

Integration of macromolecular diffraction data using radial basis function networks

Boris Pokrić,^a Nigel M. Allinson^{a*} and John R. Helliwell^b

^aUMIST, Department of Electrical Engineering and Electronics, PO Box 88, Manchester M60 1QD, UK, and ^bUniversity of Manchester, Department of Chemistry, Manchester M13 9PL, UK. E-mail: allinson@umist.ac.uk

(Received 14 June 2000; accepted 18 September 2000)

This paper presents a novel approach for intensity calculation of X-ray diffraction spots based on a two-stage radial basis function (RBF) network. The first stage uses pre-determined reference profiles from a database as basis functions in order to locate the diffraction spots and identify any overlapping regions. The second-stage RBF network employs narrow basis functions capable of local modifications of the reference profiles leading to a more accurate observed diffraction spot approximation and therefore accurate determination of spot positions and integrated intensities.

Keywords: spot integration; RBF networks; profile fitting; diffraction data processing.

1. Introduction

X-ray diffraction data processing proceeds through several stages, namely indexing, pre-refinement of camera parameters and crystal orientation, intensity integration, post-refinement and scaling. This paper addresses issues regarding intensity integration, the process of obtaining estimates of intensities for each diffraction spot. Two different methods can be used to estimate the integrated intensities: summation integration and profile fitting (Rossmann, 1979; Leslie, 1999). The summation integration involves adding the pixel values within the peak area and subtracting the estimated underlying background. In the profile-fitting procedure, known spot profiles are fitted to the observed diffraction spots in order to determine the integrated intensity. The summation integration method is computationally inexpensive, but it fails in situations where two or more diffraction spots overlap. The profile-fitting method improves the accuracy for weak intensities and is generally more robust for overlapping spots. However, the critical stage of devising standard profiles greatly influences the results. A more detailed discussion of this method and possible improvements are given below.

2. Profile fitting

There are two general approaches for deriving the reference profiles that can be used for the approximation of observed diffraction spots. The first approach is to use analytical functions that are sufficiently flexible in modelling diffraction spots over wide variations in spot shape. Many different functions have been proposed, such as the pseudo-Voigt (Thompson *et al.*, 1987), Pearson VII (Hall

et al., 1977) and Gaussian–Hermite polynomials (Sanchez-Bajo & Cumbre, 1999). However, many of these approaches require the use of non-linear optimization that is computationally extensive for the large number of necessary parameters. The second approach is to build a library of reference profiles by averaging a large number of Bragg spots located near to the reflection under evaluation (Rossmann, 1979; Bourgeois, 1999; Leslie, 1999). The averaging can be performed in two-dimensional or three-dimensional space (*i.e.* thick or thin slicing) (Pflugrath, 1999). The intensities of the reflections are then estimated by fitting the corresponding reference profile to the spot data by calculating the position and scale factor that minimizes the mean square error (Durbin & Gog, 1989; Messerschmidt & Pflugrath, 1987; Kabsch, 1988). Improvements in the profile-fitting process have also been achieved using the profile-interpolation method (Bourgeois, 1999).

Regardless of what method is used, it is very difficult to devise the reference profiles that are consistently the best representations of the observed Bragg spots. The spots can overlap, especially when the Laue diffraction method is used (Cruickshank *et al.*, 1991), or their profiles can change from frame to frame (in thin slicing) and experiment to experiment leading to inaccurate determinations of the integrated intensities. This can occur due to variation in parameters such as diffraction power of the crystal (*e.g.* due to crystal degradation), detector-to-sample distance, position-dependent point-spread functions (Bourgeois, 1999) and X-ray beam divergence (Bourgeois, 1999; Leslie, 1999). The intensity of the diffraction spots is a most important parameter which, in order to refine a crystal structure, needs to be determined with a mean uncertainty better

than 10% (Westbrook, 1988). Furthermore, in order to determine a new structure using multiple-energy anomalous dispersion (MAD) methods, the measurement uncertainty must be better than 2% (Ross *et al.*, 1998).

Therefore, a more robust approach is suggested, based on radial basis function (RBF) neural networks, a well established technique that has been applied to a wide variety of problems such as image processing (Saha *et al.*, 1990), speech recognition (Ng & Lippmann, 1990), adaptive equalization (Chen *et al.*, 1992) and medical diagnosis (Lowe & Webb, 1990). The technique for diffraction data processing presented in this paper extends the conventional reference profile-fitting method by using a concentrated RBF network design (Pokrić *et al.*, 1999) in order to compensate for possible changes in the observed spot profiles.

3. Overview of modified profile-fitting approach

Fig. 1 shows the main stages in the overall diffraction data processing procedure. An X-ray detector acquires diffraction data images which are first corrected for static instrument distortions, namely non-uniformity intensity response, non-linearity and spatial distortion. The corrected image is processed in blocks of pre-defined size in order to reduce computational effort. The following stage is

based on the reference profile-fitting technique. The incremental RBF network, with reference profiles used as basis functions, is employed for estimation of the initial positions of diffraction spots together with the background signal. The initial positions of spots are used to identify the possible overlapping regions between spots. The next stage estimates the noise probability density function (PDF) from the background intensity levels in the spot neighbourhood. The noise PDF is used to design a concentrated RBF network by determining the minimum contribution boundary that the basis functions need to satisfy inclusion in the network. The final stage calculates the Bragg spot integrated intensities and positions. A more detailed description of these stages is given in the following sections.

