THE ROYAL SOCIETY

# Characterization of structures from X-ray scattering data using genetic algorithms

By Matthew Wormington, Charles Panaccione, Kevin M. Matney and D. Keith Bowen

*Bede Scientific Incorporated, 14 Inverness Drive East, Suite G-104, Englewood, CO 80112, USA*

We have developed a procedure for fitting experimental and simulated X-ray reflectivity and diffraction data in order to automate and to quantify the characterization of thin-film structures. The optimization method employed is a type of genetic algorithm called 'Differential Evolution'. The method is capable of rapid convergence to the global minimum of an error function in parameter space even when there are many local minima in addition to the global minimum. We show how to estimate the pointwise errors of the optimized parameters, and how to determine whether the model adequately represents the structure. The procedure is capable of fitting some tens of adjustable parameters, given suitable data.

**Keywords: X-ray scattering; data fitting; genetic algorithms; Differential Evolution**

## 1. Introduction

Inverse problems are commonplace in the physical and engineering sciences. If we know, or assume, the structure of an object, then we can often calculate many of its effects. However, the inverse problem of going directly from the effects to determine their cause is often intractable. Unfortunately, in many cases all we have are the effects such as certain experimental measurements. In the case of X-ray scattering, we would require both the amplitude and phase for all scattering angles in order to determine directly the structure of the scattering object. However, we can usually measure only the intensity over a limited range of scattering angles. Hence, we must usually resort to some indirect route, even though this must imply some loss of uniqueness in the deduced structure.

A typical way to determine the structure of an object from its X-ray scattering is to construct a model that we hope reasonably describes its structure and from which we can simulate the X-ray scattering. Using the model, we simulate the X-ray scattering, for example the X-ray reflectivity or diffraction curve, and calculate the difference between the experimental and simulated curves using some error function, $E$. The model is then adjusted by some optimization method in order to minimize the difference between the two curves. This procedure is repeated until the difference between the two curves is judged to be sufficiently small, at which point we accept the model to be an accurate representation of the structure.

The field of data fitting and parameter optimization has a long and fruitful history. The earliest successes were for linear problems that possessed a single minimum in the error function. The mean-squared difference between the experimental and

 

simulated data was commonly used as the error function because of its computational simplicity in the days before fast digital computers. More recent research has focused on nonlinear problems, and on those with local minima in the error function in addition to the global minimum. A variety of data-fitting and parameter-optimization strategies has been developed for such systems (Bevington 1969; Press *et al*. 1989), and the following are the most commonly encountered.

(1) *Direct search*. The parameter space is divided up into small, but finite, regions. The error function is calculated for each region and the region that gives the smallest value for $E$ is said to give the best-fit (optimum) parameter values.

(2) *Downhill simplex*. An initial guess at the parameter values is made. The simplex (a geometrical construction) then moves in directions that decrease the value of $E$. The parameters that yield the smallest value of $E$ in the neighbourhood of the initial guess are said to be the best-fit parameters.

(3) *Levenberg–Marquardt method*. An initial guess at the parameter values is made by the user. The algorithm then combines linearization and gradient searching of the error function to minimize $E$ in the neighbourhood of the initial guess. The parameter values giving the smallest value for $E$ are then selected as the best-fit parameters.

(4) *Monte Carlo method*. The parameter space is again divided into small regions. Regions are selected at random and the error function is evaluated. After a certain number of regions have been chosen, or when $E$ is smaller than some specified value, the algorithm is stopped. The region with the smallest value for $E$ is said to yield the best-fit parameter values.

(5) *Simulated annealing*. This uses the physical principles governing annealing (i.e. the slow cooling of a liquid so that it forms a crystal) to search for the error function and obtain the best-fit parameters. There is a finite probability in any step that the parameters can move in a direction so as to increase $E$, so the method does escape from local minima, but slowly.

All of the above methods run into severe difficulties when fitting X-ray scattering data such as X-ray reflectivity or diffraction curves. The parameter space is simply too vast for direct searches and becomes uncomputable for all but the simplest cases. The downhill simplex and Levenberg–Marquardt methods work well for nonlinear problems because they are guided by the geometry of the error function in parameter space. However, the initial estimate of the parameter values needs to be very close to the optimized values if local minima are present, as they will become trapped in the first local minimum that they encounter. These two methods are therefore only effective when the parameters are initially contained within the multidimensional 'well' of the global minimum, and in most practical cases in X-ray scattering we have found them to be of little use. The Monte Carlo and simulated annealing methods do not get trapped in local minima. However, they are very inefficient at searching the parameter space, since they search it randomly without taking into account the geometry of the error function. We consider that the practical conditions for an optimization to be guided by the geometry of the error function are as follows.

