was wise, even though it led to disastrous consequences, if the evidence at hand indicated it was the best one to make; and a decision was foolish, even though it led to the happiest possible consequences, if it was unreasonable to expect those consequences."

Bayes' Theorem

Bayes' theorem, which was actually written down in its present-day form by Laplace and not Bayes, is the cornerstone of scientific inference. It provides the bridge between the inductive logic we require and the deductive logic we know how to use. Its status is somewhat akin to the position of Newton's second law of motion in mechanics: seemingly tame and innocuous, but powerful enough to analyze a wide variety of problems when the relevant details and assumptions are given. In mechanics we may be taught that $s = \frac{1}{2}gt^2$ is the relationship between the vertical distance s that a body falls under a gravitational field g after a time t when released from rest at $t = 0$. We may also be told that the speed of sound v through a gas with pressure P and density ρ is given by $v^2 = P/\rho$. Although these two formulae look quite different and apply to different situations, it is satisfying to know that both of them are derived from the same physical law: Force is equal to the rate of change of momentum. Similarly the Bayesian approach to probability and statistics provides the logical foundation for the conventional teaching of statistics we are given as undergraduates. A Bayesian analysis often leads us to use the same procedure as advocated by the "cookbook" school of statistics, but it forces us to state clearly the

assumptions (usually forgotten) made in going from the fundamental rule for inductive inference (Bayes' theorem) to the particular statistical prescription we use.

But what is Bayes' theorem? Simply stated, it says that the conditional probability of A (being true) given B (is true), written as $\Pr(A|B)$, is proportional to the conditional probability of B given A times the probability of A:

$$\Pr(A|B) \propto \Pr(B|A) \times \Pr(A). \tag{1}$$

Bayes' theorem is easy to prove for problems in which A and B are "macroscopic" events that can be realized in a large number of equally probable "microscopic" ways. In such problems the probability of an event is the number of ways in which the event can occur divided by the total number of possibilities. For example, suppose the space of "microscopic" possibilities is all the possible sequences of heads and tails that can occur if a fair coin is flipped ten times. Since the coin is fair, each of the possible sequences is equally probable. "Macroscopic" event A might then be the event that the total number of heads was less than four, and B might be the event that a head was obtained on the third and seventh tosses. Figure 2 shows a schematic representation of the space of all "microscopic" possibilities and the portions of that space occupied by realizations of event A and event B. Now, let N be the total number of possibilities, N_A be the number of possibilities resulting in event A, N_B be the number of possiblities resulting in event B, and N_{AB} be the number of possibilities resulting in both event A and event B. Then the probabilities of the various outcomes of interest become

$$\Pr(A) = N_A/N, \quad \Pr(B) = N_B/N, \quad \Pr(A|B) = N_{AB}/N_B, \quad \text{and} \quad \Pr(B|A) = N_{AB}/N_A.$$

We can then write the probability of both A and B occurring, $\Pr(A,B)$, in two different ways:

$$\Pr(A,B) = N_{AB}/N = \Pr(A|B) \times \Pr(B) = \Pr(B|A) \times \Pr(A). \tag{2}$$

Bayes' theorem, as stated in Eq. 1, follows immediately from the two expressions for $\Pr(A,B)$ in Eq. 2, provided we associate $1/\Pr(B)$ in Eq. 2 with the proportionality constant in Eq. 1.

Although this proof is simple, the full implications of Bayes' theorem do not become apparent until we discover that the theorem applies equally well to cases in which A and B are any arbitrary propositions and the probabilities assigned to them represent merely our belief in the truths (or otherwise) of the propositions. This remarkable generalization, which is certainly not obvious, was proved by Cox (1946) while he was considering the rules necessary for logical and consistent reasoning.

Suppose we have a set of propositions. For example, a: It will rain tomorrow; b: King Harold died by being hit in the eye with an arrow during the battle of Hastings in 1066 A.D.; c: This is a fair coin; d: This coin is twice as likely to come up heads as tails; and so on. The minimum requirement for expressing our relative beliefs in the truth of the various propositions in a consistent fashion is that we rank them in a transitive manner. That is to say, if we believe proposition a more than b and b more than c, then we necessarily believe a more than c. Such a transitive ranking can easily be obtained by assigning a real number to each of the propositions in a manner so that the larger the numerical value associated with a proposition, the more we believe it. Cox went on to put forward two more axioms for logical, consistent reasoning: (1) If we first specify our degree of belief that A is true and then specify how much we believe B is true given that A is true, then we have implicitly defined our degree of belief for both A and B being true; and (2) If we specify how much we believe that A is true, then we have implicitly specified how much we believe that A is false. Cox showed that if we accept these remarkably mild desiderata, then
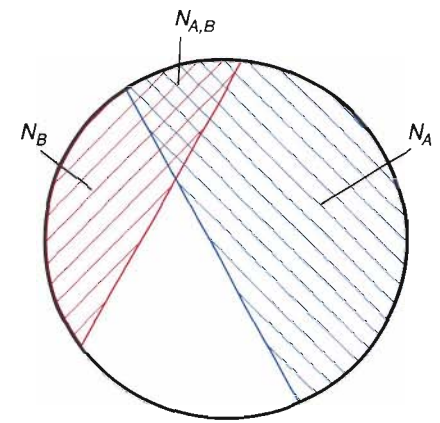

SAMPLE SPACE AND PROBABILITIES

Fig. 2. The sample space occupied by all N equally probable microscopic possibilities is depicted schematically here as a circle of area N. The microscopic possibilities result in various macroscopic events, such as A and B. The number of possibilities that result in A and the number of possibilities that result in B are represented by portions of the circle with areas N_A and N_B (hatched regions). The probability of A, $\Pr(A)$, is given by the fraction N_A/N; similarly, $\Pr(B)$ is given by the fraction N_B/N. The probability of A and B, $\Pr(A, B)$, is given by $N_{A,B}/N$, where $N_{A,B}$, represented as an area of overlap between N_A and N_B, is the number of possibilities that result in both A and B.

INDIRECT PROBABILITIES:
THE BIAS WEIGHTING OF A COIN

Fig. 3. (a) Shown here are three of the
prior probability distributions that might
be assigned to *H*, the bias weighting of
a coin: the "ignorant" (or uniform) prior,
which reflects the belief that all values of *H*
(0 ≤ *H* ≤ 1) are equally probable; a "fair"
prior, which reflects a belief that the coin is
likely to have both a head and a tail and to be
unbiased, or, in other words, a belief that the
most likely value of *H* is 0.5; and a "crooked"
prior, which reflects a belief that the coin is
likely to be double-headed or double-tailed,
or a belief that the most likely values of *H*
are 1 or 0. The series of graphs in (b) shows
how the posterior probability distributions
corresponding to the priors in (a) evolve
as the number of data increases. The data
were generated by using a random-number
generator in a computer and a value of 0.2 for
the bias weighting. Note that, as the number
of data increases, all the posteriors converge
to a delta function centered at *H* = 0.2. In
other words, as the experimental evidence
increases, the assumptions embodied in the
priors have less effect on our estimate of *H*.

(a) The Priors

(b) Evolution of Posteriors
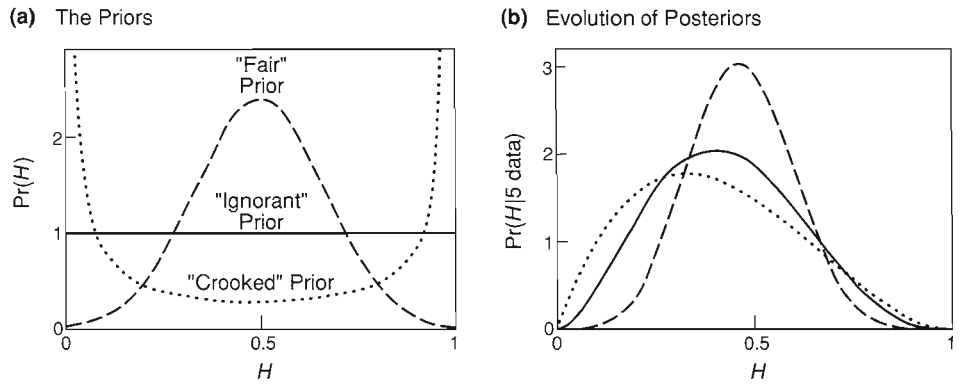
there must be a mapping that transforms the real numbers we have associated with
the various propositions (to express our beliefs in them) to another set of positive real
numbers that obeys the usual rules of probability theory:

$$\Pr(A, B) = \Pr(A|B) \times \Pr(B) \text{ and } \Pr(A) + \Pr(\bar{A}) = 1,$$

where $\bar{A}$ represents the proposition that A is false. In other words, any method of
logical and consistent reasoning (no matter what the context) must be equivalent to
the use of ordinary probability theory, where the probabilities represent our beliefs or
state of knowledge about various propositions or hypotheses in the Bayes-Laplace-
Jeffreys sense.

Bayes' theorem itself is just a simple corollary of these rules, but what does
it really mean and why is it so powerful? Let us return to the coin-flipping prob-
lem as a concrete but simple example. Again we are told that a coin was flipped n
times and came up heads r times, but we don't know whether the coin was fair. Our
problem is to *infer* the coin's bias-weighting for heads, call it H. We will say that
$H = 0$ represents a double-tailed coin (that is, a coin such that a head never appears),
$H = 0.5$ represents a fair coin (that is, a coin such that its head is likely to come up
as often as its tail), $H = 1$ represents a double-headed coin, and all other values of H
(between 0 and 1) correspond to some intermediate bias-weighting.

To carry out the inference, we need to specify our beliefs in the set of proposi-
tions that, given the data, the value of H lies in a narrow range between h and $h + \delta h$,
where h can take on values between 0 and 1. In terms of a probability distribution
for H, $\Pr(H = h|\{\text{data}\})$, or simply $\Pr(H|\{\text{data}\})$, we write

$$\lim_{\delta h \to 0} \Pr(h \leq H \leq h + \delta h|\{\text{data}\}) = \Pr(H|\{\text{data}\})dh.$$

Thus $\Pr(H|\{\text{data}\})$, known as the posterior probability distribution (or simply the
posterior), represents our state of knowledge about the bias-weighting for heads in
light of the data. The value of h at which the posterior is a maximum gives our best
estimate of the bias-weighting, and the spread of the posterior about the maximum
gives our confidence in that estimate. If the posterior is sharply peaked, we are sure
about our estimate; if it is broad, we are fairly uncertain about the true value of H.

In order to determine $\Pr(H|\{\text{data}\})$, we need to use Bayes' theorem,

$$\Pr(H|\{\text{data}\}) \propto \Pr(\{\text{data}\}|H) \times \Pr(H),$$

which relates the posterior to two other probability distributions, one of which can be
"calculated" from the data and the other "guessed."

The probability distribution $\Pr(H = h)$, or simply $\Pr(H)$, which also is defined
for $0 \leq h \leq 1$, represents our state of knowledge about the value of H *before* we
are given the data. It is thus called the prior probability distribution (or simply the
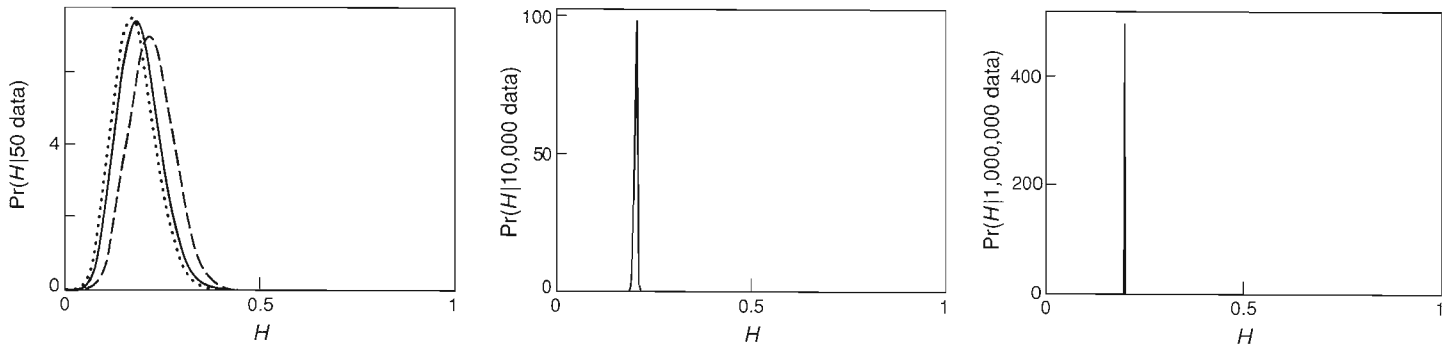
prior). In the coin-tossing problem, if we are completely ignorant about the coin, we would assign a uniform prior, $Pr(H) = \text{constant} = 1$ for all values of h between 0 and 1, to indicate that a priori all possible values of H are equally probable. If we do have other prior information, perhaps the results of previous data, then this information should be reflected in the nonuniform character of $Pr(H)$. (Actually our statement of Bayes' theorem should read $Pr(H | \{\text{data}\}, I) \propto Pr(\{\text{data}\} | H, I) \times Pr(H | I)$, where I represents other prior information or prior assumptions.) Figure 3a shows three possible assignments for $Pr(H)$, each reflecting a different assumption about the coin: the uniform, or ignorant, prior; a prior that assumes the coin is most likely to be double-headed or double-tailed; and a prior that assumes the coin has a head and a tail and is probably fair.

Having specified our prior, we need now to consider the other probability distribution in Bayes' theorem, $Pr(\{\text{data}\} | H)$, which reflects the nature of the "experiment." This probability distribution can be computed because it involves deductive logic. It is called the *likelihood* function because it tells us how likely it is that we would have obtained the data that we did if we had been given the value of H. For our problem we are told that a coin was flipped n times and came up heads r times. If we assume that the data are independent (that is, the outcome of one flip did not affect the result of another) and that the bias-weighting is H, then the likelihood function is simply a binomial distribution:

$$Pr(\{\text{data}\} | H) = {}^nC_r \times H^r \times (1 - H)^{n-r},$$

where ${}^nC_r = n!/r!(n - r)!$ is the number of ways of picking r objects (independent of order) from a choice of n. (Figure 1 shows such a binomial distribution.)

Multiplying $Pr(H)$ and $Pr(\{\text{data}\} | H)$, we obtain the posterior $Pr(H | \{\text{data}\})$, which summarizes all that we can infer about the value of H given the data. Figure 3b shows how the posterior for each of the three priors in Fig. 3a changes as we are given more and more data. The data in this example were generated by using a random-number generator in a computer and setting H to 0.2. We find that as we obtain more data, we become more confident in our prediction for the inferred value of H (that is, the width of each posterior decreases) and our prior state of knowledge, as expressed in $Pr(H)$, becomes less important (that is, no matter what our prior assumptions were, the posteriors converge to the same answer when enough data are available).