4. Reference profile fitting using incremental RBF network

This stage estimates initial positions and amplitudes of the diffraction spots together with the background signal plane in the image region under consideration. Prior to this phase it is possible to remove strong outliers in the data so that further data processing is not affected. This can be performed using a technique based on Wilson statistics (Read, 1999). The successive placement of the reference profiles is performed using RBF network methodology

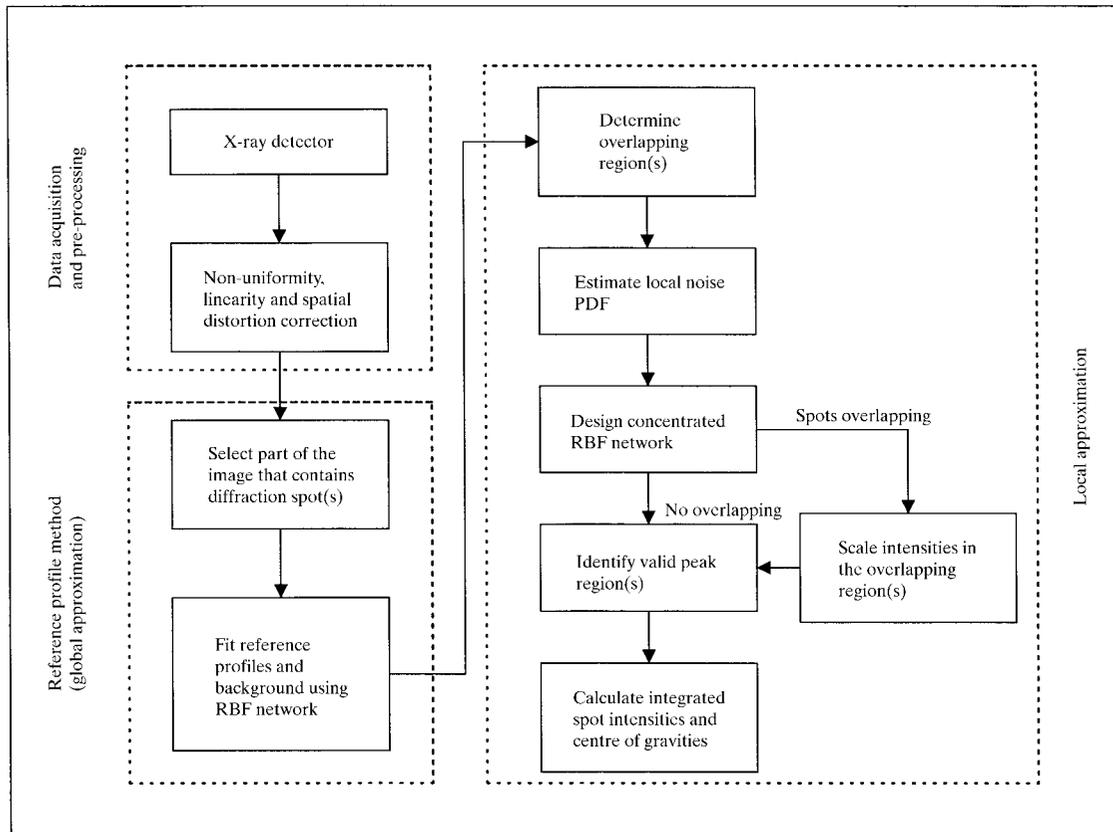


Figure 1
Block diagram of proposed diffraction data processing.

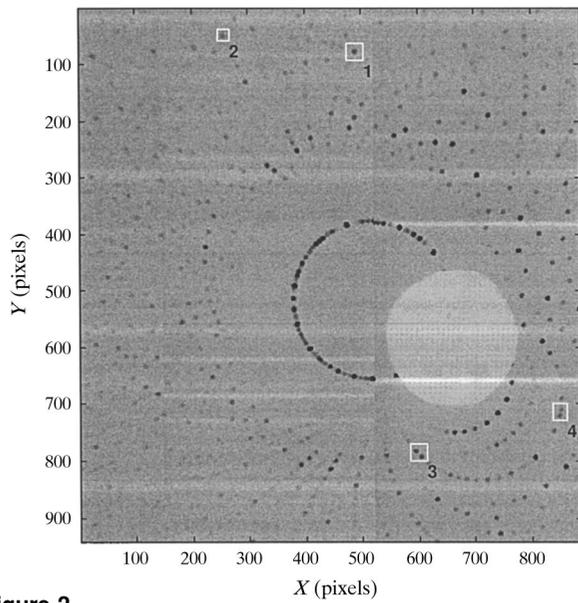


Figure 2
Laue diffraction pattern of concanavalin crystal with specified processed image segments marked.

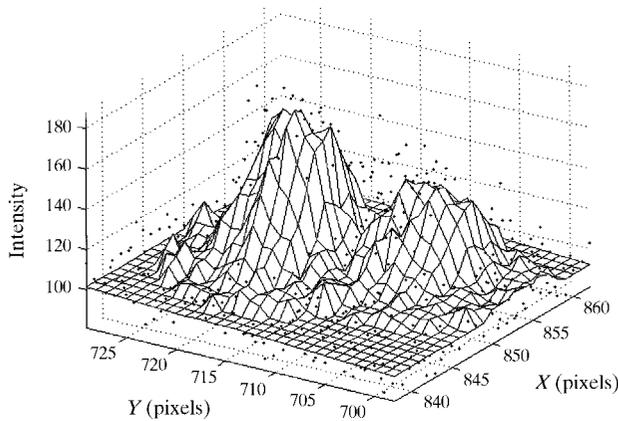


Figure 3
Reference profile fitting using incremental RBF network.

