

## 2.5. ELECTRON DIFFRACTION AND ELECTRON MICROSCOPY IN STRUCTURE DETERMINATION

## 2.5.7.7. Maximum entropy and likelihood

Maximum entropy has been applied to electron crystallography in several ways. In the sense that images are optimized, the entropy term

$$S = -\sum_i P_i \ln P_i,$$

where  $P_i = p_i / \sum_i p_i$  and  $p_i$  is a pixel density, has been evaluated for various test electron-microscope images. For crystals, the true projected potential distribution function is thought to have the maximum value of  $S$ . If the phase contrast transfer function used to obtain a micrograph is unknown, test images (*i.e.* trial potential maps) can be calculated for different values of  $\Delta f_{\text{trial}}$ . The value that corresponds to the maximum entropy would be near the true defocus. In this way, the actual objective lens transfer function can be found for a single image (Li, 1991) in addition to the other techniques suggested by this group.

Another use of the maximum-entropy concept is to guide the progress of a direct phase determination (Bricogne & Gilmore, 1990; Gilmore *et al.*, 1990). Suppose that there is a small set  $H$  of known phases  $\phi_{\mathbf{h} \in H}$  (corresponding either to origin definition, or the Fourier transform of an electron micrograph, or both) with associated unitary structure-factor amplitudes  $|U_{\mathbf{h} \in H}|$ . [The unitary structure factor is defined as  $|U_{\mathbf{h}}| = |E_{\mathbf{h}}|/(N)^{1/2}$ .] As usual, the task is to expand into the unknown phase set  $K$  to solve the crystal structure. From Bayes' theorem, the procedure is based on an operation where  $p(\text{map}|\text{data}) \propto p(\text{map})p(\text{data}|\text{map})$ . This means that the probability of successfully deriving a potential map, given diffraction data, is estimated. This so-called posterior probability is approximately proportional to the product of the probability of generating the map (known as the prior) and the probability of generating the data, given the map (known as the likelihood). The latter probability consults the observed data and can be used as a figure of merit.

Beginning with the basis set  $H$ , a trial map is generated from the limited number of phased structure factors. As discussed above, the map can be immediately improved by removing all negative density. The map can be improved further if its entropy is maximized using the equation given above for  $S$ . This produces the so-called maximum-entropy prior  $q^{\text{ME}}(X)$ .

So far, it has been assumed that all  $|U_{\mathbf{h} \in K}| = 0$ . If large reflections from the  $K$  set are now added and their phase values are permuted, then a number of new maps can be generated and their entropies can be maximized as before. This creates a phasing 'tree' with many possible solutions; individual branch points can have further reflections added *via* permutations to produce further sub-branches, and so on. Obviously, some figure of merit is needed to 'prune' the tree, *i.e.* to find likely paths to a solution.

The desired figure of merit is the likelihood  $L(H)$ . First a quantity

$$\Lambda_{\mathbf{h}} = 2NR \exp[-N(r^2 + R^2)]I_0(2NrR),$$

where  $r = |U_{\mathbf{h}}^{\text{ME}}|$  (the calculated unitary structure factors) and  $R = |U_{\mathbf{h}}^{\text{O}}|$  (the observed unitary structure factors), is defined. From this one can calculate

$$L(H) = \sum_{\mathbf{h} \notin H} \ln \Lambda_{\mathbf{h}}.$$

The null hypothesis  $L(H_0)$  can also be calculated from the above when  $r = 0$ , so that the likelihood gain

$$LLg = L(H) - L(H_0)$$

ranks the nodes of the phasing tree in order of the best solutions.

Applications have been made to experimental electron-crystallographic data. A small-molecule structure starting with phases from an electron micrograph and extending to electron-diffraction

resolution has been reported (Dong *et al.*, 1992). Other experimental electron-diffraction data sets used in other direct phasing approaches (see above) also have been assigned phases by this technique (Gilmore, Shankland & Bricogne, 1993). These include intensities from diketopiperazine and basic copper chloride. An application of this procedure in protein structure analysis has been published by Gilmore *et al.* (1992) and Gilmore, Shankland & Fryer (1993). Starting with 15 Å phases, it was possible to extend phases for bacteriorhodopsin to the limits of the electron-diffraction pattern, apparently with greater accuracy than possible with the Sayre equation (see above).

## 2.5.7.8. Influence of multiple scattering on direct electron crystallographic structure analysis

The aim of electron-crystallographic data collection is to minimize the effect of dynamical scattering, so that the unit-cell potential distribution or its Fourier transform is represented significantly in the recorded signal. It would be a mistake, however, to presume that these data ever conform strictly to the kinematical approximation, for there is always some deviation from this ideal scattering condition that can affect the structure analysis. Despite this fact, some direct phasing procedures have been particularly 'robust', even when multiple scattering perturbations to the data are quite obvious (*e.g.* as evidenced by large crystallographic residuals).