The power of Bayes' theorem is that it effectively provides the only consistent bridge between the inductive logic (or indirect probabilities) required for scientific inference and the deductive logic (or direct probabilities) that we know how to use. Generalizing, we see that Bayes' theorem encapsulates the process of "learning":

$$Pr(\text{"hypothesis"} | \{\text{data}\}, I) \propto Pr(\{\text{data}\} | \text{"hypothesis"}, I) \times Pr(\text{"hypothesis"} | I),$$

where the "hypothesis" is the quantity that we wish to infer (the bias-weighting of a coin, for example, or the neutron scattering law for some sample) and I represents any prior knowledge we may have about the "hypothesis." The prior probability distribution, Pr("hypothesis", I), reflects our knowledge (or ignorance) about the "hypothesis" before we obtained the data. This prior state of knowledge is modified by the likelihood function, Pr({data}|"hypothesis", I), which encodes the nature of the "experiment" and involves the use of deductive logic, to yield our posterior probability distribution, Pr("hypothesis"|{data},I), which represents our state of knowledge about the "hypothesis" after we have obtained the data. What we infer about some quantity of interest depends not only on the data we have but also on what we know or assume about it a priori! If the data are accurate, abundant, and sensitive to the quantity of interest, then the likelihood function will be sharply peaked and will dominate the posterior probability distribution. No matter what our prior state of knowledge, the data force us to the same conclusion. If the data are inaccurate, few in number, or insensitive to the quantity of interest, then the posterior will depend crucially on our prior. In other words, if the data do not tell us very much, then our state of knowledge after we have obtained the data will be governed largely by our state of knowledge (or ignorance) before the experiment.

Just as Newton's second law of motion is central to all classical mechanics, Bayes' theorem provides the fundamental rule for all logical and consistent inductive inference. Many statistical tests and procedures can be derived, justified, or at least understood from Bayes' theorem when one states the relevant assumptions and details about the situation under consideration. Model fitting, least squares, maximum likelihoods, singular-value decomposition, the maximum entropy method, Tikhonov regularization, Fourier filtering, the χ^2 test, the F test, Student's t test, and other statistical procedures for analyzing data can all be seen as suitable courses of action for different choices or assumptions about three things: the prior information I, which can even determine what we mean by "hypothesis"; the prior probability distribution, Pr("hypothesis"|I); and the nature of the experiment, which is enshrined in the likelihood function, Pr({data}|"hypothesis", I).

The Maximum Entropy Method

The data-analysis method known as maximum entropy (MaxEnt) arises in the context of a specific but commonly occurring problem—that of making inferences about *positive* and *additive* distributions. The neutron scattering law $S(Q,E)$ for a sample is an example of such a positive and additive distribution. It is positive because $S(Q,E)dQdE$ is proportional to the number of neutrons scattered with momentum transfer between Q and $Q + dQ$ and energy transfer between E and $E + dE$. It is additive because the number of neutrons scattered into a large $\Delta Q \Delta E$ interval is equal to the sum of the neutrons scattered into the small $dQdE$ intervals that compose the large $\Delta Q \Delta E$ interval. Other examples of positive and additive distributions include probability distribution functions, the radio-frequency brightness function of an astronomical source, the electron density in a crystal, the intensity of incoherent light as a function of position in an optical image, and so on. (By contrast, the amplitude of incoherent light is positive but not additive.) Given only the information I that the quantity of interest is a positive and additive distribution f, what should we assign as the prior probability distribution $\Pr(f|I)$? The assignment of a prior is often a difficult problem. Bayes' theorem tells us that the prior is a necessary and integral part of making a scientific inference, but the theorem does not tell us how to assign it. Methods that seem to avoid the use of a prior merely make an implicit choice (usually of a uniform distribution) rather than state an explicit choice. (Luckily, as mentioned above, the prior does not matter very much when we have "good" data.)

The choice of a prior usually involves somewhat obscure arguments and frequently involves a consideration of the allowed transformation groups that specify our ignorance about the quantity of interest. For example, consider the problem of estimating the length L of a biological molecule. What prior $\Pr(L)$ should be assigned to express complete ignorance about the value of L before we have carried out any measurements? Well, if we are really ignorant about the size of the molecule, then we should assign the same prior for the numerical value of L irrespective of whether we make the measurement in meters, inches, cubits, or whatever. The variable L would then be a so-called *scale parameter*. To express our complete ignorance about the value of a scale parameter, we say that the prior must be invariant under a change of scale in the measurement units. Mathematically we require that $\Pr(L)dL = \Pr(\beta L)d(\beta L)$ for all values of $\beta \geq 0$, leading us to the assignment $\Pr(L) \propto 1/L$, or a uniform prior for $\log L : \Pr(\log L) = $ constant.

The appropriate prior for a positive and additive distribution is, again, not immediately obvious. Many different types of arguments, however, including logical consistency, information theory, coding theory, and combinatorial arguments, lead us to believe that the prior is of a rather special form:

$$\Pr(f | I, \alpha, m) \propto \exp[\alpha S(f, m)]. \tag{3a}$$

Here the (prior) information I assumes only that f is positive and additive, and S is the generalized Shannon-Jaynes entropy:

$$S(f, m) = \int \{f(x) - m(x) - f(x)\log[f(x)/m(x)]\}dx. \tag{3b}$$

In this general expression for entropy, $m(x)$ is a Lebesgue measure on x, the space of the distribution, and α is a dimensional constant (initially unknown). We will say more about what this entropic prior means (and the value of α) a little later, but let us continue by considering $m(x)$ further.

In the absence of any data, the posterior becomes directly proportional to the prior, and our best estimate of f is given by the maximum of the entropy function S, which occurs at $f(x) = m(x)$. The function $m(x)$ is therefore a *default model* (that is, the solution to which f will default unless the data say otherwise) and can be thought of as representing our prior state of knowledge, or ignorance, about f. The default model is usually taken to be uniform (that is, constant), but the use of a nonuniform $m(x)$ can be important for such difficult problems as protein crystallography or for introducing spatial correlations across the positive and additive distribution we want to infer. If we know that f is normalized, so that $\int f(x)dx$ is fixed, and if the Lebesgue measure is uniform ($m(x) = $ constant), then the entropy formula above reduces to the form

$$S(f) = -\int f(x)\log[f(x)]dx,$$

which is the form of the entropy familiar from statistical mechanics.

The other quantity we need in order to make an inference about the distribution f is the likelihood function $\Pr(\{\text{data}\}|f, I)$. The likelihood function incorporates the information about the experiment, whether it is a neutron-scattering experiment, a nuclear-magnetic-resonance experiment, a radio-astronomy experiment, or whatever. It relates the quantity of interest to the data we have, thereby encoding details about the type of experiment and the accuracy of the measurements.

Let us consider the common case in which the data are independent (one measurement does not affect another) and are subject to additive Gaussian noise. The likelihood function then takes the form

$$\Pr(\{\text{data}\}|f,I) \propto \exp(-\frac{1}{2}\chi^2), \qquad (4a)$$

where χ^2 is the familiar misfit statistic, which measures how well a trial distribution f fits the actual data:

$$\chi^2 = \sum_{k=1}^{N} \frac{(D_k - F_k)^2}{\sigma_k^2}. \qquad (4b)$$

Here D_k is the kth datum (say the number of neutron counts in the kth bin), σ_k is the noise, or expected error, in that datum (for a neutron-scattering experiment, $\sigma_k = \sqrt{D_k}$), and F_k is the value for the kth datum that a trial distribution f would have produced in the absence of any noise. The noise in the neutron counts, though really described by a Poisson distribution, is approximated well by a Gaussian distribution because the number of counts is usually large (≥ 10). Thus the usual model fitting corresponds to assuming the likelihood function in Eq. 4 and maximizing that function to obtain the best "fit" to the data (that is, implicitly assuming a uniform prior so that the posterior becomes directly proportional to the likelihood function).

According to Bayes' theorem we must combine Eq. 3, the entropic prior, with the likelihood function of Eq. 4 to find the posterior probability distribution for f:

$$\Pr(f|\{\text{data}\},I,\alpha,m) \propto \exp(\alpha S - \frac{1}{2}\chi^2).$$

Then, given the data and only the prior knowledge that f is a positive and additive distribution, our best estimate of f is given by the distribution that maximizes this posterior probability distribution. Since the exponential is a monotonic function, we obtain the solution by maximizing $\alpha S - \frac{1}{2}\chi^2$ (a general algorithm to do this is given in Bryan and Skilling 1984). This procedure can be interpreted as maximizing the entropy S subject to some constraint on the value of the misfit statistic χ^2, where the initially unknown constant α is seen as a Lagrange multiplier. Hence the name *maximum entropy* method. The method is illustrated schematically in Fig. 4.

In past applications of the maximum entropy method, the constant α was chosen such that $\chi^2 = N$, where N is the number of data. This choice seems intuitively reasonable since any proposed distribution f should give data consistent with those actually measured, as defined by the constraint that $\chi^2 < \approx N$. The MaxEnt method was thus seen as choosing a distribution f that, while "fitting the data," had the most entropy. More recent thinking (Skilling and Gull 1989), however, carries the Bayesian logic one step further: Since α is unknown, it becomes just one more parameter that needs to be estimated in the same sense that we are trying to estimate f. This approach, which leads to a slightly more complicated (but less ad hoc) criterion for the choice of α, has the advantage that the increased rigor allows us to automatically determine σ, the level of the noise, or expected error, in the measured data if it is not known. We leave these and other recent advances, including a discussion of practical reliability estimates of the inferred distribution f, to the avid reader (see Further Reading) and continue to pursue the more traditional approach to the MaxEnt method and its applications.

The Meaning of Maximum Entropy

Well, we have talked about the entropic prior, but what is its significance and what does it mean? To answer this question, we will use two very simple examples. The first, known as the *kangaroo problem*, is an example of having accurate but in-
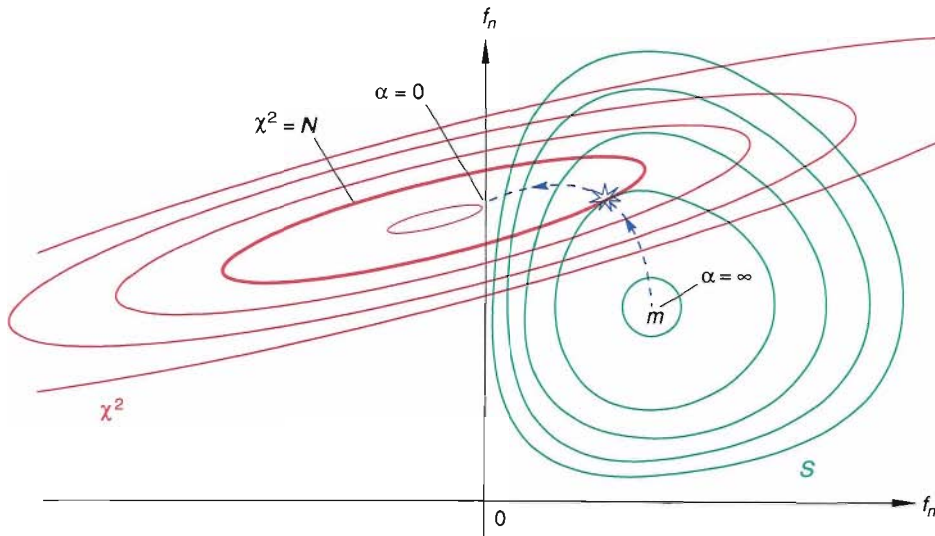
THE MAXIMUM ENTROPY METHOD

Fig. 4. Suppose that we are trying to find the "best" estimate for some positive and additive distribution $f(x)$. Suppose further that the hypothesis space of f is defined by the values of f specified on a grid finely discretized with respect to x into N pixels. In other words, the hypothesis space of f is the N-dimensional space whose coordinate axes are the set $\{f_j\}$, where f_j is the value of f at pixel j. Shown here is a schematic two-dimensional section, namely the $f_m f_n$ plane, through the hypothesis space. Plotted (in red) are contours along which χ^2, (twice) the logarithm of the likelihood function, is constant; the set $\{f_k\}$ for which $\chi^2 < \approx N$ (the number of data) compose the feasible set of distributions allowed by the data. Also plotted (in green) are contours along which the entropy S (the logarithm of the prior probability distribution) is constant; the entropy is a maximum at the default model $f = m$ (where m is a Lebesgue measure on the hypothesis space) and rapidly approaches $-\infty$ as any part of f becomes negative. The MaxEnt solution is that f for which the posterior probability distribution is maximum, that is, the f for which $\partial/\partial f_j(\alpha S - \frac{1}{2}\chi^2) = 0$. The blue line indicates the trajectory of the MaxEnt solution as the value of the Lagrange multiplier α goes from ∞ to 0; the blue star represents the traditional choice of α, which satisfies the condition that $\chi^2 = N$.

sufficient data. Nevertheless the problem is small enough that common sense tells us what constitutes a "sensible" solution. It will be shown that the MaxEnt choice, unlike several commonly used alternatives, concurs with our common sense. We will then use a second example, known as the *monkey argument*, to try to give a more general interpretation of the MaxEnt solution.

The Kangaroo Problem. We have said that in the MaxEnt method we choose, as our best estimate of a positive and additive distribution f, the f that agrees with the data and has the most entropy. This method of choosing a solution by maximizing some function of the desired distribution is known as regularization. The Shannon-Jaynes entropy is an example of a regularizing function, but several others are also commonly used. We will follow Gull and Skilling (1984) in using the kangaroo problem to demonstrate our preference for the choice of the Shannon-Jaynes entropy over the alternatives. The kangaroo problem, a physicists' perversion of a formal mathematical argument (Shore and Johnson 1980), shows that the Shannon-Jaynes entropy is the only regularizing function that yields self-consistent results when the same information can be used in different ways (for example, the choice of coordinate system should not matter). The kangaroo problem is as follows.

Information: One-third of all kangaroos have blue eyes, and one-third of all kangaroos are left-handed.

Question: On the basis of this information alone, what proportion of kangaroos are both blue-eyed and left-handed?

