

THE PHASE PROBLEM OF X-RAY CRYSTALLOGRAPHY

For almost 40 years physicists thought it impossible, even in principle, to determine complex crystal structures directly from the intensities of diffracted x rays.

Herbert A. Hauptman

With the invention some 200 years ago of the goniometer, an instrument for measuring the angles between the faces of a crystal, the science of crystallography was born. The goniometer made possible the discovery of the fundamental laws of descriptive crystallography: that the angles between the facial planes are determined by the chemical composition of the crystal and that the relative orientations of the facial planes follow a simple rule, the law of rational positions.

Wilhelm Röntgen's discovery of x rays in 1895 set the stage for the creation of the modern science of x-ray crystallography. In 1912 Paul Ewald, working in Munich under the direction of Arnold Sommerfeld, completed his doctoral dissertation on the optical properties of a medium consisting of a regular anisotropic arrangement of isotropic resonators. Because a crystalline solid (such as that shown in figure 1) may be regarded as a regular, triply periodic arrangement of atoms or molecules (like the one in figure 2), it serves as a realization of Ewald's abstraction, an anisotropic arrangement of isotropic resonators. When Max von Laue learned of Ewald's dissertation it occurred to him that a crystal might serve as a three-dimensional diffraction grating for x rays, because the smallest interatomic distances are of the same order of magnitude as x-ray wavelengths. He thereupon persuaded Walter Friedrich and Paul Knipping to perform a scattering experiment of the sort shown schematically in figure 3.

The experiment demonstrated that when x rays strike a crystal, they do indeed scatter, and will blacken a photographic plate according to their directions and

intensities. In this way one obtains the crystal's diffraction pattern; a typical example is shown in figure 4. This fundamental experiment marked the birth of x-ray crystallography, and because of that technique's essential role in the determination of crystal and molecular structures, the experiment must be regarded as a watershed event in modern science.

In this article I give a brief historical account of the phase problem of x-ray crystallography, which forms the major obstacle in the path leading from the observable diffraction pattern to the desired crystal structure.

Key to molecular structures?

Figure 2 shows how a crystal may be regarded on the submicroscopic level as a regular, triply periodic arrangement of an array of atoms. Three families of parallel planes decompose the crystal's space, the planes within each family being equidistant from one another. Any one of the resulting collection of congruent parallelepipeds is called a fundamental parallelepiped, or unit cell, of the crystal.

If one imagines each unit cell to be filled with atoms that collectively constitute a molecule and are arranged in precisely the same way in all the unit cells, then each unit cell and its contents are indistinguishable from every other unit cell and its contents.

Associated with each atom is an electron density function, so that the distribution of the atoms in the unit cell of a crystal yields, by the superposition of the individual atomic density functions, an overall electron density function $\rho(\mathbf{r})$. This is a nonnegative function that gives the number of electrons per unit volume at the position $\mathbf{r}$. Owing to the geometric construction, it is clear that the electron density function in any unit cell is identical to that in every other unit cell, so that $\rho(\mathbf{r})$ is a

Herbert Hauptman is president of the Medical Foundation of Buffalo, in Buffalo, New York.



Crystals of butein, $C_{15}H_{12}O_5$, at $63\times$ magnification. The precise geometry, or arrangement, of atoms in the butein molecule was determined by the direct x-ray diffraction techniques described in the text. **Figure 1**

triply periodic function of position. This may be taken as the mathematical definition of a crystal.

Alternatively, one may regard a crystal as a triply periodic arrangement of an array of atoms or molecules. With this interpretation one means by a crystal structure simply the arrangement and identities of the atoms in the unit cell, and by a molecular structure the arrangement and identities of the atoms in the molecule.

It was recognized almost from the beginning that a crystal's structure uniquely determines its diffraction pattern—that is, the directions and intensities of the x rays that it scatters—and that one could calculate the diffraction pattern completely if one knew the crystal structure. As it turns out—although this was not known until many years later—the converse is also true: In general, diffraction patterns define unique crystal and molecular structures. In other words, the information content of a typical molecular structure coincides precisely with the information content of its diffraction pattern. It is a major triumph of modern x-ray crystallography that it permits one routinely to transform a diffraction pattern such as the one in figure 4 into a structure such as that in figure 5, provided the molecule contains fewer than about 150 nonhydrogen atoms.

