

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  $\phi$  (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  $\phi$ .

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.