Clearly, we do not have enough information to know the correct answer: All solutions of the type shown in the 2×2 contingency table of Fig. 5a agree with the data and thus constitute the *feasible set* of solutions. Without additional information, each solution is equally likely because they all fit the data exactly. Figure 5b shows three among the myriad of feasible solutions: namely, the one with no correlation between being blue-eyed and left-handed and the ones with the maximum positive and negative correlation. Although the data do not allow us to say which is the *correct* answer, our common sense compels us to choose the uncorrelated solution if we are forced to make a choice. That is to say, unless we have prior knowledge to the contrary, we do not expect that knowing the eye color of a kangaroo will tell us anything about whether the kangaroo is left-handed or right-handed. Thus our best estimate is that one-ninth of the kangaroos will be blue-eyed and left-handed.

Table 1 shows the results of selecting the solution by maximizing four commonly used regularizing functions. Note that the integral in the formula for the entropy, for example, has been replaced by a summation because the space of the distribution, x, is not continuous but discrete. In fact, it consists of just four pixels— the four boxes in the 2×2 contingency table. For this very simple example, where

TRUTH TABLES FOR
THE KANGAROO PROBLEM

Fig. 5. (a) This truth table illustrates the general feasible solution to the kangaroo problem. That solution is obtained by letting $x \equiv f_1$ be the fraction of kangaroos that are blue-eyed and left-handed, where $0 \leq x \leq \frac{1}{3}$. Then the fractions corresponding to the other contingencies (f_2, f_3, and f_4) can be expressed in terms of x. (b) These truth tables illustrate three specific solutions derived by setting x to $\frac{1}{9}$, which corresponds to no correlation between being blue-eyed and left-handed, and by setting x to $\frac{1}{3}$ or 0, which correspond respectively to maximum positive or negative correlation between the traits.

(a) General Feasible Solution

Left-Handed

	T	F
T	$f_1 = x$ $0 \leq x \leq \frac{1}{3}$	$f_2 = \frac{1}{3} - x$
F	$f_3 = \frac{1}{3} - x$	$f_4 = \frac{1}{3} + x$

Blue-Eyed

(b) Three Specific Solutions

Left-Handed

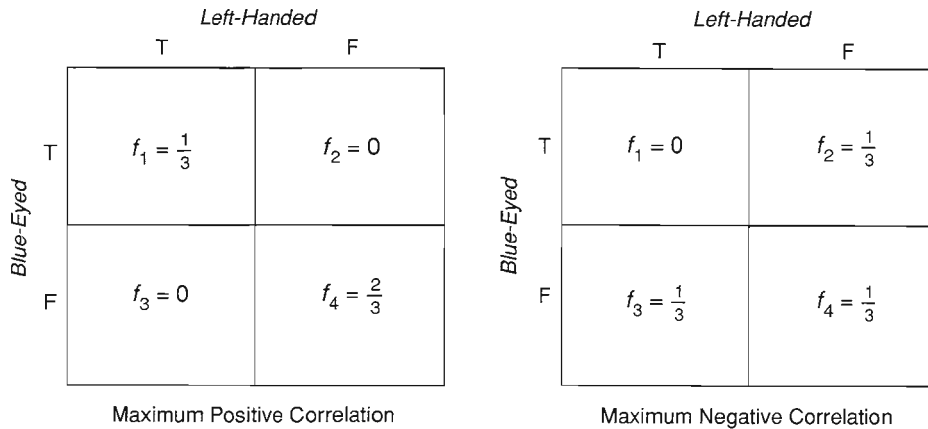|	T	F
T	$f_1 = \frac{1}{9}$	$f_2 = \frac{2}{9}$
F	$f_3 = \frac{2}{9}$	$f_4 = \frac{4}{9}$

Blue-Eyed

No Correlation

common sense tells us the "best" answer when faced with insufficient (but noise-free) data, it is only the Shannon-Jaynes entropy that yields a sensible answer! (Although we have considered only four regularizing functions, it can be shown that the Shannon-Jaynes entropy is the only one that has this desired property.)

Before going on to consider a more general interpretation of the MaxEnt choice, it is worth commenting on the frequently heard statement that in data analysis (or image reconstruction) positivity is the important constraint, not how you enforce it. For large problems that statement is very often true. Our small kangaroo problem, however, magnifies the differences among the regularizing functions and shows that we get more from MaxEnt than just positivity. The way we have set up the problem in Fig. 5a has the positivity constraint already built in, but it is still not sufficient to make a choice on the basis of the data we are given. The $\sum f^2$ regularizing function, for example, which for the kangaroo problem corresponds to the "Tikhonov with positivity" that some people seek, does not yield the same solution as our common sense—only MaxEnt does! Many general image-processing methods (both ad hoc and sound) often give similar results. The similarity merely reflects the fact that the prior probability distribution does not usually matter very much when the data are "good." However, if we assume only that the quantity of interest is a positive and additive distribution and ask what is the appropriate choice for the prior, the answer is the entropic prior.

The Monkey Argument. Our common sense recommended the uncorrelated solution to the kangaroo problem because, intuitively, we knew that it was the most noncommittal choice. The data did not rule out correlation, but, without actual evidence, it was a priori more likely that the genes controlling handedness and eye color were on different chromosomes than on the same one. Crudely speaking, if we consider randomly scattering two genes among eight chromosomes, they are seven times more likely to land on different chromosomes than on the same one. Although we cannot usually appeal to specific knowledge such as what is known about genes and chromosomes, we can use the monkey argument (Gull and Daniell 1978) to see more generally that the MaxEnt choice is the one that is *maximally noncommittal* about the information we do not have.

The monkey argument can again be thought of as a physicists' perversion of formal mathematical work, that of Shannon (1948) showing that entropy is a unique measure of "information content." The words "information content" are being used here in the information-theory sense and have somewhat the opposite sense of their everyday use! We might better think of entropy as a measure of uncertainty (rather than as a measure of information) because uncertainty is closer to the idea of the lack of order that characterizes entropy. However, a system that has more entropy has a

Left-Handed

	T	F
T	$f_1 = \frac{1}{3}$	$f_2 = 0$
F	$f_3 = 0$	$f_4 = \frac{2}{3}$

Blue-Eyed

Maximum Positive Correlation

Left-Handed

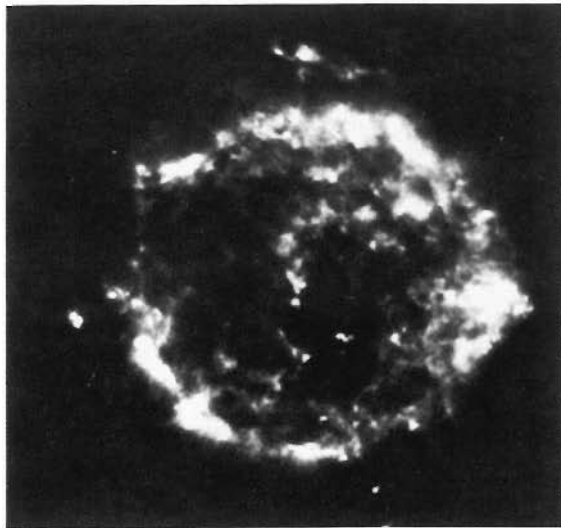|	T	F
T	$f_1 = 0$	$f_2 = \frac{1}{3}$
F	$f_3 = \frac{1}{3}$	$f_4 = \frac{1}{3}$

Blue-Eyed

Maximum Negative Correlation

Table 1

REGULARIZATION-FUNCTION SOLUTIONS OF KANGAROO PROBLEM

Listed here are values of x (fraction of kangaroos that are blue-eyed and left-handed) derived by maximizing four commonly used regularizing functions. Of the four only the Shannon-Jaynes entropy, $-\sum f_j \log f_j$, yields a value for x that agrees with our common sense, which tells us that, in the absence of relevant data, the two traits are most likely to be uncorrelated.

Regularizing Function	x	Correlation
$-\sum f_j \log f_j$	0.111..., or $\frac{1}{9}$	None
$-\sum f_j^2$	0.083..., or $\frac{1}{12}$	Negative
$\sum \log f_j$	0.13013	Positive
$\sum \sqrt{f_j}$	0.12176	Positive

greater degree of randomness, and its description requires more information (more bits in a computer). It is in this sense, then, that entropy is a measure of information.

The monkey argument presents the MaxEnt solution in graphic terms. Imagine a large team of monkeys who make images, or positive and additive distributions, by randomly throwing small balls of light at a rectangular grid. After a while, the grid is removed and replaced by another and so on. Eventually, the monkeys will generate all possible images, and many copies of each one. If we want an image of an object about which we have some experimental data, we can reject most of the monkey images because they give data that are inconsistent with the experimental measurements. Those images that are not rejected constitute the feasible set. If we are to select just one image from this feasible set as representing our best estimate of the object, the image that the monkeys generate most often would be a sensible choice. Because our hypothetical team of monkeys is presumed to have no particular bias, such a choice represents the image that is consistent with the measured data but, at the same time, is most noncommittal about the information we do not have. This preferred image is the MaxEnt solution, because the entropy is just the logarithm of the number of ways in which the image could have been generated (and, hence, the number of times it was).

Applications of MaxEnt at LANSCE

MaxEnt has been used successfully in image reconstruction in a wide variety of fields (see, for example, Gull and Skilling 1984). A small selection of its diverse applications, shown in Fig. 6, include forensic deblurring, radio astronomy, medical tomography, and nuclear-magnetic-resonance spectroscopy. We are now starting to use this powerful technique, and Bayesian ideas in general, to enhance the analysis of neutron-scattering data at LANSCE.

The Filter-Difference Spectrometer. The first example of the use of MaxEnt at LANSCE is the analysis of data from the Filter-Difference Spectrometer, or FDS. This example has the form of a standard convolution problem. That is, the data are related to the quantity of interest through a blurring process, so that they are a blurred (and noisy) version of what we want.

Our own eyes produce such a convolution, or blurring. Because the pupils of our eyes have a finite size, we do not see point sources of light as infinitesimal dots but as small fuzzy disks. (The angular size of the disk is roughly λ/d, where λ is the wavelength of the light and d is the diameter of the pupil.) If two point sources of light are so close that the disks overlap, we can no longer distinguish them as separate entities. Such blurring, response, resolution, or point-spread functions occur al-
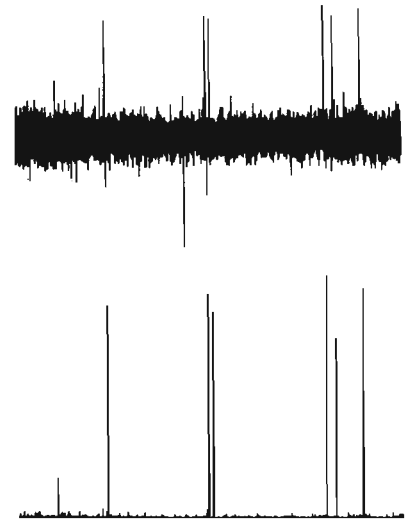
MaxEnt deblurring of a photographic image

MaxEnt x-ray tomograph of a human skull

MaxEnt image of radio-frequency (5-gigahertz) emissions from the supernova remnant Cassiopeia A

Comparison of (top) conventional Fourier reconstruction and (bottom) MaxEnt reconstruction of a nuclear-magnetic-resonance spectrum

SOME APPLICATIONS OF THE MAXIMUM ENTROPY METHOD

Fig. 6. The examples presented here, which are reproduced through the courtesy of J. Skilling and S. F. Gull, show how the maximum entropy method can be used to clarify the information extracted from a variety of data.

most universally in all experiments. The resolution functions of particular interest in neutron-scattering experiments arise from various aspects of the experimental setup, such as the finite size and temperature of the moderator and the finite angle of collimation of the neutron beam. In the case of the FDS, the major contribution comes from the transmission spectrum of the polycrystalline filters used to select for recording those inelastically scattering neutrons with certain final energies.

For those not familiar with the idea of a convolution, or the performance of MaxEnt, we start with a simple simulated example computed on a grid of 128 points. Suppose that the "true" object, or neutron scattering law, consists of two spikes on the left separated by a small plateau and a broader peak on the right, as shown in Fig. 7a. Also suppose that a noisy data set (Fig. 7c) is generated by first convolving the scattering law with a resolution function (Fig. 7b) that is similar to the transmission spectrum of the filters used in the FDS and then adding to the resulting blurred signal a small background count and random noise. In a convolution each point of the object (pixel) is replaced with a copy of the resolution function scaled by the "height" of the object at that point; the data are then the sum of all the scaled copies of the resolution function. As can be seen from Figs. 7a and 7c, a large single spike can give much the same data as a smaller broad peak. Mathematically, using matrix and vector notation, we can write the "experiment" as

$$\mathbf{d} = O \cdot \mathbf{f} + \mathbf{b} \pm \sigma.$$

Here $\mathbf{d}$ is the data vector, the matrix O is the convolution operator ($O_{jk} = r_{k-j}$, where $\mathbf{r}$ is the resolution function), $\mathbf{f}$ is the scattering law, $\mathbf{b}$ is the background, and σ is the root-mean-square value of the random noise ($\langle \sigma_k^2 \rangle = d_k$). Given the data set and a knowledge of the resolution function and background, we wish to infer the underlying scattering law. A simple way of performing the deconvolution is to ap-