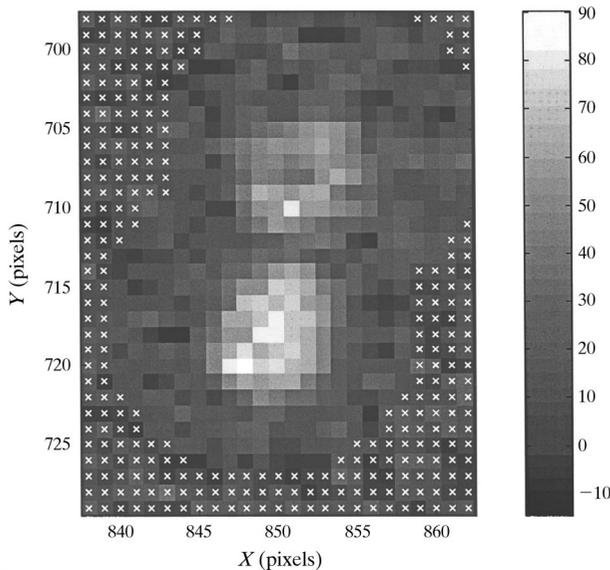


Figure 4
Selected pixels used for noise PDF estimation.

where the reference profiles are used as basis functions. RBF networks are capable of non-linear processing if their basis function placements or spreads are modifiable (or if they contain more than one hidden layer). For least-mean-square (LMS) supervised training, the need is to determine the minimum of the cost function (Haykin, 1994),

$$C = \sum_{i=1}^p [\hat{z}_i - f(x_i, y_i) - ax_i - by_i - c]^2, \quad (1)$$

where

$$f(x_i, y_i) = \sum_{j=1}^m \hat{w}_j h_j(x_i, y_i),$$

a , b and c are background plane constants, $\{w_j\}_{j=1}^m$ are second layer weights, $\{h_j\}_{j=1}^m$ are basis functions (*i.e.* reference profiles) and $\{(x_i, y_i, \hat{z}_i)\}_{i=1}^p$ is a paired training set.

The minimization of C leads to a set of $(m + 3)$ simultaneous linear equations, for which the optimum set of weights, \hat{w} , and the background plane constants in the LMS error sense can be expressed as

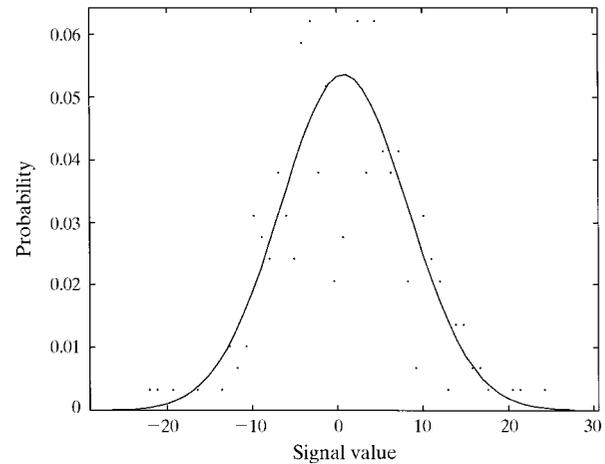


Figure 5
Estimated noise PDF with the following parameters: $P_n = 0.73$, $\mu_n = 1$ and $\sigma_n = \pm 6.98$.

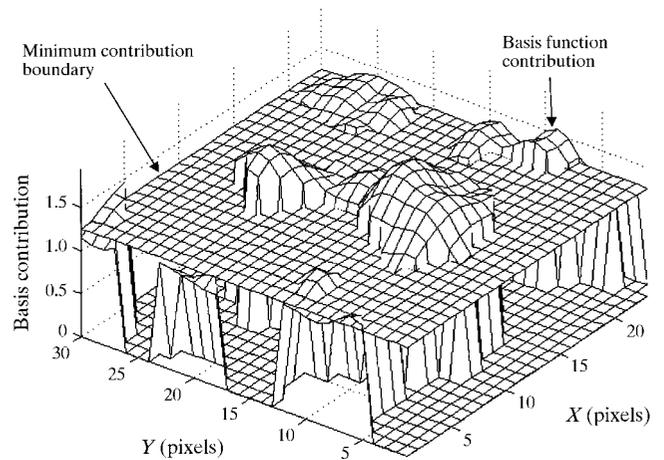


Figure 6
Minimum contribution boundary and individual basis function contribution to the final approximation.

$$\begin{bmatrix} \hat{w}_1 \\ \hat{w}_2 \\ \vdots \\ \hat{w}_m \\ a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \sum h_1^2 & \sum h_1 h_2 & \cdots & \sum h_1 h_m & \sum h_1 \mathbf{x} & \sum h_1 \mathbf{y} & \sum h_1 \\ \sum h_1 h_2 & \sum h_2^2 & \cdots & \sum h_2 h_m & \sum h_2 \mathbf{x} & \sum h_2 \mathbf{y} & \sum h_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \sum h_m h_2 & \sum h_m h_2 & \cdots & \sum h_m^2 & \sum h_m \mathbf{x} & \sum h_m \mathbf{y} & \sum h_m \\ \sum h_1 \mathbf{x} & \sum h_2 \mathbf{x} & \cdots & \sum h_m \mathbf{x} & \sum \mathbf{x}^2 & \sum \mathbf{x} \mathbf{y} & \sum \mathbf{x} \\ \sum h_1 \mathbf{y} & \sum h_2 \mathbf{y} & \cdots & \sum h_m \mathbf{y} & \sum \mathbf{x} \mathbf{y} & \sum \mathbf{y}^2 & \sum \mathbf{y} \\ \sum h_1 & \sum h_2 & \cdots & \sum h_m & \sum \mathbf{x} & \sum \mathbf{y} & m \cdot p \end{bmatrix}^{-1} \begin{bmatrix} \sum h_1 \hat{z} \\ \sum h_2 \hat{z} \\ \vdots \\ \sum h_m \hat{z} \\ \sum \mathbf{x} \hat{z} \\ \sum \mathbf{y} \hat{z} \\ \sum \hat{z} \end{bmatrix}. \quad (2)$$