X rays, it should be recalled, are electromagnetic waves, and have not only an intensity but also a phase, like any other wave disturbance. As it turns out, to work backwards from diffraction patterns to crystal and molecular structures it is necessary to measure not only the intensities but also the phases of the x rays scattered by the crystal.

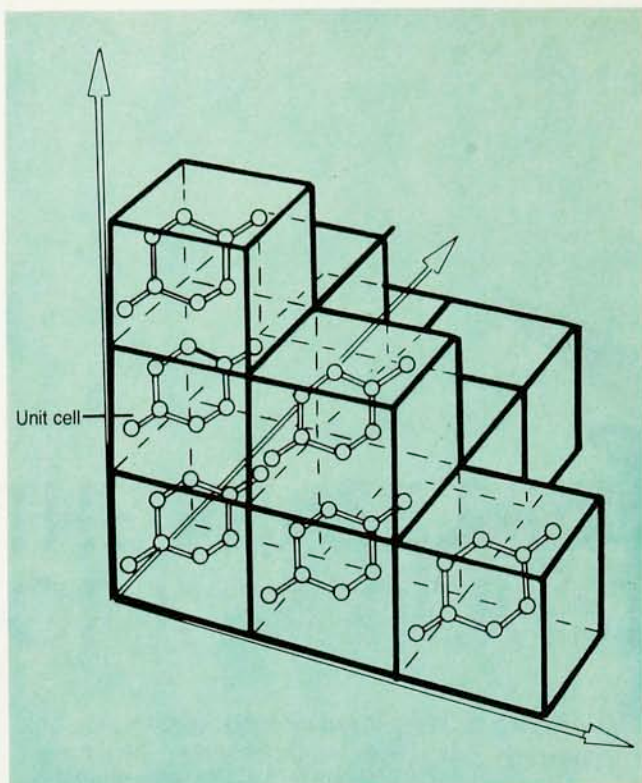
Unfortunately, one can measure only the intensities of the scattered x rays. The phases are lost in the usual kind of diffraction experiment. Nevertheless, immediately following Friedrich and Knipping's experiment, Wil-

liam and Lawrence Bragg, father and son, were able to determine the arrangement of atoms in a few simple substances such as NaCl, KCl and KBr by interpreting their diffraction patterns as due to specular reflection from the atomic planes within the crystals. Despite this initial success, all attempts to find a general method for going directly from the diffraction pattern to the crystal structure—a method that would be useful for the more complex structures of interest to chemists, biologists and mineralogists—were defeated for almost 40 years.

In fact, because the needed phase information was lost in the diffraction experiment, it was thought that one could use arbitrary values for the phases associated with the measured intensities of the scattered x rays. This approach yielded a myriad of different crystal structures, all consistent with the known intensities. It therefore came to be generally believed that a procedure for going directly from the measured intensities to crystal structures could not, even in principle, be devised. By the same thinking, the problem of deducing the values of the individual phases from the diffraction intensities, the so-called phase problem, was unsolvable, even in principle. It wasn't until the early 1950s, through the exploitation of special properties of molecular structures and through a simple mathematical argument, that these erroneous conclusions were finally refuted.

The phase problem

The special properties that all crystal and molecular structures possess may be summed up in one word: atomicity. Thus the electron density function $\rho(\mathbf{r})$ in a molecule not only is nonnegative everywhere but takes on large positive values at the atomic position vectors and drops to small values between atoms. If our goal is merely to determine the positions of the atoms—that is, the



Representation of a crystal. A crystalline solid may be regarded as a regular, triply periodic array of atoms or molecules. **Figure 2**

positions of the maxima of $\rho(\mathbf{r})$ —rather than the much more complicated electron density function associated with the distribution of atoms in the crystal, then our problem is greatly simplified; it turns out to be not only determinate but actually greatly overdetermined by the available x-ray diffraction intensities.