(a) Test Scattering Law

(b) Resolution Function

(c) Simulated Data

(d) Direct Inverse

(e) "Smoothed" Direct Inverse

(f) MaxEnt Reconstruction

DECONVOLUTION OF SIMULATED DATA

Fig. 7. The power of the maximum entropy method is illustrated by its application to a simulated data set. The simulated data set (c) was obtained by convolving a test scattering law (a) with an instrumental resolution function (b) and then adding a small background and random noise. (The instrumental resolution function shown in (b) is similar to the transmission spectrum of the filters used in the Filter-Difference Spectrometer at LANSCE). The series (d), (e), and (f) compares reconstructions, or deconvolutions, of the mock scattering law produced by three methods.
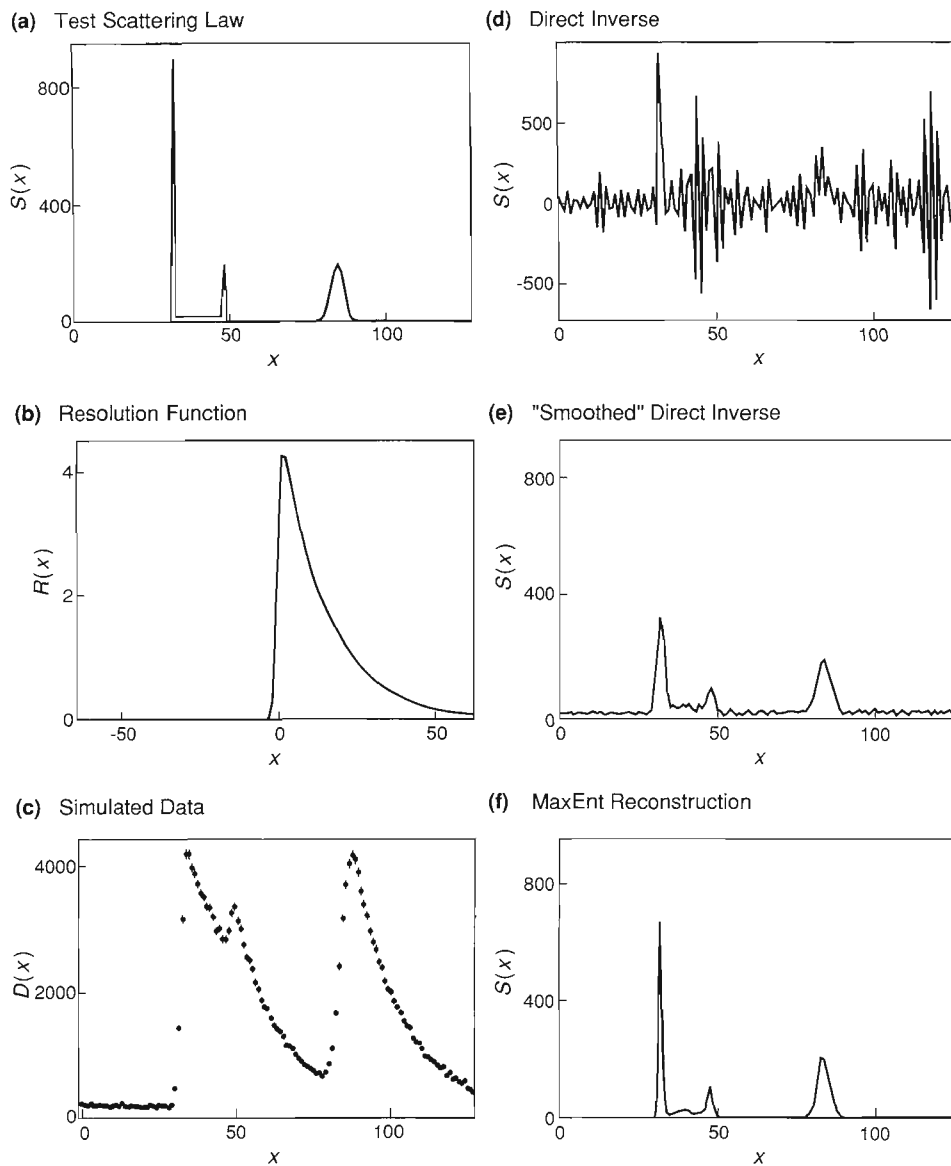
ply the inverse convolution operator O^{-1} to $\mathbf{d} - \mathbf{b}$ by using Fourier transforms. This procedure is equivalent to making the assumption that the prior is uniform, $\Pr(f) =$ constant, and determining the *maximum-likelihood* solution. Unfortunately, the inverse solution may not exist. For example, the maximum-likelihood solution may not be unique because of missing data. Furthermore, even when the inverse does exist, it produces a reconstruction of the scattering law (Fig. 7d) that has a lot of high-frequency ringing (wiggles). To overcome this difficulty, it is common practice to use a smoothed (or slightly blurred) version of the direct inverse, a procedure known as *Fourier filtering* (Fig. 7e). In the grand scheme of things, Fourier filtering can be regarded as an example of *singular-value decomposition*. An alternative approach is to use the fact that the scattering law is a positive and additive distribution and hence choose an entropic prior ($\Pr(f|I) \propto \exp(\alpha S)$) and thus obtain the MaxEnt solution shown in Fig. 7f. We find that the maximum entropy method has suppressed the level of the artifacts without sacrificing as much detail in the reconstruction as does Fourier filtering.

Now, let us turn from simulated data to real data. The FDS is an instrument used to perform molecular rotational-vibrational spectroscopy with neutrons rather than with photons, as in infrared or Raman spectroscopy. Figure 8a shows data taken with a beryllium filter imposed between the sample and the detector. Those data

DECONVOLUTION OF INELASTIC-NEUTRON-SCATTERING DATA

Fig. 8. Shown in (a) are inelastic neutron-scattering data obtained with the Filter-Difference Spectrometer at LANSCE. Such data are the basis for deducing the energy levels of the molecular vibrations and rotations excited in a sample by the incident neutrons. (Here the sample is hexamethylene tetramine at 15 kelvins; its well-known rotational-vibrational spectrum is used to calibrate the energy-transfer values deduced from the recorded times of flight and the energy-cutoff points of the filters.) The raw data shown in (a) are a convolution of the true rotational-vibrational spectrum of the sample with the transmission spectrum of a beryllium filter located between the sample and the detector. (That transmission spectrum is similar to the resolution function shown in Fig. 7b.) Shown in (b) is the "filter-difference" spectrum, a hardware deconvolution of the data in (a) derived by subtracting the raw data in (a) from raw data obtained with a beryllium oxide filter. (The transmission spectrum of a beryllium oxide filter differs from that of a beryllium filter mainly in being slightly shifted in energy.) The filter-difference spectrum is inverted relative to the raw data plot because the abscissa in (a) is (essentially) the time of flight of the scattered neutrons whereas the abscissa in (b) (and (c) and (d)) is the energy transferred to the sample. Shown in (c) is the MaxEnt reconstruction of the data in (a). The MaxEnt reconstruction and a filtered inverse, or "Mezei," reconstruction (dots) are compared in (d).

(a) Raw Data (Beryllium Filter)

(b) "Filter-Difference" Spectrum

(c) MaxEnt Reconstruction

**(d)** MaxEnt and Mezei Reconstructions
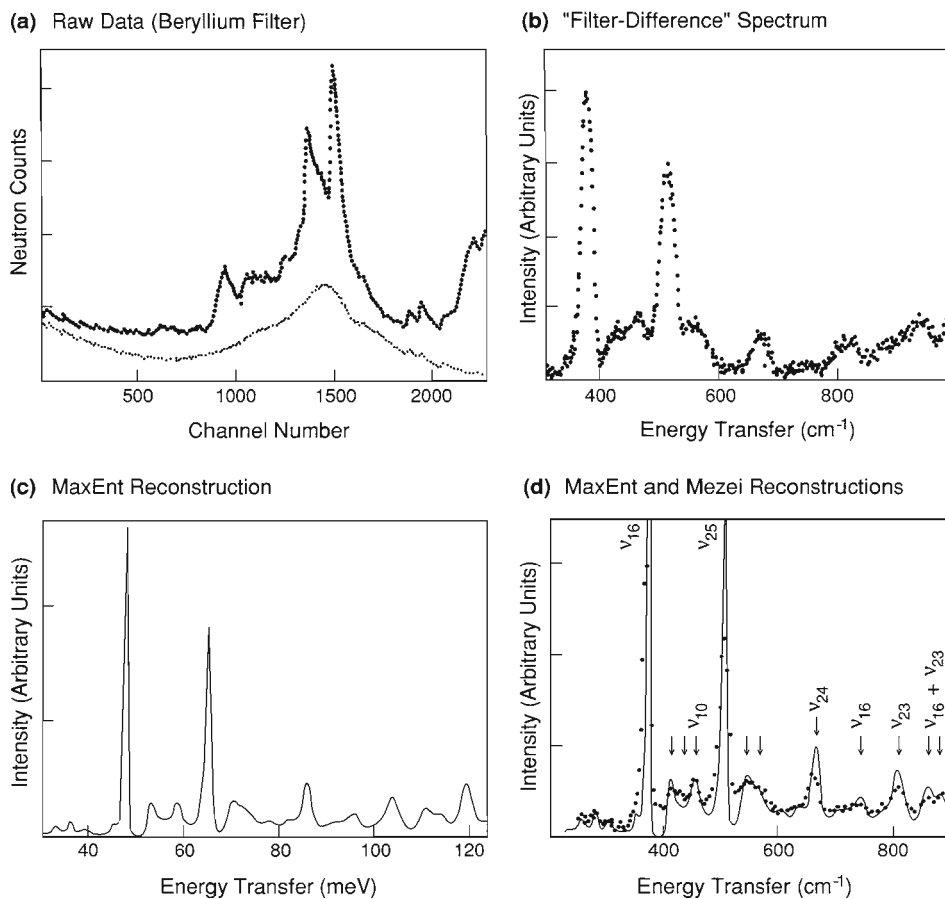
show the effects of the sharp edge and long decaying tail of the transmission spectrum of the filter (see Fig. 7b). The earliest method used to remove the blurring produced by such a resolution function is a hardware solution. Two data sets are collected, one consisting of the scattered neutrons transmitted through a beryllium filter and the other consisting of the scattered neutrons transmitted through a beryllium oxide filter. The transmission spectra of the two filters have almost the same shape, but their sharp energy cutoffs are slightly offset. Therefore, the data set obtained with one filter differs from the data set obtained with the other filter mainly in being shifted in energy by a small amount. When the two data sets are subtracted, the contributions from the long decaying tails (and background) tend to cancel, and only the significant features defined by the sharp rising edges remain. Figure 8a shows raw data obtained with only the beryllium filter plotted in data channels corresponding to increasing neutron time of flight. Figure 8b shows the corresponding "filter-difference" spectrum plotted as a function of energy transfer. The filter-difference spectrum is inverted relative to the data plot because increasing time of flight is equivalent to decreasing energy transfer.

Given only the data obtained with the beryllium filter and knowledge of the filter's transmission spectrum and the background, the deconvolution can be carried out mathematically (in software) by using the maximum entropy method. The MaxEnt reconstruction thus obtained is shown in Fig. 8c and is compared in Fig. 8d with a conventional reconstruction (due to Mezei) that can be interpreted as a filtered inverse. As expected, the MaxEnt reconstruction is an improvement over both the filter-difference and the Mezei deconvolutions in that it shows finer detail and fewer noise artifacts. The improvement is obvious but not dramatic because the data have good statistical accuracy. Noisier data causes the filtered inverse solution to deteriorate much more rapidly than the MaxEnt solution.
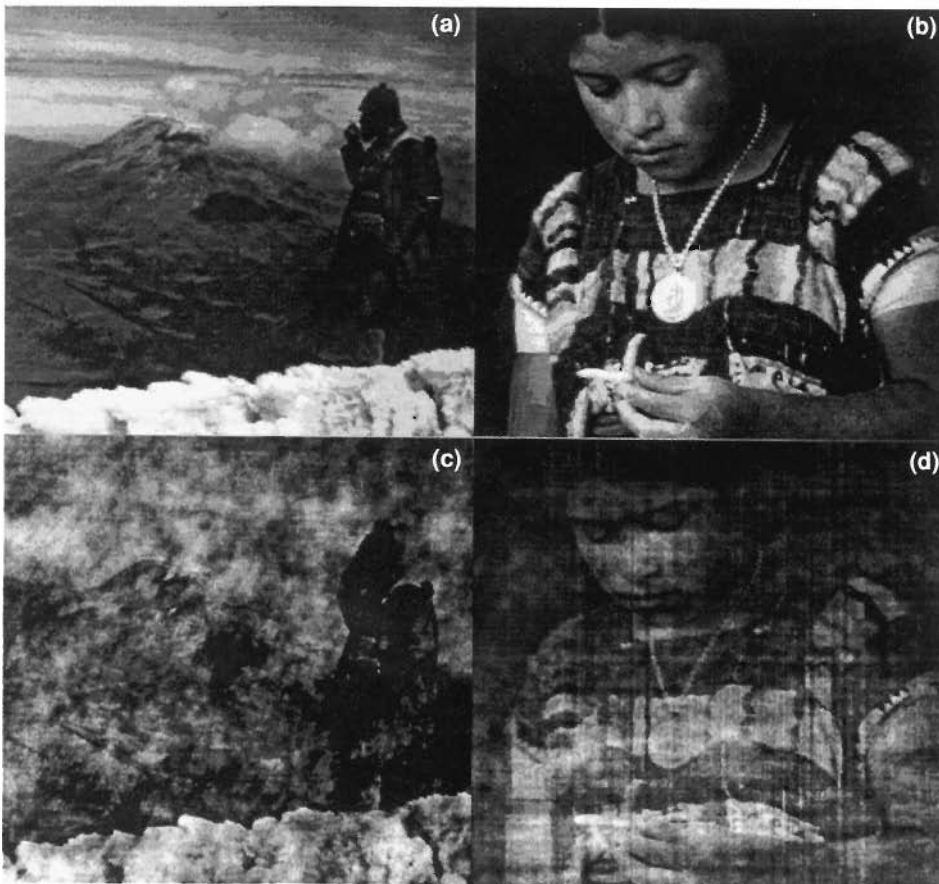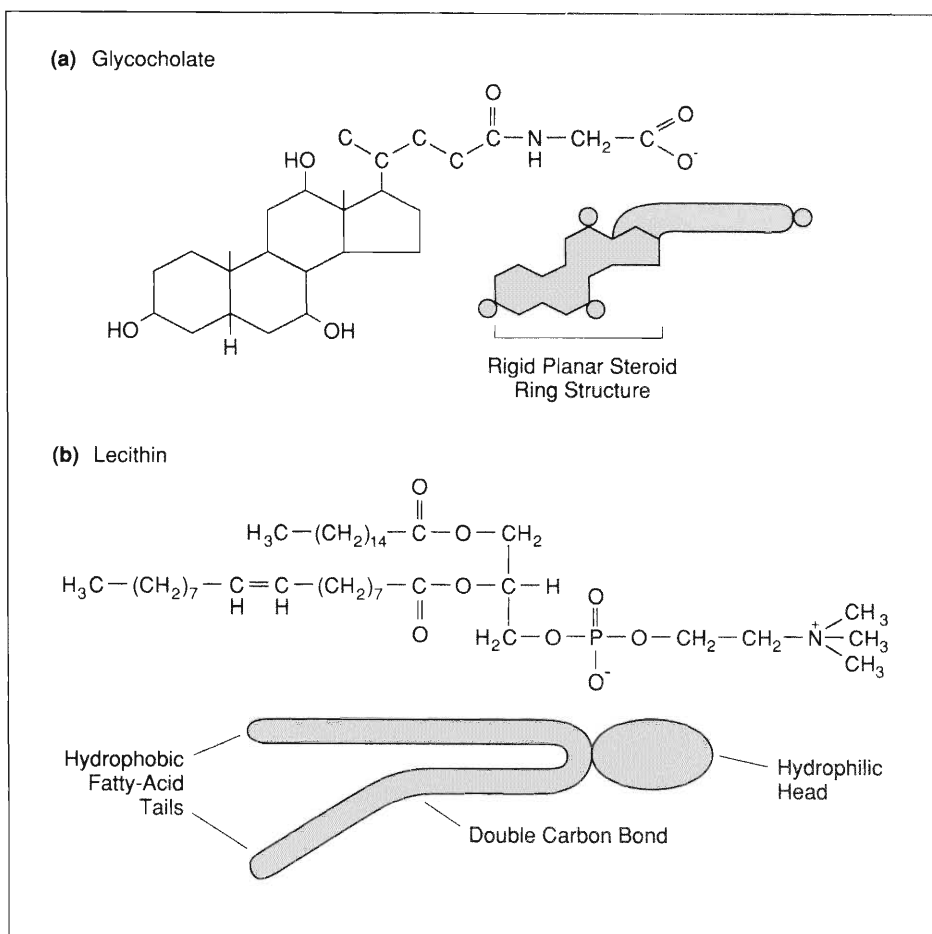
The Low-Q Diffractometer. The next application of the maximum entropy method involves the analysis of data that reflect the aggregation of biological macromolecules in solution. These data were taken on the Low-Q Diffractometer, a small-angle neutron-scattering (SANS) instrument useful for studying structures with dimensions ranging from 10 to 1000 angstroms. The spatial distribution of particles in a sample (including their size, shape, and location) is related to the neutron scattering law through a Fourier transform—in general, a complex quantity. (The elements of the O matrix for a Fourier transform are of the form $O_{jk} = \exp(i2\pi jk/N)$, where $i^2 = -1$ and N is the number of points in the discrete Fourier transform.) The neutron counts we measure are, of course, given by the Fourier intensities (or a blurred and noisy version thereof). We are thus brought face-to-face with the dreaded *Fourier phase problem*! The Fourier phase problem entails trying to make an inference about some quantity of interest given information about only the amplitudes (but not the phases) of its Fourier transform. It is a notoriously difficult problem, well known in x-ray crystallography, because the many local maxima of the likelihood function make it hard for us to find the global maximum of the posterior probability. The gravity of the situation is illustrated by Fig. 9. Luckily, we are not interested in determining the relative locations of the particles but only the number of particles of a given size and shape. Thus our problem is analogous to the problem in x-ray crystallography of determining not the electron-density map but only the autocorrelation (or Patterson) function, for which the Fourier intensities alone are sufficient.