The approximated output, \hat{z}_i , for the i th component, is given by

$$\begin{aligned} \hat{z}_i = f_i = f(x_i, y_i) &= ax_i + by_i + c + \sum_{j=1}^m \hat{w}_j h_j(x_i, y_i) \\ &= ax_i + by_i + c + \bar{\mathbf{h}}_i^T \hat{\mathbf{w}}, \end{aligned} \quad (3)$$

where

$$\bar{\mathbf{h}}_i = \begin{bmatrix} h_1(\mathbf{x}_i, \mathbf{y}_i) \\ h_2(\mathbf{x}_i, \mathbf{y}_i) \\ \vdots \\ h_m(\mathbf{x}_i, \mathbf{y}_i) \end{bmatrix}.$$

The entire approximated output, \mathbf{r} , for all p training pairs can be expressed as

$$\mathbf{r} = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_p \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{h}}_1^T \hat{\mathbf{w}} \\ \bar{\mathbf{h}}_2^T \hat{\mathbf{w}} \\ \vdots \\ \bar{\mathbf{h}}_p^T \hat{\mathbf{w}} \end{bmatrix} + ax_i + by_i + c = \mathbf{H} \hat{\mathbf{w}}, \quad (4)$$

where \mathbf{H} is the design matrix.

In developing an RBF network, one common strategy is to use the method of forward selection (Orr, 1996), which starts with an empty subset of potential basis functions to which one function is added at a time until some criterion is satisfied. The complete set of potential basis functions can be expressed as

$$\mathbf{F} = [\mathbf{f}_1 \quad \mathbf{f}_2 \quad \dots \quad \mathbf{f}_M], \quad (5)$$

where basis functions in this case are reference profiles at corresponding positions within the image region under consideration. After the full design matrix is determined, the forward selection process of the basis functions can proceed. The basis function that most reduces the cost function is selected first and the next function will be that which most reduces the residual cost function, and so on. Using this approach it is important to note that each selected basis function influences the contribution of other basis functions in the neighbourhood. Therefore, the contribution of each basis function towards the cost function is calculated using the contribution matrix, \mathbf{K} . This matrix is initially identical to the full design matrix, \mathbf{F} , and it is modified every time a new basis function, \mathbf{f}_s , is selected. The update rule for the contribution matrix at each iteration i is as follows,

$$\mathbf{K}_{(i)} = \mathbf{K}_{(i-1)} - \mathbf{f}_s \mathbf{g}, \quad (6)$$

where

$$\mathbf{g} = \mathbf{f}_s^T \mathbf{K}_{(i-1)} / (\mathbf{f}_s^T \mathbf{f}_s).$$

The contribution, \mathbf{e} , of all basis functions towards the cost function is calculated as follows,

$$\mathbf{e} = \mathbf{K}^T \hat{\mathbf{z}} + ax + by + c. \quad (7)$$

The basis function, which maximizes this contribution, is selected from the full design matrix, \mathbf{F} , and placed in the design matrix, \mathbf{H} . This process is repeated until one of the following criteria is satisfied:

- (i) The generalized cross-validation (GCV) stops decreasing (Pokrić *et al.*, 1999; Golub *et al.*, 1979; Craven & Wahba, 1979);
- (ii) The weight of the last selected basis function is below a pre-defined minimum value (typically equal to the estimated noise RMS value);
- (iii) The approximation error RMS value is below the estimated noise RMS value;
- (iv) The number of basis functions m exceeds a pre-defined maximum value.

Fig. 2 shows a region of a Laue diffraction pattern of concanavalin crystal obtained using the back-illuminated CCD direct detection detector system (Allinson *et al.*, 1994; Allinson, 1994). The region under consideration is marked by a small rectangular box. Fig. 3 shows the result of reference profile fitting using an incremental RBF network. As can be seen, two main diffraction spots have been detected and approximated using scaled reference profiles.

5. Fitting refinement using a concentrated RBF network

Fitting refinement is performed using a ‘local approximation’ RBF network that is defined by its full design matrix $\mathbf{F}_{\text{local}}$ and design matrix $\mathbf{H}_{\text{local}}$ containing typically narrow Gaussian basis functions. This RBF network is used to estimate the residual signal after reference profile fitting. Once the initial positions and amplitudes of diffraction spots are estimated (*i.e.* fitted reference profile positions and their amplitudes), it is possible to determine regions in the image region where approximation refinement is necessary. This ‘local approximation’ is based on the best estimate of the noise in the neighbourhood of the identified spots. Fig. 4 shows the initial pixel positions whose intensities are used in the noise PDF estimation procedure using the expectation–maximization algorithm (Vaseghi, 1996;

Proakis *et al.*, 1992; Dempster *et al.*, 1977). The pixel intensities that deviate for more than three standard deviations, σ_n , are rejected and the noise PDF is re-evaluated. This process is repeated until all pixel intensities fall within the $\pm 3\sigma_n$ region. The final PDF of a random signal \mathbf{n} can be represented as

$$f_{n|\Theta}(n|\hat{\Theta}) = (1/2\pi\sigma_n^2) \exp[-(x - \mu_n)^2/2\sigma_n^2], \quad (8)$$

where $\hat{\Theta}$ are parameters of the Gaussian mixture representing prior probability (P_n), mean (μ_n) and standard deviation (σ_n) of each function.

The final estimated noise PDF is shown in Fig. 5. There is only a need to consider those basis functions (and their positions) which can materially contribute to the final approximating function, \mathbf{z}_a . This requires that the influence of a basis function must exceed the system noise level. Hence, the design matrix employs only those basis functions that satisfy the relationship

$$\eta \sum_{i=1}^p \mathbf{b}_{ji} < |(\hat{\mathbf{z}} - \mathbf{r})^T| \mathbf{b}_j, \quad (9)$$

where η is calculated from a pre-defined confidence limit δ using

$$\int_{-\eta}^{\eta} f(n) dn = \delta/100,$$

where $0\% \leq \delta \leq 100\%$, \mathbf{r} is the full approximated output of the reference profile RBF network and \mathbf{b} are the basis functions in the full design matrix $\mathbf{F}_{\text{local}}$.