The most effective direct phasing procedures seem to be those based on the  $\Sigma_2$  triple invariants. These phase relationships will not only include the symbolic addition procedure, as it is normally carried out, but also the tangent formula and the Sayre equation (since it is well known that this convolution can be used to derive the functional form of the three-phase invariant). The strict ordering of  $|E_{\mathbf{h}}|$  magnitudes is, therefore, not critically important so long as there are no major changes from large to small values (*or vice versa*). This was demonstrated in direct phase determinations of simulated  $n$ -beam dynamical diffraction data from a sulfur-containing polymer (Dorset & McCourt, 1992). Nevertheless, there is a point where measured data cannot be used. For example, intensities from *ca* 100 Å-thick epitaxially oriented copper perchlorophthalocyanine crystals become less and less representative of the unit-cell transform at lower electron-beam energies (Tivol *et al.*, 1993) and, accordingly, the success of the phase determination is compromised (Dorset, McCourt, Fryer *et al.*, 1994). The similarity between the Sayre convolution and the interactions of structure-factor terms in, *e.g.*, the multislice formulation of  $n$ -beam dynamical scattering was noted by Moodie (1965). It is interesting to note that dynamical scattering interactions observed by direct excitation of  $\Sigma_2$  and  $\Sigma_1$  triples in convergent-beam diffraction experiments can actually be exploited to determine crystallographic phases to very high precision (Spence & Zuo, 1992, pp. 56–63).

While the evaluation of positive quartet invariant sums (see Chapter 2.2) seems to be almost as favourable in the electron diffraction case as is the evaluation of  $\Sigma_2$  triples, negative quartet invariants seem to be particularly sensitive to dynamical diffraction. If dynamical scattering can be modelled crudely by a convolutional smearing of the diffraction intensities, then the lowest structure-factor amplitudes, and hence the estimates of lowest  $|E_{\mathbf{h}}|$  values, will be the ones most compromised. Since the negative-quartet relationships require an accurate prediction of small 'cross-term'  $|E_{\mathbf{h}}|$  values, multiple scattering can, therefore, limit the efficacy of this invariant for phase determination. In initial work, negative quartets have been mostly employed in the NQUEST figure of merit, and analyses (Dorset, McCourt, Fryer *et al.*, 1994; Dorset & McCourt, 1994a) have shown how the degradation of weak kinematical  $|E_{\mathbf{h}}|$  terms effectively reduced its effectiveness for

## 2. RECIPROCAL SPACE IN CRYSTAL-STRUCTURE DETERMINATION

locating correct structure solutions *via* the tangent formula, even though the tangent formula itself (based on triple phase estimates) was quite effective for phase determination. Substitution of the minimal function  $R(\phi)$  for NQUEST seems to have overcome this difficulty. [It should be pointed out, though, that only the  $\Sigma_2$ -triple contribution to  $R(\phi)$  is considered.]

Structure refinement is another area where the effects of dynamical scattering are also problematic. For example, in the analysis of the paraelectric thiourea structure (Dorset, 1991*b*) from published texture diffraction data (Dvoryankin & Vainshtein, 1960), it was virtually impossible to find a chemically reasonable structure geometry by Fourier refinement, even though the direct phase determination itself was quite successful. The best structure was found only when higher-angle intensities (*i.e.* those least affected by dynamical scattering) were used to generate the potential map. Later analyses on heavy-atom-containing organics (Dorset *et al.*, 1992) found that the lowest kinematical  $R$ -factor value did not correspond to the chemically correct structure geometry. This observation was also made in the least-squares refinement of diketopiperazine (Dorset & McCourt, 1994*a*). It is

obvious that, if a global minimum is sought for the crystallographic residual, then dynamical structure factors, rather than kinematical values, should be compared to the observed values (Dorset *et al.*, 1992). Ways of integrating such calculations into the refinement process have been suggested (Sha *et al.*, 1993). Otherwise one must constrain the refinement to chemically reasonable bonding geometry in a search for a local  $R$ -factor minimum.

Corrections for such deviations from the kinematical approximation are complicated by the presence of other possible data perturbations, especially if microareas are being sampled, *e.g.* in typical selected-area diffraction experiments. Significant complications can arise from the diffraction incoherence observed from elastically deformed crystals (Cowley, 1961) as well as secondary scattering (Cowley *et al.*, 1951). These complications were also considered for the larger (*e.g.* millimeter diameter) areas sampled in an electron-diffraction camera when recording texture diffraction patterns (Turner & Cowley, 1969), but, because of the crystallite distributions, it is sometimes found that the two-beam dynamical approximation is useful (accounting for a number of successful structure analyses carried out in Moscow).