The particles under study are involved in the digestion and transport of fats. My biologist colleague at LANSCE, Rex Hjelm, introduced me to the problem by saying: "Your body is mostly water. If you visit your favorite ice-cream parlor, then the fat in the ice cream will form a greasy blob at the bottom of your stomach and you will soon die!" He then told me that bile salts, produced in the liver, had hy-

DIGESTION OF FATS

Fig. 10. The digestion of dietary fats begins with their emulsification by complexes of bile-salt and lecithin molecules. Bile salts (glycocholate and taurocholate) are polar derivatives of cholesterol. Shown in (a) are the structural formula and a schematic diagram of glycocholate. (Polar groups are denoted in the schematic diagram by circles.) Taurocholate differs in that the terminal carboxyl group (CO_2^-) is replaced by the group $H_2C—SO_3^-$. Although glycocholate, say, is itself an effective emulsifier, complexes of glycocholate and the lipid lecithin are, for reasons not yet known, even more effective. Shown in (b) are the structural formula and a schematic diagram of lecithin, or phosphatidyl choline. Aiding the digestion of fats is not the only physiological function of lecithin; it also is a major constituent of the lipid bilayers that compose biological membranes.


(a) Glycocholate

Rigid Planar Steroid
Ring Structure

(b) Lecithin

Hydrophobic
Fatty-Acid
Tails

Double Carbon Bond

Hydrophilic
Head

drophylic heads and hydrophobic tails (Fig. 10a). "So the body dumps in some bile salts to act as detergents," I remarked, somewhat relieved. "No, that's what an engineer would do!" came the reply. For reasons that we do not fully understand, nature uses a conglomerate of bile salts and the fat lecithin (Fig. 10b) to begin the digestion process.

An understanding of the action of bile salts in lipid digestion and in the transport of liver products such as cholesterol has potential applications in industrial processes and in the development of drug-delivery systems and model membranes. As a step in this direction, Hjelm et al. (1990) have been investigating the nature of particle growth in aqueous solutions of lecithin and the bile salt glycocholate. Figure 11a shows SANS data sets for three increasingly dilute solutions. Hjelm asked the following question: If I assume that the particles in the sample can be modeled as cylinders of uniform density, what is my "best" estimate of their size distribution, given the data and a knowledge of the experimental setup? Since SANS data are not sensitive to fine structure, the sharp edges of the cylinders are of little consequence; all that we are really assuming is that the particles are "blobs" of uniform density defined by a length and a diameter. Moreover, the fact that the distribution of particle sizes is a positive and additive quantity means that the relevant prior for the distribution of particle sizes is an entropic prior! Figure 11b shows the particle-size distributions derived by using MaxEnt on the data in Fig. 11a. The distribution for the highest lipid concentration indicates the presence of only a single type of particle, roughly globular, with a diameter of about 50 angstroms. As the sample is diluted, evidence for a second type of particle appears, a rod-like structure with a diameter of about 50 angstroms and a length of about 100 angstroms, or twice the original length. Even greater dilution leads to the appearance of even more elongated particles with

(a) Small-Angle Neutron-Scattering Data
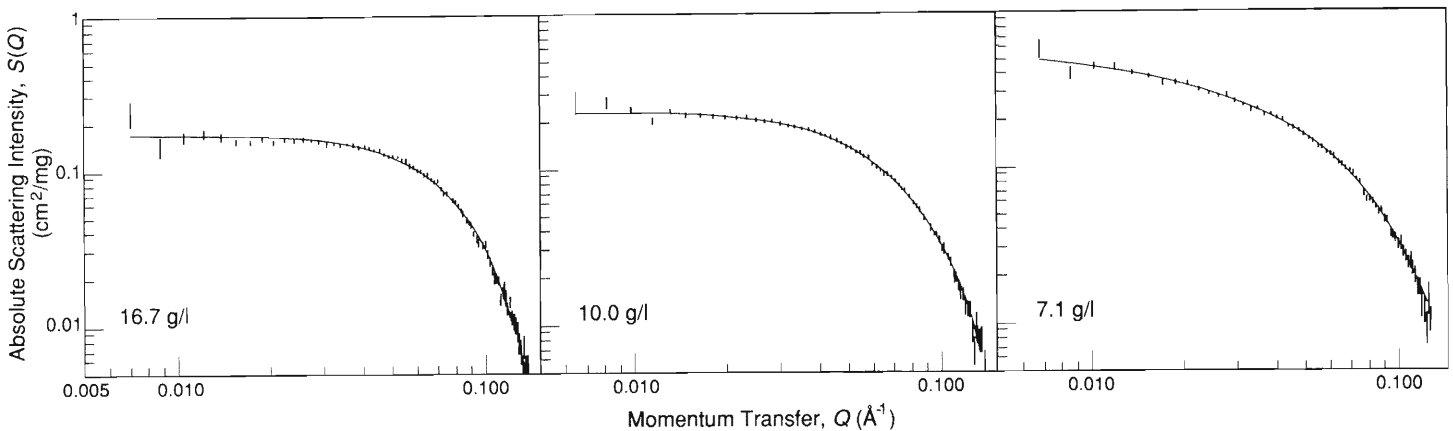
(b) Cylinder-Size Distributions
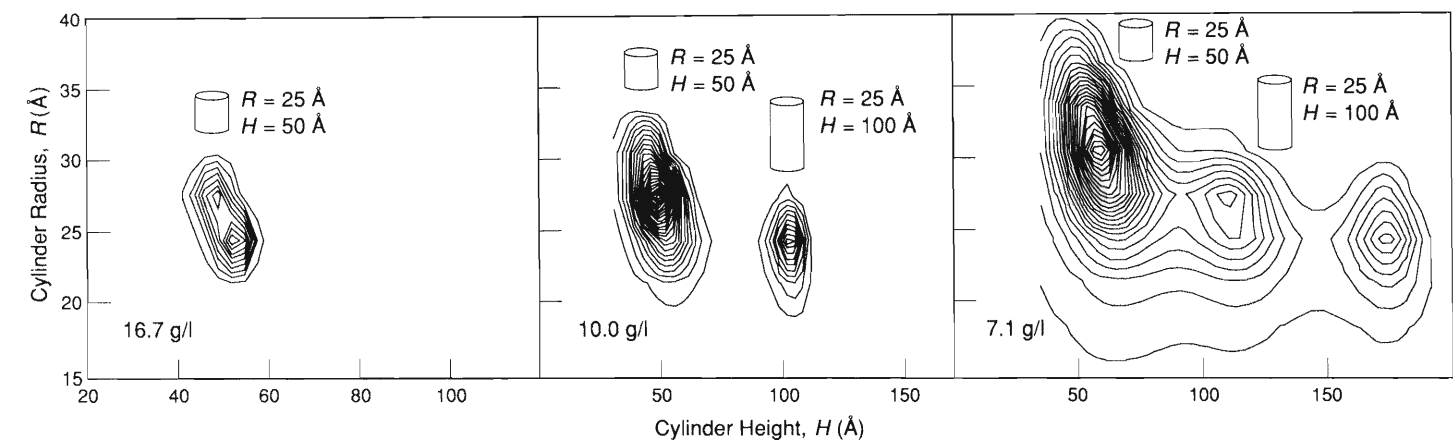
a length of about 170 angstroms, or three to four times the original length. These results lead us to believe that particle growth occurs through aggregation of preformed subunits with a size of about 50 angstroms (which corresponds nicely to the thickness of lecithin bilayers) rather than through the aggregation of individual bile-salt or lecithin molecules.

This example shows that the data need not bear any visual resemblance to the information extracted; in other words, MaxEnt is a method for data analysis, or scientific inference, and not just image enhancement.

The Constant-Q Spectrometer. Our last example involves data from the Constant-Q Spectrometer (CQS), an instrument designed to investigate phonons and magnons in single-crystal samples. The example illustrates a more advanced use of MaxEnt—*multichannel entropy*. This method is needed for convolution problems in which we want to determine not only the (sharp) scattering law of interest but also a broad, unknown background signal. We will begin with a simple simulation to illustrate multichannel entropy and then demonstrate its use on real data from the CQS.

For our simulation we convolve the scattering law of Fig. 7a with the resolution function of Fig. 7b (scaled down by a factor of about 10) and then add a large background, assumed to be unknown, to generate the noisy data set shown in Fig. 12a. To analyze these data we use the technique of *two-channel entropy*. We assume that the unknown background b is also a positive and additive quantity and is fairly broad compared with the scattering law f. What we are attempting to do is an example of multichannel entropy because we are trying to make our best inference about several different "images" simultaneously. In this case we have only two channels: one for the background and the other for the scattering law. We set up two image channels,

PARTICLE GROWTH

Fig. 11. Information about the sizes of glycocholate-lecithin complexes in an aqueous solution can be obtained by analysis of small-angle neutron-scattering data for the solution. Shown in (a) are such data (Hjelm et al. 1990) for increasingly dilute solutions. (The concentrations indicated are total concentrations of glycocholate plus lecithin.) The corresponding particle-size distributions, shown in (b), were derived by assuming that the complexes are adequately represented by cylinders of radius *R* and height *H* and then using the maximum entropy method to determine the most probable distribution of cylinder sizes. Note that increasing dilution is accompanied by the appearance of populations of cylinders whose radii do not change significantly but whose heights increase by approximately integral multiples.

(a) Simulated Data

(b) MaxEnt Reconstruction: Background

(c) MaxEnt Reconstruction: Signal
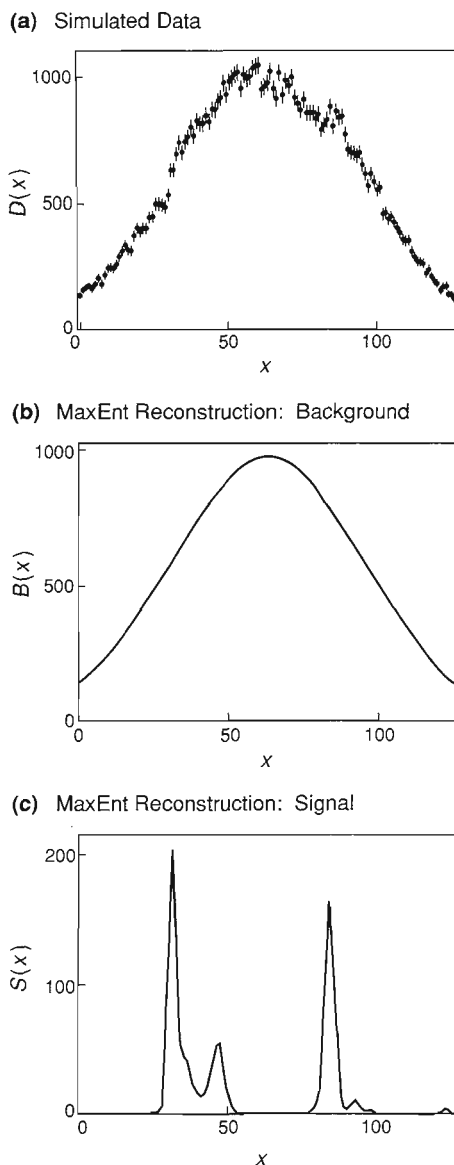
TWO-CHANNEL DECONVOLUTION OF SIMULATED DATA

Fig. 12. (a) Two-channel entropy is an example of an advanced use of the maximum entropy method. It allows deconvolution of data into two components, such as a scattering law with sharp features and a relatively featureless background. Application of the two-channel entropy method to the simulated data in (a), which were generated by convolving the scattering law and the resolution function shown in Fig. 7a and Fig. 7b and then adding a large, unknown background, yields the background and scattering-law reconstructions shown, respectively, in (b) and (c).

f_1 and f_2. One channel is allowed to have only broad structure (by construction); the other is permitted the full resolution of the 128-pixel grid. We also arrange the problem so that the "entropic cost" of putting structure in the broad channel is very low relative to the cost of putting structure in the high-resolution channel. What do we mean by entropic cost? Recall that the absolute maximum of entropy occurs when f is the same as the default model m. But as f deviates from m, in order to become consistent with the data, the entropy decreases, and that decrease in entropy is what we mean by entropic cost. Thus by making the entropic cost of putting structure in the broad channel relatively low, we ensure that if a broad distribution can account for the data, it will appear in the broad channel. If sharp structure is required, it can appear only in the high-resolution channel. We identify the high-resolution channel with the scattering law and the broad channel with the unknown background. Carrying out this procedure (for details see Sivia 1990), we obtain the MaxEnt reconstructions for the background and scattering law shown in Figs. 12b and 12c. Although the image of Fig. 12c is not as good as that of Fig. 7f, it is still a very impressive reconstruction in light of the given data (compare Fig. 7c with Fig. 12a!).