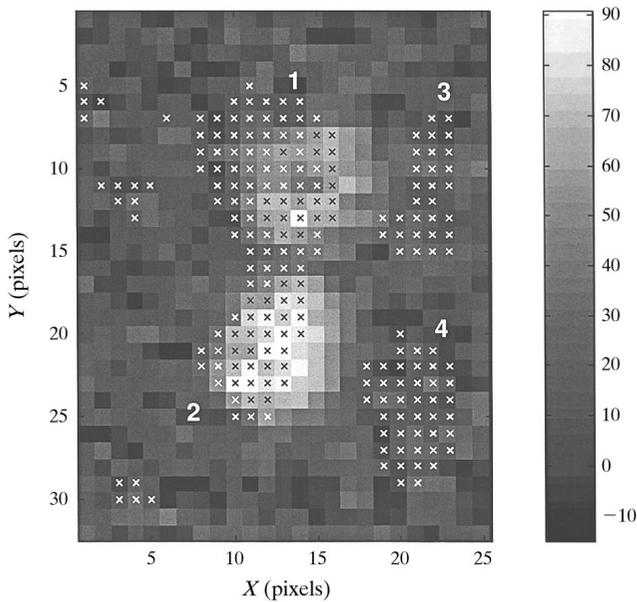


Figure 7
Selected basis function (marked by crosses) used in concentrated RBF network (local approximation). Main peak regions are marked by numbers 1 and 2. Regions marked by numbers 3 and 4 result from the inaccurate background model within the reference profile.

The left-hand side of (9) for the image region under consideration is graphically represented in Fig. 6 as a flat plane. Basis function contributions [right-hand side of (9)] are shown on the same plot. Only basis functions, which exceed the minimum contribution boundary plane, are incorporated in the design matrix $\mathbf{H}_{\text{local}}$. The positions of selected basis functions are marked in Fig. 7 as crosses. As can be seen, there are four distinct regions where local approximation basis functions are placed (marked 1, 2, 3 and 4). Regions marked with numbers 1 and 2 are the main peak regions where reference peaks could not provide a sufficiently accurate model of the observed spots. Regions marked with numbers 3 and 4 result from an inaccurate background model in the reference peak profiling. However, this does not influence the final result as the local approximation basis functions correct for these irregularities. Final approximation of the residual signal using the concentrated RBF network and the approximation error are shown in Figs. 8 and 9, respectively. The standard deviation of the approximation error is ± 6.11 , which is very similar to the estimated noise standard deviation of ± 6.98 .

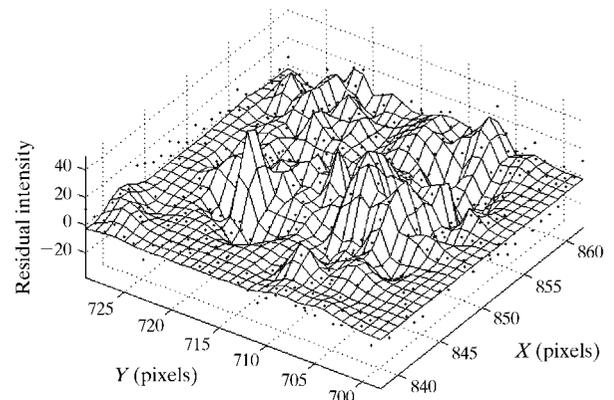


Figure 8
Concentrated RBF network approximation of residual signal (*i.e.* original data subtracted from reference profile approximation signal).

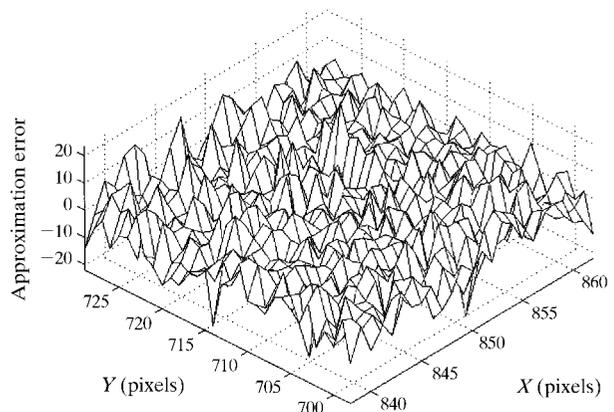


Figure 9
Approximation error with $\sigma = \pm 6.11$.

This, of course, is the objective of the RBF approximation, namely to approximate the ‘true’ signal without over-fitting. The final approximation using a two-stage RBF network (*i.e.* reference profile plus local approximation stages) is shown in Fig. 10.

6. Calculation of diffraction spot parameters

The parameters that need to be extracted for each diffraction spot are the exact position in the image and the integrated intensity. Providing that there are no overlapping spots in the image region under consideration, parameter calculation can be performed in the following way: (i) identify the valid spot region(s); (ii) calculate the integrated intensity within the valid region(s); (iii) calculate the centre of gravity for each spot.

The valid spot regions are within the boundary line that divides the valid spot signal values and background values. The boundary line is calculated from individual spot profiles that are obtained by taking the approximation intensities along the line originating from the reference spot centre point. The orientation of the line is changed from 0 to 360° in discrete steps. The boundary point for the corresponding line angle lies at the position where the spot profile along that line reaches the background intensity level or global minimum in the case of overlapping spots. The spot profiles can be calculated with a sub-pixel accuracy using the analytical representation of the concentrated RBF network, \hat{r}_{local} , and interpolated reference profiles \hat{r} . The interpolated reference profile is calculated using the standard bi-cubic interpolation technique (Press *et al.*, 1995). The analytical representation of concentrated RBF network for Gaussian basis functions is

$$\hat{r}_{\text{local}}(x_i, y_i) = \sum_{j=1}^m \hat{w}_j \exp\left[-(x_i - \mu_{xj})^2 / \sigma_x^2\right] \times \exp\left[-(y_i - \mu_{yj})^2 / \sigma_y^2\right], \quad (10)$$

