17.X-1 PROBLEMS IN SPACE-GROUP ASSIGNMENTS. By R. E. Marsh, A. A. Noyes Laboratory of Chemical Physics, California Institute of Technology, Pasadena, Calif. 91125, *USA*.

Three aspects of a proper space-group assignment will be discussed: finding the correct lattice, determining the appropriate Laue symmetry, and deciding whether or not the structure is centrosymmetric. An incorrect lattice may be chosen if reflections are overlooked, perhaps because they are systematically weak, or are thought to be present when they are not, perhaps because of interference from neighboring reflections. Incorrect Laue symmetry may result if a cell reduction is not carried out correctly, or if errors in cell dimensions are large (perhaps due to absorption or to a poor choice of reference reflections), or if an insufficient sampling of data has been examined. Resolving the centrosymmetric-noncentrosymmetric ambiguity is a more complicated problem and, in terms of molecular structure, a more serious one, as an incorrect choice may lead to large errors in atom positions. Some procedures for reducing the chance of error in all three areas will be suggested.

17.X-2 ERRORS IN ABSOLUTE-STRUCTURE DETERMINATION. by G. Bernardinelli and H. D. Flack, Laboratoire de Cristallographie, University of Geneva, 24 quai Ernest Ansermet, CH-1211 Genève 4, Switzerland.

The term "Absolute Structure" englobes the notions of both chirality and polarity (viz absolute configuration, absolute conformation, enantiomorph). A significant improvement in absolute-structure determination has recently resulted from the use of a single variable parameter which can be refined by least squares.

Absolute structure is a property of any non-centrosymmetric structure and a principal source of error is to ignore it. Systematic errors and effects in intensity data such as: data region, absorption correction, neglect of light atoms and stability constant have been investigated by the use of the refineable parameter. The stability constant causes the major perturbations. Attention will also be drawn to the following error sources in treating non-centrosymmetric structures: (1) Polar dispersion error; (2) Hand of axes; (3) Relation to physical and chemical properties.

In many cases, insufficient attention to the problems of absolute structure is being paid in the determination and publication of non-centrosymmetric structures.

17.X-3 ANALYSIS AND USE OF CRYSTALLOGRAPHIC DATA. By Frank H. Allen, Crystallographic Data Centre, University Chemical Labs., Lensfield Road, Cambridge, England.

The results of early X-ray analyses had a fundamental impact on the development of theories of chemical bonding in all its aspects. Concepts of ionicity, covalency, H-bonding, van der Waals interactions, etc. were formalized and 'geometrized' by systematic study of these early data. Over the past 25 years it is doubtful if modern inorganic chemistry would have progressed so rapidly had it not been underpinned by some 30,000 crystal structures.

As the number of reported structures rushed towards 100,000 attempts to systematize and interpret this data, beyond the confines of individual studies, abated considerably. The effort to locate and organize original data, against a backdrop of increased demand for more and more crystal structures, was a significant barrier. Increasing availability of crystallographic databases has done much to reduce this barrier : over 100 papers on 'molecular systematics' have appeared in the literature since 1980.

The sheer volume of data now available for analysis has, however, created its own barriers. The urgent need now is for numerical and statistical methods to aid in the extraction of meaningful crystallographic or chemical results. A number of such techniques are now in use, and examples will be given.

In the organic and biological areas considerable use of crystallographic data is now being made by non-specialists. The rapid expansion of molecular graphics facilities, particularly in the pharmaceutical industry, has created a demand for coordinate-based models of both 'known' and 'unknown' molecules. In the latter case a knowledge of likely dimensions and conformational preferences for many common substructural units is of vital importance. The derivation of these quantities, by application of methods noted above, is an important adjunct to purely computational techniques. Prospects for such a 'fragmentary' approach will be summarized.

17.X-4 OVERVIEW OF DIRECT METHODS, WITH ANOMALOUS DISPERSION. By H. A. Hauptman, Medical Foundation of Buffalo, 73 High St., Buffalo, NY 14203.

The electron density function $\rho(r)$ in a crystal determines its diffraction pattern, that is, both the magnitudes and phases of its X-ray diffraction maxima, and conversely. If, however, as is always the case, only magnitudes are available from the diffraction experiment, then the density function $\rho(r)$ cannot be recovered. If one invokes prior structural knowledge, usually that the crystal is composed of discrete atoms of known atomic numbers, then the observed magnitudes are, in general, sufficient to determine the positions of the atoms, that is, the crystal structure.

The intensities of a sufficient number of X-ray diffraction maxima determine the structure of a crystal. The available intensities usually exceed the number of parameters needed to describe the structure. From these intensities a set of numbers $|E_H|$ can be derived, one corresponding to each intensity. However, the elucidation of the crystal structure also requires a knowledge of the complex numbers $E_H = |E_H| \exp(i\phi_H)$, the normalized structure factors, of which only the magnitudes $|E_H|$ can be determined from experiment. Thus, a "phase" ϕ_H, unobtainable from the diffraction experiment, must be assigned to each $|E_H|$, and the problem of determining the phases when only the magnitudes $|E_H|$ are known is called "the phase problem". Owing to the known atomicity of crystal structures and the redundancy of observed magnitudes $|E_H|$, the phase problem is solvable in principle.

The values of the individual phases are determined by the crystal structure and the choice of origin. However, there always exist certain linear combinations of the phases whose values are determined by the structure alone and are independent of the choice of origin. These linear combinations of the phases are

called the structure invariants.