Finally, we show the application of this two-channel entropy algorithm to real data on the inelastic scattering of neutrons from phonons and magnons in a sample of iron. The data (Yethiraj et al. 1990) are shown in Fig. 13a as a function of the experimental variables: time of flight and detector angle. The data suffer from a combination of broadening and an unknown background signal (in addition to $\sqrt{N}$ noise) that obscures the scattering law of interest. The MaxEnt reconstruction of the signal, or high-resolution, channel (Fig. 13b) shows a dramatic improvement in both the detail seen in the scattering law and in the reduction of background artifacts. When the scattering law is plotted in terms of the physically meaningful coordinates of energy and momentum transfer (Fig. 13c), we can easily identify the dispersion curves for the magnon and phonon excitations characteristic of iron.

Instrument Design

The examples of the use of MaxEnt given in the last section are all cases of doing the "best" with the data we have. Usually that is all we can do. The instrumentation and hardware already exist at facilities like LANSCE, and often the only freedom a user has to improve the quality of the data is to increase its statistical accuracy by collecting data for a longer time. Let us suppose, however, that we are going to build a new facility, or just a new spectrometer. How should we design it to get the "best" data? This is an important question since a new facility can cost a hundred million dollars or more, and even a single spectrometer can cost a million or two!

Silver, Sivia, and Pynn (1989) have addressed this question from a heuristic viewpoint and have also suggested a quantitative answer based on elementary signal-to-noise ratio arguments from a power-spectrum error analysis. They posed the following question: Given that the neutron-scattering data are usually a blurred and noisy version of the scattering law we want, what are the optimal characteristics of the instrumental resolution (blurring) function? Conventional wisdom suggests that the most important characteristic of the resolution function is its width: the wider the resolution function, the poorer the quality of the data in the sense that it is more difficult to determine reliably the underlying scattering law. Such thinking is based on a visual, or "what-you-see-is-what-you-get," consideration of the data. A more formal analysis based on statistical inference, or image processing, leads to the conclusion that the overall shape of the resolution function is more important than its width.

We now outline the formal Bayesian approach to the question of instrument design; the algebra is presented in Sivia (1990). We will cast the problem in the same way as did Silver et al. and arrive at the same results; what we add here is the Bayesian rationale for their results. The real advantage of the Bayesian approach is

(a) Inelastic Scattering Data for Iron

Intensity

21.0
19.0
17.0
15.0
13.0
11.0
9.00
7.00
5.00
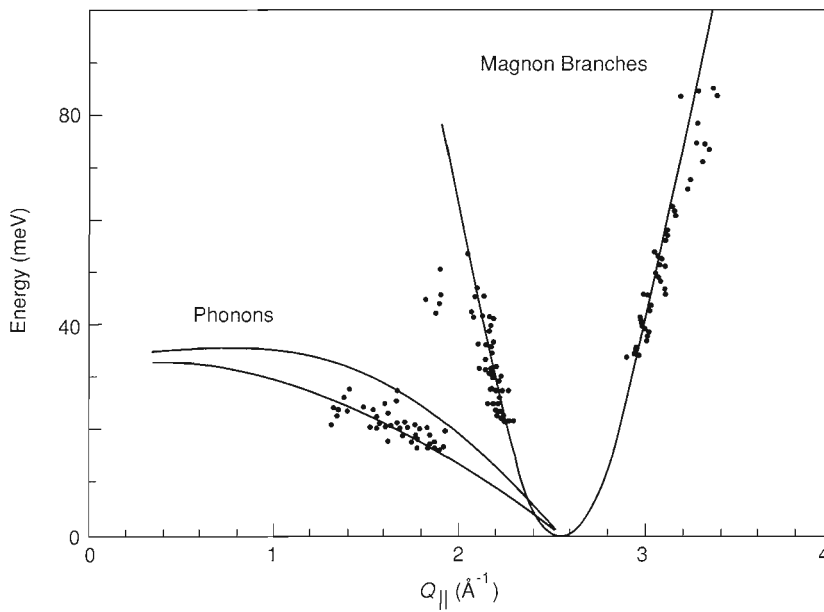3.00
1.00

Detector Angle

Time of Flight
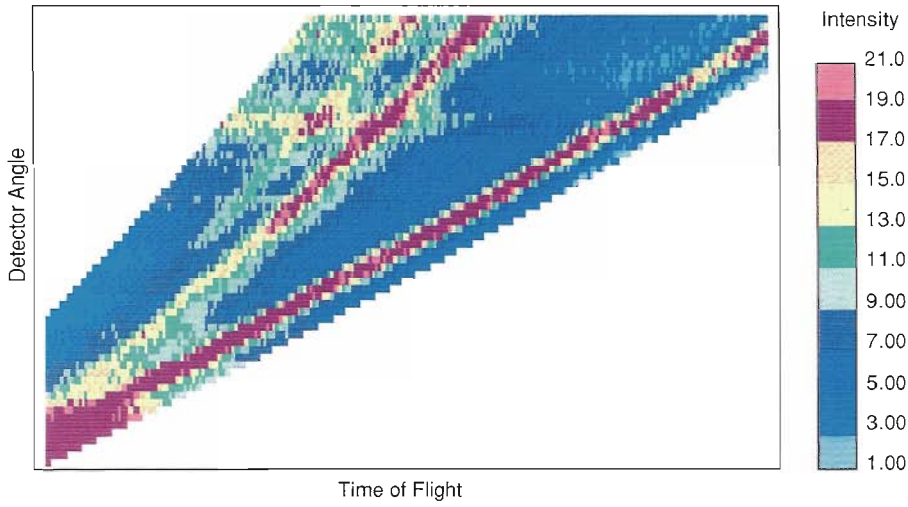
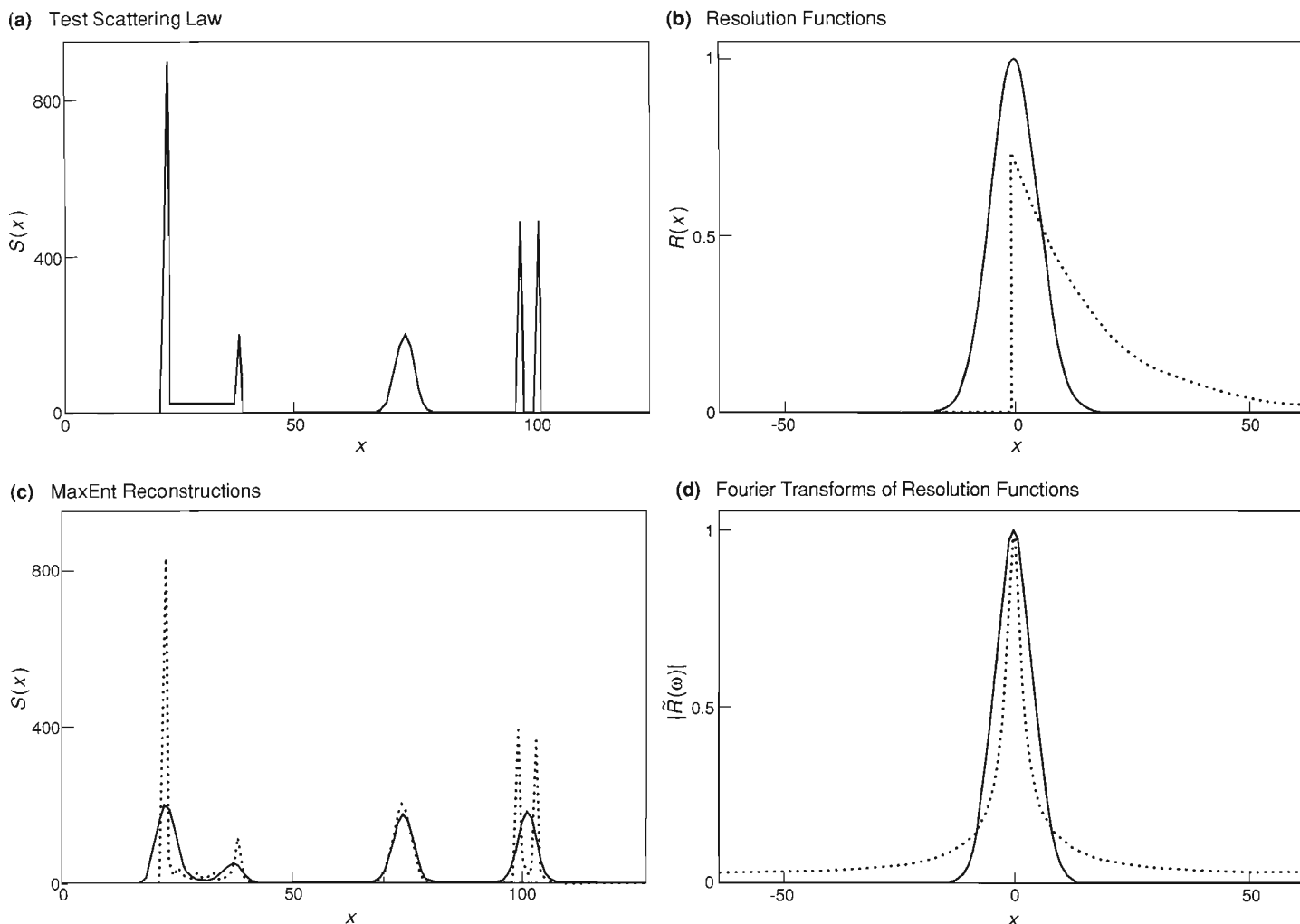## TWO-CHANNEL DECONVOLUTION OF INELASTIC-NEUTRON-SCATTERING DATA

Fig. 13. Application of the two-channel entropy algorithm to the inelastic-neutron-scattering data for iron shown in (a) (Yethiraj et al. 1990) yields the deconvolved "signal" channel shown in (b). Transformation of the lines in (b) to the physically meaningful coordinates of energy transfer and $Q_{parallel}$ (the component of the momentum transfer parallel to the incident neutron beam) reveals both branches of a magnon and some phonons (c).

(b) MaxEnt Reconstruction

Intensity

21.0
19.0
17.0
15.0
13.0
11.0
9.00
7.00
5.00
3.00
1.00

Detector Angle

Time of Flight
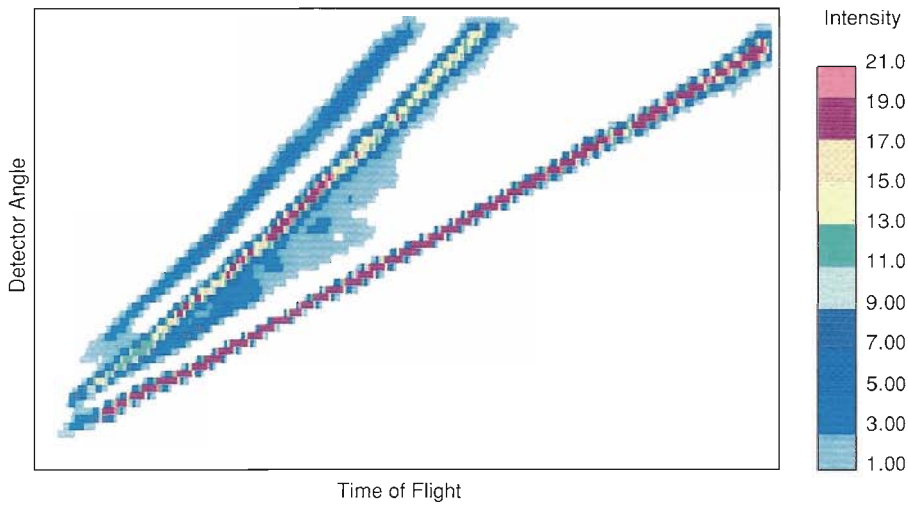
(c) Magnon and Phonon Dispersion Curves

(a) Test Scattering Law

(b) Resolution Functions

(c) MaxEnt Reconstructions

(d) Fourier Transforms of Resolution Functions

## THE FIGURE-OF-MERIT PROBLEM: RESOLUTION FUNCTIONS WITH THE SAME FWHM

Fig. 14. Application of the maximum entropy method to data sets obtained by convolving the test scattering law in (a) with one or the other of the two resolution functions in (b) yields the reconstructions shown in (c). Both resolution functions have the same full width at half maximum (FWHM) and the same integrated intensity and hence have the same conventional figure of merit. Nevertheless, the reconstruction corresponding to the sharp-edged resolution function more nearly matches the original scattering law than does the reconstruction corresponding to the Gaussian resolution function. Also shown, in (d), is the Fourier transform of each resolution function. As discussed in the text, the Fourier transform of a resolution function, and not its full width at half maximum, is most relevant to defining a versatile figure of merit.

its generality; an almost identical analysis can be used to address questions about experimental design in many other contexts (not just convolutions). Moreover, our conclusions are relevant not only to neutron-scattering experiments but also to any other type of experiment involving some element of an instrumental resolution.

The first step in any data analysis is the formulation of the precise question we wish to answer. Formally, we must define the space of possible answers, or choose the hypothesis space. In the case of neutron scattering, we may say that we wish to know the scattering law of our sample, but how is the scattering law to be described? If we know (or assume) that the scattering law consists of a single Lorentzian, for example, then we have a three-dimensional hypothesis space defined by the position, height, and width of the Lorentzian. If, on the other hand, we have no functional form for the scattering law, then we might digitize it into a large number M of pixels, whereupon we have an M-dimensional hypothesis space defined by the flux in each pixel. However, the fact that our best estimate of the scattering law depends not only on the data but also on our choice of hypothesis space limits our ability to provide a universal figure of merit for instrument design. Nevertheless, we will be able to suggest at least a versatile figure of merit, one that is meaningful for many types of problems. But first let's analyze the problem using Bayesian logic.