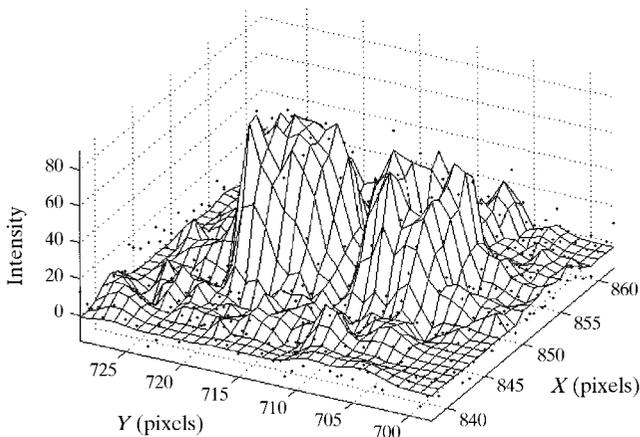


Figure 10
Final approximation of two overlapping Bragg spots.

where x_i and y_i are rational numbers, m is the number of basis functions in the design matrix $\mathbf{H}_{\text{local}}$, (μ_{xj} , σ_{xj}) and (μ_{yj} and σ_y) are the mean and standard deviation in the x and y directions, respectively, of the j th basis function.

The individual spot profile $p(\theta)$ calculated at a desired resolution Δt is therefore

$$p(\theta, x_i, y_i) = \hat{r}(x_i, y_i) + \hat{r}_{\text{local}}(x_i, y_i), \quad (11)$$

where $x_i = t \cos(\theta) + x_r$ and $y_i = t \sin(\theta) + y_r$ for $t = 0, \Delta t, 2\Delta t, \dots$ and (x_r, y_r) is the reference profile centre.

An example of the calculated boundary line is shown in Fig. 11. Once the valid spot region is identified, it is straightforward to calculate the integrated spot intensity. As the integrated intensity for the normalized reference peaks is known *a priori*, it is only necessary to calculate the contribution of the concentrated RBF network. The contribution of the concentrated RBF network towards the integrated intensity is calculated using the analytical representation within the valid spot regions. Adding this contribution to the scaled integrated intensity of the corresponding reference profile leads to the final integrated intensity of the diffraction spot. The centre of gravity (c_x, c_y) is then calculated from intensity levels within the valid peak region using

$$c_x = \frac{\sum_i x_i [\hat{r}(x_i, y_i) + \hat{r}_{\text{local}}(x_i, y_i)]}{\sum_i \hat{r}(x_i, y_i) + \hat{r}_{\text{local}}(x_i, y_i)}, \quad (12)$$

$$c_y = \frac{\sum_i y_i [\hat{r}(x_i, y_i) + \hat{r}_{\text{local}}(x_i, y_i)]}{\sum_i \hat{r}(x_i, y_i) + \hat{r}_{\text{local}}(x_i, y_i)}.$$

The centre of gravity of the spot shown in Fig. 11 is marked by a cross. Two more examples of processing of non-overlapping diffraction spots are shown in Figs. 12 and 13. As can be seen from these examples, the fitting refinement

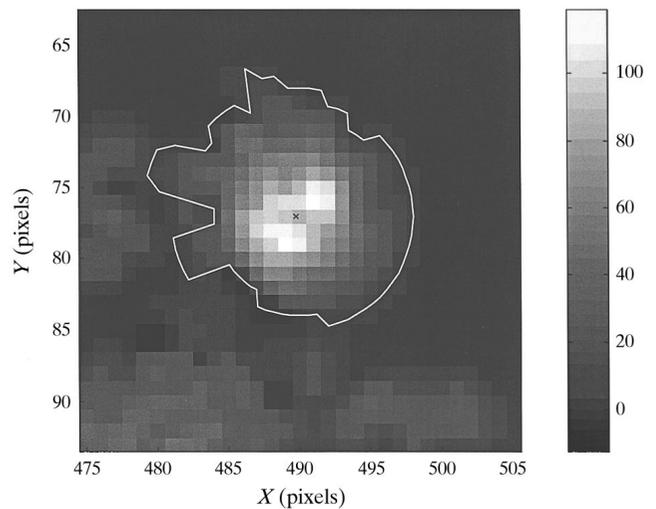


Figure 11
Boundary line with indicated centre of reference profile (marked by a cross) of spot marked by box number 1 in Fig. 2.

Table 1

Summary of results after processing of image segments marked by boxes 1, 2, 3 and 4 in Fig. 2.

Parameter	Segment number					
	1	2	3		4	
			Spot 1	Spot 2	Spot 1	Spot 2
C_x (pixel)	489.74	256.68	604.22	594.32	849.93	851.79
C_y (pixel)	76.98	48.88	790.92	780.87	717.92	707.99
Noise RMS (ADU)	11.65	9.36	11.94	11.94	7.91	7.91
Reference RMS (ADU)	22.741	24.45	17.23	21.67	16.34	19.71
Reference intensity	10733.67	1581.93	8737.06	9649.22	6211.95	4346.82
Reference $I/\sigma(I)$	472.00	64.70	507.08	445.28	380.17	220.54
Final RMS (ADU)	11.03	5.41	8.65	9.55	6.99	5.66
Final intensity	8701.99	2028.59	7271.22	7974.96	6090.49	3610.03
Final $I/\sigma(I)$	788.94	374.97	840.60	835.07	871.31	637.81

using the concentrated RBF network considerably improves the approximation RMS error over the reference profile approximation RMS error. The final approximation RMS error approaches the RMS noise value, which indicates that the concentrated network approximates the data accurately without over-fitting the residual intensity raw data. The integrated intensity calculated using only the reference profile differs from the intensity calculated using the refined approximation by as much as 20%, with marked improvement in the $I/\sigma(I)$ ratio for examples shown in Figs. 11–13. The results are summarized in Table 1.