This is most easily seen by eliminating the lost phase information from the relationships between the diffraction pattern and the crystal structure. Doing this results in a system of equations relating the diffraction intensities alone with the atomic position vectors. Because the number of these relationships far exceeds (by a factor of ten or so) the number of unknown position vectors needed to define the crystal structure, our problem is greatly overdetermined. Thus it is clear that there exist relationships between the measured diffraction intensities and the lost phases that may be exploited. It follows that the phases of the scattered x rays are also determined by their intensities. In short, the lost phase information is to be found among the available intensities, and the phase problem is therefore a solvable one, at least in principle. There remains the task of devising numerical algorithms leading from the abundance of experimentally measured diffraction intensities to the values of the individual phases. The techniques of x-ray crystallography that deduce the individual phases by exploiting relationships between measured diffraction intensities and phases are known as direct methods.

The argument just presented was in fact anticipated in 1927 by Heinrich Ott,¹ who showed by algebraic analysis and applications that the method is capable of solving simple centrosymmetric structures, in which all phases must be either 0 or π . The method was further elaborated by Kedaraswar Banerjee² in 1933 and Melvin Avrami³ in 1938 but was clearly of only limited value in

applications. While this early work of Ott, Banerjee and Avrami shed important light on the more general phase problem, it attracted little attention at the time and was not further developed; it appears now to be all but forgotten.

The beginnings

In the fall of 1947 I joined the Naval Research Laboratory and started my collaboration with Jerome Karle. Although the group to which I was assigned was called the electron diffraction section, and my initial task was to design and construct a slow-electron diffraction instrument to study the interaction of low-energy electrons with gases, within a year the lure of the x-ray crystallographic phase problem became irresistible. It had been some 35 years since Friedrich and Knipping had carried out their landmark experiment, and by 1947 the phase problem, the central problem of x-ray crystallography, was generally regarded as unsolvable. The central importance of this problem and its strong mathematical component combined to provide a challenge that could not be denied.

Then too, there was a certain air of mystery surrounding the problem. On the one hand the simplicity and logic of the argument "proving" its unsolvability, even in principle, appeared to be overwhelming. On the other hand crystal and molecular structures were being solved, although the structures studied were almost always very simple ones involving a small number of atoms or larger structures containing one or a small number of heavy atoms, for which special techniques had been devised. It had not yet been generally understood that the assumption of atomicity and the concomitant trial-and-error approach to most structure solutions had imposed a powerful restriction on the permitted values of the phases.

Nonnegativity. The first important contribution that Karle and I made was the recognition that it would be necessary to exploit prior structural knowledge to transform the phase problem from an unsolvable one to one that was solvable, at least in principle. Our first step in this direction was to exploit the nonnegativity of the electron density function $\rho(\mathbf{r})$. Before our analysis was complete, however, David Harker and John Kasper published their famous paper in which they derived inequalities in which the measured intensities restrict the possible values of the phases.⁴ This was a very mysterious paper, because nowhere in it does there appear any explicit mention of the basis for the inequality relations, and indeed the most important fact is conspicuous by its absence. It is simply

$$\rho(\mathbf{r}) \geq 0$$

That is to say, the electron density function is nonnegative everywhere. This fact is, however, implicit in Harker and