Once we have chosen the hypothesis space, we can assign a probability distribution over it to indicate our relative beliefs in the various possible scattering laws. The assignment we make before conducting the experiment is, of course, the prior, and Bayes' theorem tells us how the prior is modified by the experimental data, through the likelihood function, to yield the posterior. We also know that the position
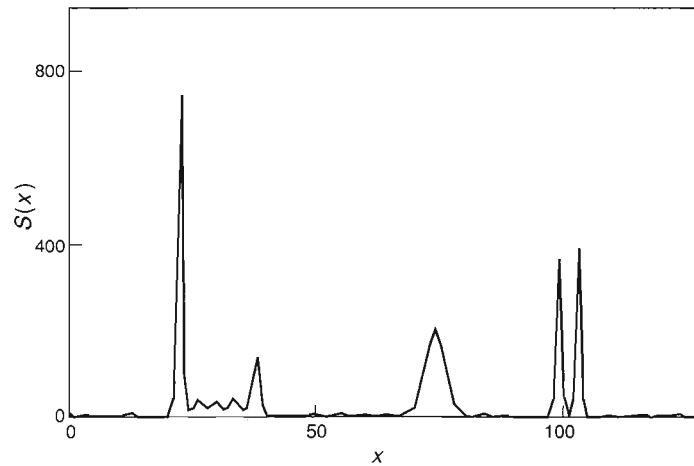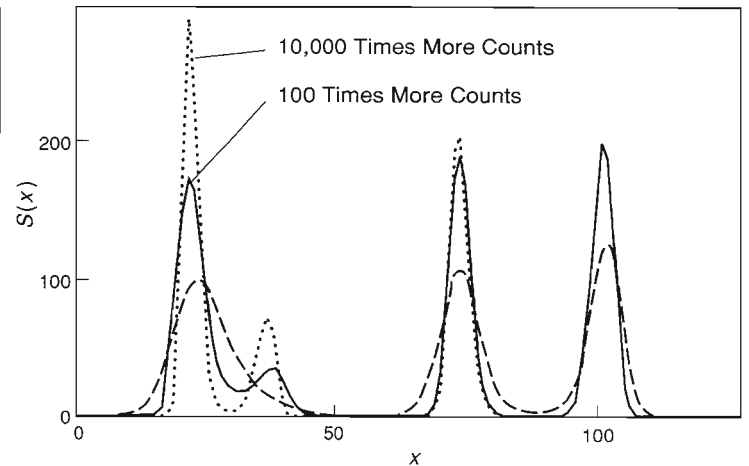
(a) Simulated Data (Narrow Gaussian)
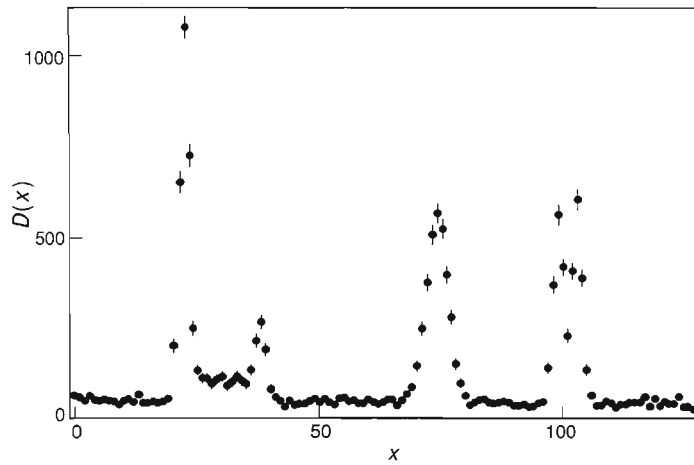
(b) Simulated Data (Wide Gaussian)
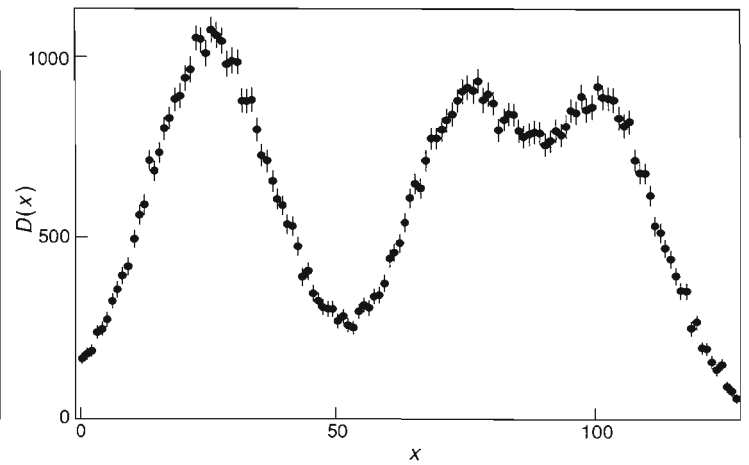
(c) MaxEnt Reconstruction (Narrow Gaussian)

(d) MaxEnt Reconstructrion (Wide Gaussian)

10,000 Times More Counts

100 Times More Counts

of the maximum in the posterior gives us our best estimate of the scattering law and that the width, or spread, of the bump in the posterior around the maximum gives us a measure of the reliability of our estimate. Both, of course, depend on our choice of hypothesis space and on our assignment of the prior probability distribution, but they also depend on the data. The question of how to optimize instrument design can thus be stated as follows: How should we choose the instrumental parameters so that the resulting data give us the most reliable estimate of the scattering law?

Since Bayes' theorem tells us that the data affect our estimate of the scattering law only through the likelihood function, we need to look at its sharpness, or spread. The sharper the likelihood function, the greater the "information content" of the experiment in the sense that the data impose a more severe constraint on what the scattering law could be.

Let us begin by considering a very simple situation. Suppose we know that the scattering law consists of a single delta-function excitation, $A\delta(x - x_0)$, of known magnitude A and unknown postion x_0. In other words, we have a one-dimensional hypothesis space defined by x_0. Suppose also that the experimental data are the result of a convolution between this scattering law and a Gaussian resolution function $T \exp(-x^2/2w^2)$. The height T of this Gaussian resolution function is determined by the length of time for which the data are collected, and its width w is some function of the instrumental parameters, such as flight-path length and collimation angle. The question now is: What restrictions do the data impose on the value of x_0? The width of the likelihood bump, viewed in the one-dimensional space of x_0, gives us the uncertainty in $x_0, \delta x_0$, allowed by the data. After some algebra we find that δx_0

THE FIGURE-OF-MERIT PROBLEM: RESOLUTION FUNCTIONS OF SIMILAR SHAPE

Fig. 15. Shown in (a) and (b), respectively, are noisy data obtained by convolving the scattering law shown in Fig. 14a with a narrow Gaussian resolution function and another Gaussian resolution function ten times wider. Deconvolution of (a) and (b) yields (c) and the dashed curve in (d), respectively. Also shown in (d) are deconvolutions of data sets with 100 times (solid curve) and 10,000 times (dotted curve) the number of counts shown in (b). Contrary to conventional wisdom, increasing the number of data by a factor of 100 does not compensate (in terms of recovering sharp structure) for an increase in FWHM by a factor of 10.

A LIKELIHOOD FUNCTION
IN A TWO-DIMENSIONAL
HYPOTHESIS SPACE

Fig. 16. Suppose that the scattering law of interest is characterized by specific values of only two parameters, A and x_0. (An example of such a scattering law is a delta function of unknown amplitude and position.) The likelihood function $Pr(\{data\}|A, x_0)$ is then a bump in a two-dimensional hypothesis space. Shown here schematically in that hypothesis space are contours along which the logarithm of $Pr(\{data\}|A, x_0)$ is constant. The shape of the likelihood function can be described either by a covariance matrix or by the eigenvectors and eigenvalues of the logarithm of the likelihood function. The elements of the covariance matrix tells us the expected uncertainties allowed by the data in our estimates of A and x_0 and how our estimate of one affects our estimate of the other. The eigenvectors, which specify the directions of the principal axes of the likelihood bubble, tell us which properties, or linear combinations, of A and x_0 can be determined independently. The eigenvalues, which are proportional to the widths of the likelihood function in the directions of the eigenvectors, tell us how reliably each independent property can be estimated.
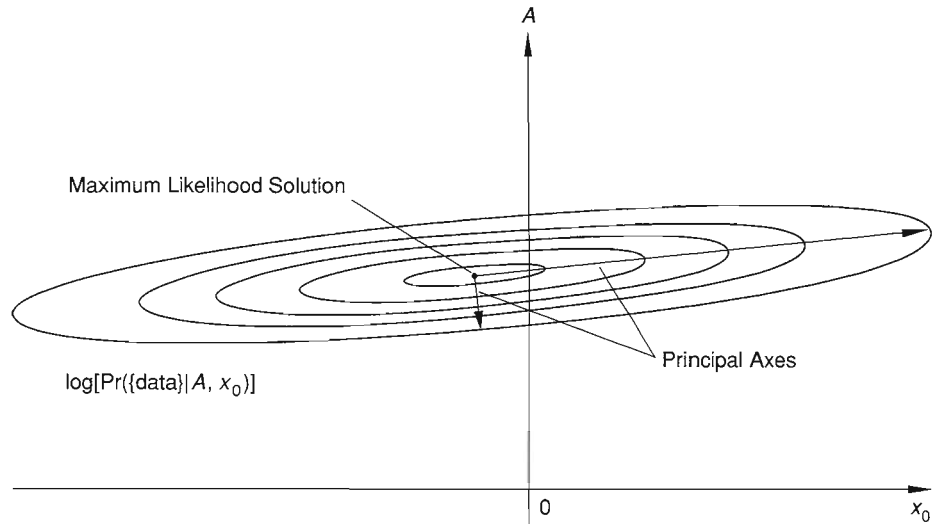
depends on the instrumental design, enshrined in the resolution-function parameters T and w, in the following manner: $\langle(\delta x_0)^2\rangle \propto w/T$. The inverse of this quantity can be used as a figure of merit and has been quoted in the neutron-scattering literature:

$$\text{Conventional Figure of Merit} = \frac{\text{Total Number of Neutrons}}{(\text{FWHM})^2} \propto \frac{T}{w},$$

where FWHM is the full width of the resolution function at half maximum and the total number of neutrons detected is proportional to Tw.

We now show, by means of the examples presented in Figs. 14 and 15, that, although the conventional figure of merit is the correct answer to the question posed above, it is quite unsuitable for general use. Figure 14 presents the MaxEnt reconstructions derived from two data sets obtained by convolving a test object with one or the other of two resolution functions. Even though the resolution functions have identical figures of merit according to the equation above, the reconstruction from the data set obtained by convolution with the sharp-edged resolution function is clearly far superior to the reconstruction from the data set obtained by convolution with the Gaussian resolution function. But the figure of merit above was based on a Gaussian resolution function, you might complain, and so is not valid here. Figure 15 counters that argument by showing the MaxEnt reconstructions derived from two data sets obtained by convolving a test object with one or the other of two Gaussian functions whose FWHMs differ by a factor of 10. According to conventional thinking, the figures of merit can be equalized by increasing the total number of counts for the wide Gaussian by a factor of 100. But Fig. 15 shows instead that, to recover the sharpest features, the number of neutrons counts must be increased by many orders of magnitude!

Next, we move on to consider a slightly more complicated case. Let the situation be exactly the same as before, except that now the scattering law is known to consist of a single delta function of not only unknown position but also unknown magnitude. That is, we have a two-dimensional hypothesis space, defined by the magnitude A and position x_0 of the delta function. Again, we want to know what restrictions the data impose on the value of A and x_0. The likelihood function is now a bump in a two-dimensional space, as illustrated schematically in Fig. 16. To describe the shape of this probability bubble, we need at least three numbers: two for the width in each of the two dimensions and one for the orientation. One way of specifying these numbers is to give the so-called *covariance matrix*, a symmetric 2×2 matrix whose elements tell us the expected uncertainty in the position, $\langle(\delta x_0)^2\rangle$, the expected uncertainty in the magnitude, $\langle(\delta A^2)\rangle$, and how the uncertainty in one af-

fects the uncertainty in the other, $\langle \delta x_0 \delta A \rangle$. After doing some algebra, we find that the correlation term is zero, $\langle \delta x_0 \delta A \rangle = 0$. In other words, the reliability with which we can estimate the position of the delta function has no bearing on the reliability with which we can estimate its magnitude. Thus, in terms of the general schematic picture of Fig. 16, the principal axes of the likelihood probability bubble should lie along the A and x_0 directions. We also find that the instrumental parameters T and w affect the reliability of the inferred magnitude and position of the delta function as follows:

$$\langle (\delta x_0)^2 \rangle \propto w/T \quad \text{and} \quad \langle (\delta A)^2 \rangle \propto 1/Tw.$$

This raises a fundamental question: What do we mean by a figure of merit? The formulae above say that to improve our estimate of the position of the delta function, we should make the width of the Gaussian resolution function as narrow as possible, but to improve our estimate of its magnitude, we should make the resolution function as wide as possible!

We can, of course, keep working through specific problems, but we will only come up with the conclusion that different questions, or different choices of hypothesis space, have different answers. So let us try to ask a *generalized* question. We accept that it will not give the exact answer in every specific case but hope that it will yield a sensible figure of merit for a wide range of situations.

Let us say that the experimental parameters (moderator material, moderator temperature, flight-path length, collimation angle, and so on) all combine to give some resolution function $R(x)$ (not necessarily Gaussian). The question we will ask is: Given that the data are the result of a convolution between the sample scattering law $S(x)$ and the resolution function $R(x)$, how reliably can we estimate the scattering law assuming no particular functional form for $S(x)$?

Since we do not have a functional form for the scattering law, as we did before, an obvious hypothesis space to choose is the one defined by the values of $S(x)$ specified on a grid finely digitized in x. That is, we have an M-dimensional hypothesis space, where M is very large. The likelihood function is now a bump in a multi-dimensional space, and we can consider Fig. 16 as a schematic two-dimensional slice through that space if the axis labels are changed to read $S(x_i)$ and $S(x_j)$ instead of A and x_0. The spread of this multi-dimensional probability bubble about its maximum will, of course, give us a measure of how well the data constrain the permissible scattering laws. However, since the likelihood bubble is, in general, skew with respect to our $\{S(x_j)\}$ axes, its width is difficult to describe. It is convenient, therefore, to rotate our coordinate axes from the original $\{S(x_j)\}$ axes to another set of axes that lie along the principal axes of the probability bump; the spread of the bubble is then given simply by its widths along the new coordinate axes. These principal axes are vectors in the coordinates $\{S(x_j)\}$ and hence represent relative pixel heights in our digitized x coordinate—they are discretized functions of x. Formally, the principal axes are called *eigenvectors* or, if we go to the continuum limit by making the digitized grid infinitesimally fine, *eigenfunctions* .