For fixed enantiomorph, the observed magnitudes $|E|$ determine, in general, unique values for all the structure invariants. The latter in turn, as certain well-defined linear combinations of the phases, lead unambiguously to unique values for the individual phases. Thus the structure invariants serve to link the known magnitudes $|E|$ with the desired phases ϕ (the fundamental principle of direct methods). By the term "direct methods" is meant that class of methods which exploits relationships among the structure factors in order to go directly from the observed magnitudes $|E|$ to the needed phases ϕ.

For fixed enantiomorph, the value of any structure invariant T is primarily determined, in favorable cases, by the values of one or more small sets of observed magnitudes $|E|$, the neighborhoods of T, and is relatively insensitive to the values of the great bulk of remaining magnitudes (the neighborhood principle). The conditional probability distribution of T, given the magnitudes in any of its neighborhoods, yields an estimate for T that is particularly good in the favorable case that the variance of the distribution happens to be small.

Most "small" crystal structures are rather routinely solvable nowadays by traditional direct methods. For the solution of macromolecular structures, on the other hand, the method of isomorphous replacement is universally used, and anomalous dispersion often plays an important supplementary role. One naturally anticipates therefore that integrating the traditional techniques of direct methods with isomorphous replacement and anomalous dispersion will strengthen our ability to solve complex structures. This goal has recently been achieved, and the initial applications suggest that the expected improvement is in fact realized.

17.X-5 OPTIMAL SYMBOLIC ADDITION. By H. Schenk and R. Peschar, Laboratory for Crystallography, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands.

In many direct program systems the selection of the starting set is based on the convergence procedure (Germain, G., Main, P. and Woolfson, M.M. (1970), Acta Cryst. B26, 274.), which starts from a set reflections with their relationships and finds the starting reflections by eliminating iteratively the weakest linked reflection. In the program SIMPEL (Schenk, H. and Kiers, C.T. (1985) in G.M. Sheldrick et al. (Eds) Crystallographic Computing 3, Oxford, 200-205) this starting set is then checked by a divergence procedure which explores the accessibility of all phases from the set. Nevertheless, in a number of cases SIMPEL fails as a result of a poor starting set and therefore we were looking for alternative procedures which build up phase sets directly. The optimal symbolic addition is such an alternative and determines systematically the theoretically most reliable phase sequences. This procedure is based on dynamic programming (Bellman, R. Dynamic Programming, Oxford, 1957) in which at each stage of a decision process, the best possible decision, according to some predefined criteria, is made. This implies for phase extension that the phases are determined by means of optimal decisions only. The criteria in the decision process are based on probabilistic arguments and result in a weighting scheme which includes triplet and quartet information and is suitable for symbolic phases as well. In general the application of the procedure results in a number of different sets of phases follows and for each of them a measure is given which indicates its expected success in the final phase extension. The mean phase error of these sets is much lower than the error in corresponding SIMPEL runs and a successful phase determination.

The research has been sponsored in part by STW, the Dutch technical research foundation.

17.X-6 MAXIMUM ENTROPY AND THE FOUNDATIONS OF DIRECT METHODS. By Gérard Bricogne, L.U.R.E., Batiment 209D, 91405 Orsay, France

This contribution will review and extend the author's previous work on a new approach to direct phase determination, presented in [1].

The Maximum Entropy (ME) method provides a practical yet optimal computational procedure for constructing conditional probability distributions of large numbers of structure factors, given assumed phases for a collection of large moduli. Its optimality follows from the equivalence of the MEM with the "saddlepoint approximation" (SPA) method of calculating asymptotic expansions of joint distributions in the presence of "large deviations", the latter being accommodated by constantly updating the prior distribution of the atoms in the cell.

This ME formalism has now been extended to the case of families of related structures made from several types of atoms, with arbitrary (complex) structure factors. The numbers of atoms of each type can be different in each structure of the family. The joint probability distribution of any "cylindrical" set of structure factors (comprising a given set of reflexions considered simultaneously across all members of the family of structures) can then be obtained, extending the recent results of Hauptman and of Karle on the incorporation into direct methods of isomorphous replacement and anomalous scattering. Other situations not hitherto considered, such as the availability of a contrast variation series, can be dealt with by this method. The equivalence between ME and SPA continues to hold in this generalised context. This derivation of statistical phase relations for arbitrary complex-valued scattering factors shows clearly that the source of such relations is the positivity of the prior probability distribution of the atoms, not the positivity of the electron density.

The ME formalism has also been extended into a statistical formulation of the molecular replacement method, by deriving joint distributions of structure factors in the presence of known structural fragments, of solvent regions, of non-crystallographic symmetries, and even in the case of multiple crystal forms. These extensions are readily merged with those concerning the treatment of families of related structures, and should provide a powerful tool for macromolecular crystallography.

Finally, the optimal Gaussian approximations of the conditional distributions given by the ME/SPA method have been used systematically to construct statistical likelihood functions from the observed data (including their error estimates). These likelihood functions afford a quantitative evaluation of the adequacy of the statistical model used to derive the conditional distribution in the first place. Their numerical optimisation affords a way of improving the statistical model, and in particular of refining the phase values associated to large moduli to make up the constraints : this refutes the commonly held view that "the ME method cannot refine phases". Furthermore, the likelihood functions have been obtained in a sufficiently general form to be able to consult not only single crystal data, but also fibre diffraction and powder diffraction data ; they can thus serve to extend the use of direct methods to these data.

It is this author's firm belief that this extended ME/SPA formalism and the associated likelihood functions constitute a powerful universal framework within which all sources of phase information can be first detected, then optimally combined, through a single basic computational mechanism in which - perhaps surprisingly - phase invariants never appear explicitly.

[1] G. Bricogne : "Maximum Entropy and the Foundations of Direct Methods" Acta Cryst. (1984) A40, 410-445.