In the situations when the image region contains two or more spots that overlap, it is first necessary to deconvolve the spots so that the individual integrated intensities and centres of gravity can be calculated. This is achieved by identifying the overlapping regions between the spots by using the initially fitted reference profiles. Fig. 14 shows two overlapping diffraction spots with centre positions of reference profiles marked by dark crosses. The pixels in the overlapping region are marked by white crosses. The spot

profiles \mathbf{p} , defined by equation (11), are scaled in the overlapping region using the reference profile geometry. The scaled spot profile \bar{p} within the overlapping region of spots a and b is calculated as

$$\bar{p}_a(\theta_o, x_o, y_o) = \frac{\hat{r}_a(x_o, y_o)}{\hat{r}_a(x_o, y_o) + \hat{r}_b(x_o, y_o)} p(\theta_o, x_o, y_o), \quad (13)$$

where θ_o , x_o and y_o are line angles, x and y are positions within the overlapping region and $\hat{r}_a(x_o, y_o)$ and $\hat{r}_b(x_o, y_o)$ are the reference profile intensities of spots a and b at the overlapping position (x_o, y_o) .

The valid regions of individual spots are determined using the combination of spot profiles in no overlapping regions and scaled spot profiles in the overlapping regions (see Fig. 15). Determination of integrated spot intensities and centres of gravity are performed in the same manner as for non-overlapping spots taking care to use scaled spot profiles in the overlapping regions.

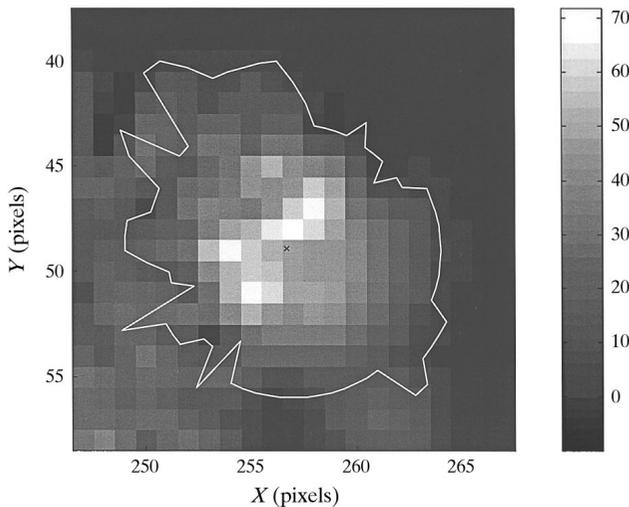


Figure 12 Boundary line with indicated centre of reference profile (marked by a cross) of spot marked by box number 2 in Fig. 2.

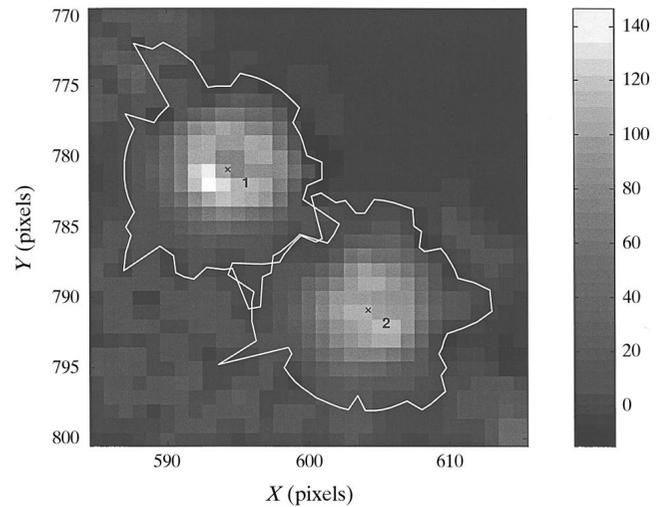


Figure 13 Boundary line with indicated centre of reference profile (marked by a cross) of spot marked by box number 3 in Fig. 2.

7. Conclusions

In this paper a novel technique for accurate Bragg spot integrated intensity calculation is presented. The method is based on the two-stage RBF network that is efficiently used to approximate complex two-dimensional Bragg spot shapes. It overcomes the limitation of conventional techniques that use pre-determined reference profiles to estimate the spot integrated intensities. The use of RBF is

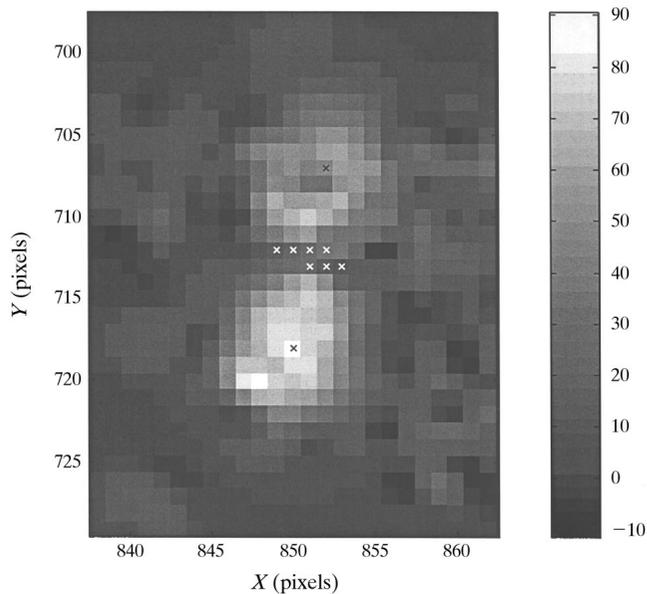


Figure 14

Two overlapping Bragg spots with pixels in overlapping region indicated by white crosses and reference profile centres marked by dark crosses.