The eigenfunctions define the natural hypothesis space for our problem because they represent the properties of the scattering law that can be estimated independently of each other. If we write the required scattering law as a linear combination of the eigenfunctions, $S(x) = \sum a_j \eta_j(x)$, where $\eta_j(x)$ are the eigenfunctions and a_j are coefficients (or parameters) that are now to be determined from the data, then we find that the reliability of our estimate of one parameter does not affect the reliability of our estimate of another; that is, the covariance matrix is diagonal ($\langle \delta a_i \delta a_j \rangle = 0$). The widths of the likelihood function along the principal directions, δa_j, tell us the reliability with which the eigenfunction properties of the scattering law can be estimated; the widths are related to the so-called *eigenvalues* λ by $\langle (\delta a_j)^2 \rangle = 2/\lambda_j$.

RESPONSE MATCHING

Fig. 17. The resolution function $R(x)$ appropriate to a neutron-scattering experiment at a spallation source is a convolution ($\otimes$) of two response functions: the moderator line shape and an instrumental contribution. The moderator line shape (the time spectrum of neutrons exiting the moderator) has a sharp leading edge, which can be regarded as the rising edge of a narrow Gaussian with a FWHM, $\sigma_{moderator}$, determined by the moderator material, and a long tail that decays roughly exponentially with a decay constant τ determined by the "poison" added to the moderator. The instrumental contribution is roughly Gaussian with a FWHM of $\sigma_{instrument}$. The analysis presented in the text indicates that $\sigma_{instrument}$ should probably be matched to $\sigma_{moderator}$ (rather than to the FWHM of the moderator line shape as a whole) to obtain the "best" $R(x)$.

$$R(X) =$$

exp(-x/τ)

$\sigma_{moderator}$

$\otimes$

$\sigma_{instrument}$

Moderator Line Shape   *x*   Instrumental Contribution   *x*
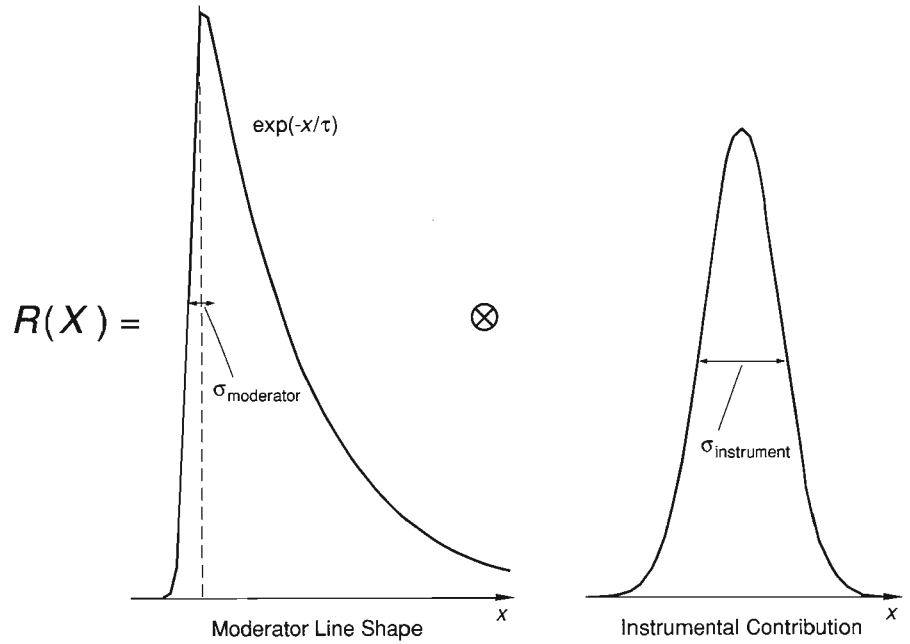
If we were to carry out the algebra for our problem, making suitable (usually reasonable) assumptions to obtain an analytic solution, we would find that the eigenfunctions $\eta_\omega(x)$ and their corresponding eigenvalues λ_ω are given by

$$\eta_\omega(x) = \cos(\omega x) \text{ and } \sin(\omega x)$$

$$\text{and} \quad \lambda_\omega = \frac{2}{\sigma^2}|\tilde{R}(\omega)|^2,$$

where $\tilde{R}(\omega)$ is the Fourier transform of the resolution function $R(x)$ and σ^2 is a measure of the average number of counts in the data. This solution tells us that if we do not have a functional form for the scattering law, then we should express it in terms of a Fourier series (a sum of sine and cosine functions). The advantage of doing so is that the reliability with which we can estimate one Fourier coefficient will not affect the accuracy with which we can determine another—it is an uncorrelated space. Since the reliability with which we can estimate any Fourier coefficent is inversely proportional to the corresponding eigenvalue, $\langle(\delta a_\omega)^2\rangle = 2/\lambda_\omega$, we can use λ_ω as a figure of merit for inferring structure in the scattering law with detail $\delta x \approx 1/\omega$.

The implications of this analysis for instrument design are as follows.

- A versatile figure of merit depends largely on the Fourier transform of the resolution function rather than on its full width at half maximum. This result is illustrated in Fig. 14: The two resolution functions in Fig. 14b have the same full width at half maximum and the same integrated intensity, but, as shown in Fig. 14d, the Fourier transform of the one with the sharp edge does not decay as rapidly with increasing frequency ω as the Fourier transform of the Gaussian resolution function. Resolution functions that have sharp features, therefore, allow high-frequency information to be recovered reliably from the data. An electrical engineer would say that the figure of merit is governed by the *bandwidth* of the resolution function.
- The figure of merit for a given resolution function is not constant but depends on the amount of detail required in the inferred scattering law.
- The background signal has not been forgotten; it enters the figure of merit through the dependence on average number of counts, or σ^2. Any long decaying tail of the resolution function reduces the figure of merit in the same way that the background does, since such a tail adds to the average number of counts but does not contribute to the Fourier term $\tilde{R}(\omega)$ at high frequency.

Since the resolution function in neutron scattering depends on details of the spectrometer and moderator, our results suggest a potential revision of ideas on the design of neutron-scattering facilities. Take, for example, the *matching* of resolution elements on a neutron spectrometer at an accelerator-based source, which is illustrated in Fig. 17. The resolution function for an experiment is the resultant of a convolution between a roughly Gaussian instrumental contribution (flight-path length, collimation angle, and so on), and the moderator line shape (the time spectrum of the pulse of neutrons leaving the moderator). The moderator line shape has a sharp rising edge, the sharpness of which is governed by the moderator material, and a long decaying tail, the decay of which is governed by the "poison" added to the moderator. The question is how to choose the width of the instrumental component so as to get the "best" resultant resolution function. Conventional wisdom recommends that we should make the width of the Gaussian-like instrumental contribution comparable to the width of the moderator line shape. The analysis above, however, suggests that following this advice could seriously impair our ability to infer (reliably) the scattering law at high resolution and that we should probably match the width of the instrumental component to the narrow width of the sharp leading edge of the moderator line shape. How such considerations translate into the optimal choice of collimation angle, of flight-path length, of moderator material, and of a moderator "poison" is the subject of ongoing research. ∎

Further Reading

Bayesian Methods: Historical Background, Fundamentals, and Tutorials

T. Bayes. 1763. *Philosophical Transactions of the Royal Society of London*, 330–418.

P. S. de Laplace. 1812. *Theorie Analytique des Probabilités*. Paris.

Harold Jeffreys. 1939. *Theory of Probability*. Oxford: Oxford University Press. (A paperback version of the third (1961) edition of this work was issued in 1983.)

R. T. Cox. 1946. Probability, frequency and reasonable expectation. *American Journal of Physics* 14: 1–13.

E. T. Jaynes. 1983. *E. T. Jaynes: Papers on Probability, Statistics and Statistical Physics*, edited by R. D. Rosenkrantz. Dordrecht, The Netherlands: D. Reidel Publishing Company. (Reprinted in 1989 as a Pallas Paperback by Kluwer Academic Publishers, Dordrecht, The Netherlands.)

E. T. Jaynes. 1986. Bayesian methods: An introductory tutorial. In *Maximum Entropy and Bayesian Methods in Applied Statistics*, edited by J. H. Justice. Cambridge: Cambridge University Press.

Maximum Entropy

C. E. Shannon. 1948. A mathematical theory of communication. *Bell System Technical Journal* 27: 379–423 and 623–656.

S. F. Gull and G. J. Daniell. 1978. Image reconstruction from incomplete and noisy data. *Nature* 272: 686–690.

John E. Shore and Rodney W. Johnson. 1980. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Transactions on Information Theory* IT-26: 26–37.

S. F. Gull and J. Skilling. 1984. Maximum entropy method in image processing. *IEE Proceedings* 131, Part F: 646–659.

J. Skilling and R. K. Bryan. 1984. Maximum entropy image reconstruction: General algorithm. *Monthly Notices of the Royal Astronomical Society* 211: 111–124.

Some Recent Developments on Maximum Entropy

Stephen F. Gull. 1989. Developments in maximum entropy data analysis. In *Maximum Entropy and Bayesian Methods (Cambridge, 1988)*, edited by J. Skilling. Dordrecht, The Netherlands: Kluwer Academic Publishers.

John Skilling. 1989. Classic maximum entropy. In *Maximum Entropy and Bayesian Methods (Cambridge, 1988)*, edited by J. Skilling. Dordrecht, The Netherlands: Kluwer Academic Publishers.
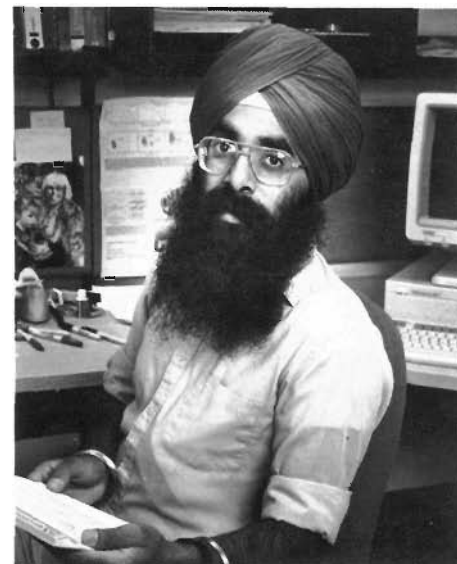
Devinderjit Singh Sivia has been a postdoctoral fellow at Los Alamos National Laboratory since 1988, holding a joint appointment with the Theoretical Division and the Manual Lujan, Jr. Neutron Scattering Center. He received his B.A. in the Natural Sciences Tripos at Cambridge University, in 1984, and continued his studies there with the Radio-Astronomy Group at the Cavendish Laboratory, receiving his Ph.D. in 1988. He has been working on the applications of maximum entropy and Bayesian methods, both as a graduate student and a postdoctoral fellow, in a wide variety of fields including x-ray crystallography, very-long-baseline interferometry, "phaseless holography," neutron scattering, medical imaging, and theoretical condensed-matter physics.

Acknowledgments

R. K. Bryan. 1990. Solving oversampled data problems by maximum entropy. In *Maximum Entropy and Bayesian Methods (Dartmouth College, 1989)*, edited by P. Fougere. Dordrecht, The Netherlands: Kluwer Academic Publishers.

S. Sibisi. 1990. Quantified MaxEnt: An NMR application. In *Maximum Entropy and Bayesian Methods (Dartmouth College, 1989)*, edited by P. Fougere. Dordrecht, The Netherlands: Kluwer Academic Publishers.

J. Skilling. 1990. Quantified maximum entropy. In *Maximum Entropy and Bayesian Methods (Dartmouth College, 1989)*, edited by P. Fougere. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Examples of the Use of MaxEnt at LANSCE and Related Papers

B. N. Brockhouse, H. E. Abou-Helal, and E. D. Hallman. 1967. Lattice vibrations in iron at $296°$K. *Solid State Communications* 5: 211–216.

J. W. Lynn. 1975. Temperature dependence of the magnetic excitations in iron. *Physical Review B* 11: 2624–2637.

A. D. Taylor, E. J. Wood, J. A. Goldstone, and J. Eckert. 1984. Lineshape analysis and filter difference method for a high intensity time-of-flight inelastic neutron scattering spectrometer. *Nuclear Instruments and Methods in Physics Research* 221: 408–418.

T. J. Newton. 1985. Blind deconvolution and related topics. Ph.D. thesis, Cambridge University.

R. A. Robinson, R. Pynn, and J. Eckert. 1985. An improved constant-Q spectrometer for pulsed neutron sources. *Nuclear Instruments and Methods in Physics Research* A241: 312–324.

F. Mezei and P. Vorderwisch. 1989. Spectroscopy with asymmetric resolution functions: Resolution improvement by an on-line algorithm. *Physica B* 156 and 157:678.

P. A. Seeger, R. P. Hjelm, Jr., and M. J. Nutter. 1989. The low-Q diffractometer at the Los Alamos Neutron Scattering Center. *Molecular Crystals and Liquid Crystals* 180A: 107–117.

R. P. Hjelm, P. Thiyagarajan, D. S. Sivia, P. Linder, H. Alken, D. Schwahn. 1990. Small-angle neutron scattering from aqueous mixed colloids of lecithin and bile salts. Accepted for publication in *Colloid and Polymer Science*.

Devinderjit Singh Sivia. 1990. Applications of maximum entropy and Bayesian methods in neutron scattering. In *Maximum Entropy and Bayesian Methods (Dartmouth College, 1989)*, edited by P. Fougere. Dordrecht, The Netherlands: Kluwer Academic Publishers.

D. S. Sivia, P. Vorderwisch, and R. N. Silver. 1990. Deconvolution of data from the filter difference spectrometer: From hardware to maximum entropy. Accepted for publication in *Nuclear Instruments and Methods in Physics Research*.

M. Yethiraj, R. A. Robinson, D. S. Sivia, J. W. Lynn, and H. A. Mook. A neutron scattering study of the magnon energies and intensities in iron. Submitted to *Physical Review B*.

Optimal Instrument Design

A. Michaudon. 1963. The production of moderated neutron beams from pulsed accelerators. *Journal of Nuclear Energy Parts A/B* 17: 165–186.

D. H. Day and R. N Sinclair. 1969. Neutron moderator assemblies for pulsed thermal neutron time-of-flight experiments. *Nuclear Instruments and Methods* 72: 237–253.

C. G. Windsor. 1981. *Pulsed Neutron Scattering*. London: Taylor and Francis Ltd.

R. N. Silver, D. S. Sivia, and R. Pynn. 1989. Information content of lineshapes. In *Advanced Neutron Sources 1988*, edited by D. K. Hyer. Bristol: Institute of Physics.

D. S. Sivia, R. N. Silver, and R. Pynn. 1990. Optimization of resolution functions for neutron scattering. *Nuclear Instruments and Methods in Physics Research* A287: 538–550.

The quotation that opens this article appears in *Statistics for Nuclear and Particle Physicists* by Louis Lyons (Cambridge University Press, 1986).