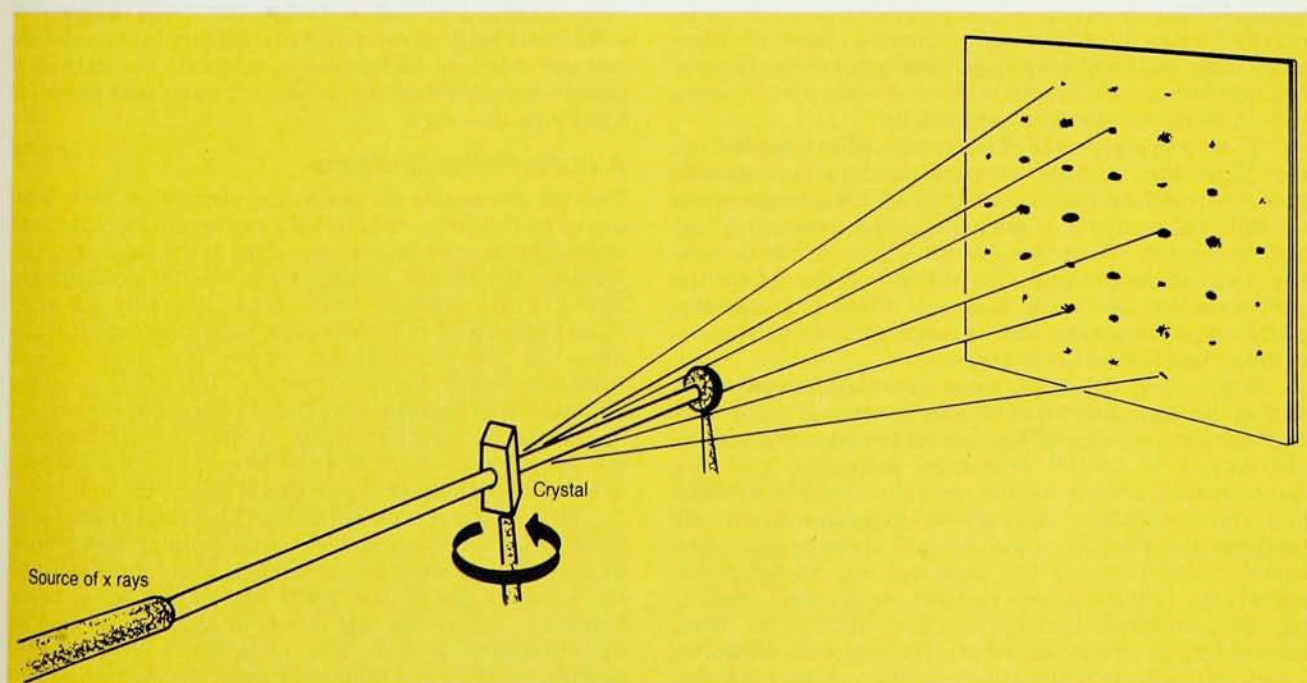
Kasper's work. Needless to say, when Karle and I discovered their omission we were very excited and delighted; it meant that we were on the right track! In very short order we completed our own analysis and derived the complete set of inequality relationships based on the nonnegativity of the electron density function.⁵ It includes the Harker-Kasper inequalities as a special case, and many others besides. Although the complete set of inequalities greatly restricts the values of the phases, the relations appear to be too intractable to be useful in applications, except for the simplest structures, and their potential has never been fully exploited.

Atomicity. The recognition in 1950 and 1951 that molecules consist of atoms that to a good approximation may be regarded as points transformed completely the nature of the phase problem. While it meant accepting as fact that the observed diffraction intensities by themselves were indeed not sufficient to determine a unique electron density function, it also meant that they were more than sufficient, by far, to determine the atomic position vectors.⁶ It meant as well that the phases corresponding

to the point atom structure were greatly overdetermined by the available intensities. Finally, it meant that a formidable psychological barrier had been removed, because it now made sense to look for a solution to the phase problem, that is, for numerical algorithms leading from measured intensities to individual phases. Our next task was to lay the mathematical foundation for the derivation of such procedures. In hindsight it is perfectly clear that owing to the great overabundance of diffraction data, a probabilistic approach is called for; some 40 years ago, however, this was not so apparent.

The value of probability

Before we could even get started, an unexpected complication arose. It turned out that because the values of the individual phases clearly depend on the atomic position vectors, they depend also on the choice of origin and in general are not uniquely determined by the crystal structure alone. It followed that the diffraction intensities alone do not determine unique values for the phases. The process leading from diffraction intensities to phases



Scheme for observing the scattering of x rays by a crystal. In the experiment proposed by Max von Laue and performed by Walter Friedrich and Paul Knipping the crystal was not rotated; instead, nonmonochromatic x rays were used. **Figure 3**

would have to include a recipe for specifying the origin. This required that we separate out two contributions to a phase, one due to the crystal structure alone and one due to the choice of origin. We clearly needed to study how a phase is transformed when the origin is shifted, a problem that was complicated by the fact that the permissible origins depend on the crystallographic elements of symmetry, which were usually known in advance.