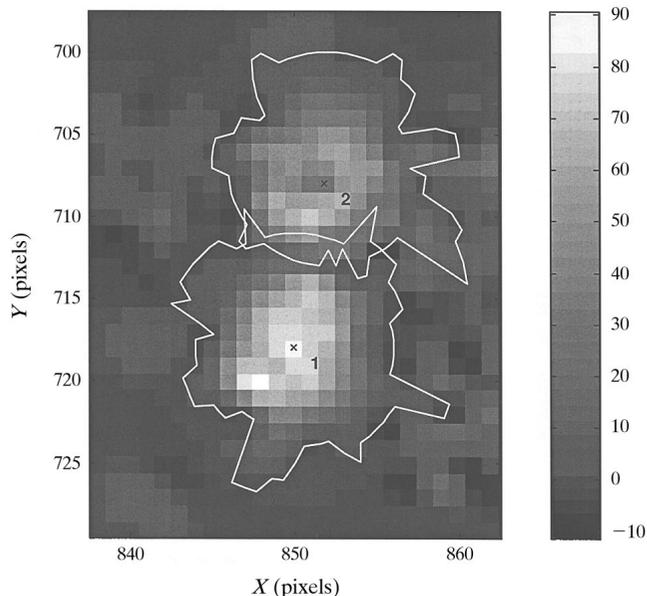


Figure 15

Calculated boundary lines of two overlapping spots marked by box number 4 in Fig. 2.

surprisingly computationally efficient. In a practical implementation the additional sophistication of this approach would be applied to any weak reflections, with estimation of the underlying noise PDF probably not calculated for every slice.

This work was undertaken as part of a programme of crystallographic detector development funded by the Foresight Initiative IMPACT grant (managed by PPARC as grant reference GR/L29576) and the Daresbury Laboratory Detector Upgrade Initiative (managed by EPSRC as grant reference GR/M39183).

References

- Allinson, N. M. (1994). *J. Synchrotron Rad.* **1**, 54–62.
- Allinson, N. M., Moon, K. J., Helliwell, J. R., Weisgerber, S., Thompson, A. M. & Colapietro, M. (1991). *Proceedings of the European Workshop on X-ray Detectors for Synchrotron Radiation Sources*, Aussois, France, 30 September–4 October 1991, edited by A. H. Walenta, pp. 156–162. Siegen, Germany: Zess.
- Bourgeois, D. (1999). *Acta Cryst.* **D55**, 1733–1741.
- Chen, S., Mulgrew, B. & McLaughlin, S. (1992). *Proceedings of the 3rd IEEE International Conference on Communications*, pp. 1267–1271. Chicago: IEEE.
- Craven, P. & Wahba, E. (1979). *Numer. Math.* **31**, 377–403.
- Cruickshank, D. W. J., Helliwell, J. R. & Moffat, K. (1991). *Acta Cryst.* **A47**, 352–373.
- Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). *J. R. Stat. Soc.* **39**, 1–38.
- Durbin, S. M. & Gog, T. (1989). *Acta Cryst.* **A45**, 132–141.
- Golub, G. H., Heath, M. & Wahba, G. (1979). *Technometrics*, **21**, 215–223.
- Hall, M. M., Veeraraghavan, V. G., Rubin, H. & Winchell, P. G. (1977). *J. Appl. Cryst.* **10**, 66–68.
- Haykin, S. (1994). *Neural Networks: A Comprehensive Foundation*. New York: Macmillan.
- Kabsch, W. (1988). *J. Appl. Cryst.* **21**, 916–924.
- Leslie, A. G. W. (1999). *Acta Cryst.* **D55**, 1696–1702.
- Lowe, D. & Webb, A. R. (1990). *Network*, **1**, 299–323.
- Messerschmidt, A. & Pflugrath, J. W. (1987). *J. Appl. Cryst.* **20**, 306–315.
- Ng, K. & Lippmann, K. (1990). *Advances in Neural Information Processing Systems*, Vol. A3, edited by R. P. Lippmann, J. E. Moody & D. S. Touretzky, pp. 970–976. San Mateo: Morgan Kaufmann.
- Orr, M. J. L. (1996). *Introduction to Radial Basis Function Networks*. Internal Report, University of Edinburgh, UK.
- Pflugrath, J. W. (1999). *Acta Cryst.* **D55**, 1718–1725.
- Pokrić, B., Allinson, N. M., Bergström, E. T. & Goodall, D. M. (1999). *IEE Proc. Vis. Image Signal Process.* **146**, 297–305.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (1995). *Numerical Recipes in C*, 2nd ed. Cambridge University Press.
- Proakis, J. G., Rader, C. M., Ling, F. & Nikias, C. L. (1992). *Advanced Digital Signal Processing*. New York: Macmillan.
- Read, R. J. (1999). *Acta Cryst.* **D55**, 1759–1764.
- Ross, S., Naday, I., Kanyo, M., Westbrook, M. L. & Westbrook, E. M. (1998). *Proc. SPIE*, **2415**, 18902.

- Rossmann, M. G. (1979). *J. Appl. Cryst.* **12**, 225–238.
- Saha, J., Christian, D. S. & Tang, C. L. (1990). *Advances in Neural Information Processing Systems*, Vol. A3, edited by R. P. Lippmann, J. E. Moody & D. S. Touretzky, pp. 728–734. San Mateo: Morgan Kaufmann.
- Sanchez-Bajo, F. & Cumbreira, F. L. (1999). *J. Appl. Cryst.* **32**, 730–735.
- Thompson, P., Cox, D. E. & Hastings, J. B. (1987). *J. Appl. Cryst.* **20**, 79–83.
- Vaseghi, S. V. (1996). *Advanced Signal Processing and Digital Noise Reduction*. Stuttgart: Wiley Teubner.
- Westbrook, E. M. (1988). Conceptual Design Report Document Number J9001-2001-SA-01. Argonne National Laboratory, Argonne, Illinois, USA.