The solution was made easier by the discovery that there are always certain linear combinations of the phases, the so-called structure invariants, that are uniquely determined by the crystal structure alone and are independent of the choice of origin. It is therefore only the values of the structure invariants that we can hope to estimate from the measured intensities. Once we have estimated a sufficient number of these we can then hope to evaluate the individual phases by a process that incorporates a recipe for specifying the origin.

What was clearly called for was devising a method for identifying the structure invariants, and then using these to come up with recipes for fixing the origin appropriate to the different elements of crystallographic symmetry that may be present. Once this was done there would remain the task of estimating the values of the structure invariants by means of their conditional probability distributions, assuming that an appropriately chosen set of diffraction intensities is known.

Beyond any doubt our most important contribution during the early 1950s was the introduction of probabilistic techniques—in particular, use of the joint probability distribution of several diffraction intensities and the corresponding phases—as the central tool in the solution of the phase problem.⁷ We assumed to begin with that all positions of the atoms in the unit cell of the crystal were equally likely, or, in the language of mathematical probability, that the atomic position vectors were random variables, uniformly and independently distributed. With this assumption the intensities and phases of the scattered x rays as functions of the atomic position vectors are also random variables, and one can use the methods of modern mathematical probability theory to calculate the joint probability distribution of any collection of intensities and phases. A suitably chosen joint probability distribution



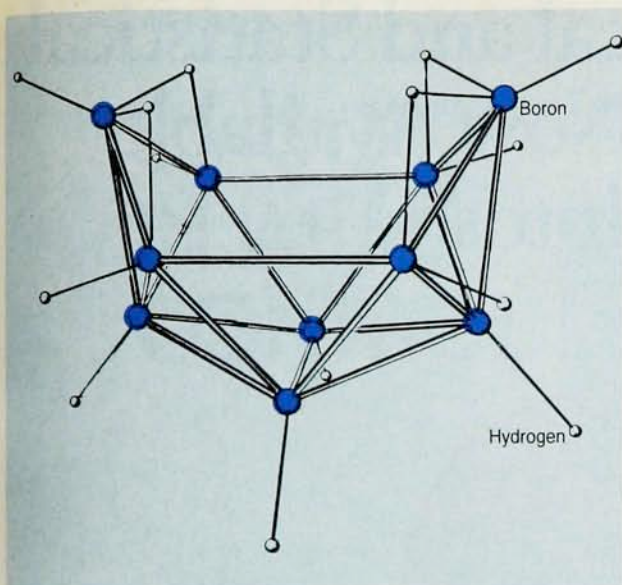
Diffraction pattern of $C_{11}H_9BrN_2O$ crystals. X rays scattered from a crystal blacken a photographic plate according to their directions and intensities. **Figure 4**

leads directly to the conditional probability distribution of a specified structure invariant, assuming again an appropriately chosen set of diffraction intensities. The conditional distribution in turn leads to the structure invariant; one estimates, for example, the invariant's most probable value. Once one has a sufficiently large number of sufficiently reliable estimates of structure invariants, one can use standard techniques to calculate the individual phases, provided that the process incorporates a recipe for specifying the origin.

A noteworthy structure

Tens of thousands of molecular structures have been solved by the direct methods of x-ray crystallography, and about 5000 new structures are added to the list every year. Whereas structures having some 50–100 nonhydrogen atoms in the molecule were rather routinely solved by these methods 10 or 15 years ago, molecules containing as many as 150 nonhydrogen atoms are more or less routinely solved today. However, the solution of the structure of gramicidin A by David Langs, a leading expert in the use of direct methods, is undoubtedly the most spectacular to date: It required determining the positions of 315 independent nonhydrogen atoms in the unit cell.⁸

Gramicidin A is an antibiotic that attacks bacteria by forming ion channels in their outer cellular membranes. These channels allow potassium ions to enter the cell, and the bacteria die in the futile attempt to expel them. However, despite the best efforts of scientists to gain an understanding of how this molecule, a string-shaped peptide, could form a tubular channel through which ions can flow, no plausible, generally accepted mechanism could be devised. Investigators made many attempts to determine its structure and proposed models to explain its activity, but without success. A plausible explanation for



Molecular structure of decaborane, $B_{10}H_{14}$, as determined by x-ray diffraction. **Figure 5**

the mechanism of action of gramicidin A based on its solid-state structure has in fact been formulated, but more work needs to be done before a definitive answer to this question can be given. Now that the solid-state structure of the molecule is known it may be possible to improve the action of this drug by suitable chemical modification. More generally, knowledge of how the gramicidin channel functions provides a framework for understanding how biological channels in living cells are controlled and regulated.

What mathematicians should know

The system of equations that describes the relationship between a crystal's structure and its diffraction pattern is simply

$$|E_{\mathbf{H}}| = \sigma_2^{-1/2} \left| \sum_{j=1}^N Z_j \exp(2\pi i \mathbf{H} \cdot \mathbf{r}_j) \right|$$

where

$$\sigma_2 = \sum_{j=1}^N Z_j^2$$

The factor $\mathbf{H}$ is an ordered triplet of integers that labels a scattered ray; $|E_{\mathbf{H}}|$, the magnitude of the complex normalized structure factor, is simply related to the intensity of the corresponding scattered ray and may therefore be presumed to be known for each $\mathbf{H}$; Z_j is the atomic number of the atom labeled j and is therefore also known; N is the number of atoms in the unit cell; and $\mathbf{r}_j$ is the position vector of the atom labeled j . Hence in this system of equations there is one equation for each scattered x ray whose intensity has been measured. The unknowns are the N atomic position vectors $\mathbf{r}_j$, each having three components; all other factors in the equation are known. Thus the problem of determining crystal structures from their diffraction patterns may be formulated in mathematical terms as the problem of solving this redundant system of transcendental simultaneous equations—redundant because in general there are far more equations, by a factor of ten or so, than there are unknowns.

While this mathematical formulation is important in showing that the crystal structure problem is in fact overdetermined, it throws little light on the solution itself

because of the complexity of the equations. Yet crystal structures are determined every day, and each time this redundant system of transcendental simultaneous equations is also solved. This fact is a well-kept secret; most mathematicians are unaware of it. It should be more widely known. After all, many mathematicians would like to learn more about a method that solves a system of several thousand equations in several hundred unknowns some 10 or 15 times a day. The methods may have more general applicability.

What next?

In 1971 my friend Michael Woolfson predicted that by 1980 structures of molecular weight 5000 would be routinely solvable by the direct methods and that some structures of molecular weight 10 000 would have been solved. It is clear now that these predictions were too optimistic; by 1989 neither one was realized.

Nevertheless, we have seen slow but steady progress over the years, and major advances since 1971. First, the theoretical underpinnings of the direct methods, based on the initial probabilistic formulation of the early 1950s, have been considerably broadened and have led to a strengthening of these methods. Second, clever and powerful computer packages incorporating the most recent theoretical advances have been written and have made routine the determination of structures of ever increasing complexity. Based on these developments I now repeat Woolfson's predictions but change the date to the year 2000—a fitting way to greet the third millennium!

References

1. H. Ott, *Z. Kristallogr.* **66**, 136 (1927).
2. K. Banerjee, *Proc. R. Soc. London, Ser. A* **141**, 188 (1933).
3. M. Avrami, *Phys. Rev.* **54**, 300 (1938).
4. D. Harker, J. S. Kasper, *Acta Crystallogr.* **1**, 70 (1948).
5. J. Karle, H. Hauptman, *Acta Crystallogr.* **3**, 181 (1950).
6. H. Hauptman, J. Karle, *Phys. Rev.* **80**, 244 (1950).
7. H. Hauptman, J. Karle, *Solution of the Phase Problem I: The Centrosymmetric Crystal*, American Crystallographic Association monograph no. 3, Polycrystal Book Service, Dayton, Ohio (1953).
8. D. Langs, *Science* **241**, 188 (1988).