(1) The problems fall into the class of NP-complete problems (Cook 1971). NP (non-deterministic polynomial) problems are those in which the number of possible solutions is uncomputably large, perhaps even infinite in principle, but the solutions can be guessed and verified rapidly.

(2) The error function should not have a constant value over large regions of parameter space.

(3) The error function needs to be continuous and single-valued in parameter space.

Condition (1) expresses the requirement that we should recognize the solution when found, but it is not in itself sufficient. Password-breaking problems are of this type, but an erroneous password simply gives a wrong answer, i.e. a constant value in parameter space, hence condition (2) is also necessary. Conditions (2) and (3) also express the form of the error function. A successful strategy for nonlinear problems containing local minima will combine both random and guided elements.

Genetic algorithms (GAs), of which there are many classes, offer remarkable potential in solving such problems and were established in 1975 by John Holland of the University of Michigan (Holland 1975). Evolutionary algorithms (EAs), one particular class of GAs, apply some of the known mechanisms of evolution to solve optimization problems. A number of EAs have been suggested, but they all share a common conceptual basis of simulating the evolution of parameter vectors by a repeated process of mutation, reproduction and selection. The algorithms are intrinsically parallel, maintaining a population (set) of parameter vectors, and the ability to explore many different parts of parameter space simultaneously is key to their success. Simplistically, mutation and diversity find the regions that contain minima while inheritance searches these regions to find the minimum.

We have employed a new EA called 'Differential Evolution' (Storn & Price 1995; Price & Storn 1997). Through simple mutation, recombination and selection schemes, parameter vectors with better 'fitness' (i.e. parameters that yield a small value of the error function, $E$) are found. Mutation is an operation that makes small random changes to one or more of the population vectors. Mutation is critical for maintaining diversity in the population of parameter vectors. Recombination is a complementary operation that creates parameter vectors (offspring) by combining two parameter vectors from the previous generation (parents) and helps focus the search on promising regions of the parameter space. Selection guarantees that the 'fittest' parameter vectors will propagate in future generations. EAs differ from the conventional parameter optimization methods listed above in several important ways.

(1) EAs optimize the trade-off between exploring new points in the parameter space (mutation) and exploiting the information discovered thus far (recombination).

(2) EAs operate on many solutions simultaneously (implicit parallelism), gathering information from current search points to direct the search. Their ability to maintain multiple solutions concurrently makes EAs less susceptible to the problems associated with local minima and noise.

(3) EAs are randomized algorithms, in that they use operators whose results are governed by probability, but they do perform purely random searches (in contrast to the Monte Carlo algorithms).

There can never be an absolute guarantee that any algorithm will find the globally optimized parameter values in a finite time. However, EAs in general and Differential Evolution (DE) in particular, do appear to be one of the most efficient algorithms yet found for this process (Storn & Price 1995).

## 2. The data-fitting procedure

### (*a*) *The Differential Evolution (DE) algorithm*

Let us assume that the experimental data contain $N$ measured points $(\theta_j, I_j)$, where $\theta_j$ is the incidence angle, $I_j$ is the intensity measured at $\theta_j$ and $j = 1, 2, \ldots, N$. Simulated data $I(\theta_j; \boldsymbol{p})$ are computed assuming a structural model with $n$ adjustable parameters represented by the vector $\boldsymbol{p} = [p_1, p_2, \ldots, p_n]$ and are compared with the experimental data using some error function $E(\boldsymbol{p})$. Guided by $E(\boldsymbol{p})$, the DE algorithm attempts to optimize the parameter vector $\boldsymbol{p}$ starting with an initial population of randomly generated parameter vectors, by a repeated cycle of mutation, recombination and selection.

A detailed flowchart for the DE algorithm used in this work is shown in figure 1. Parameters are stored in an array large enough to hold a population (set) of parameter vectors $\boldsymbol{P} = [\boldsymbol{p}_0, \boldsymbol{p}_1, \ldots, \boldsymbol{p}_{m-1}]$. The size of the population, $m$, is selected by the user and in this work we typically use $m = 10 \times n$. The parameter vector, $\boldsymbol{p}_0$, is initialized from the user's initial guess at the structure, while the remaining $m - 1$ vectors are initialized by assigning each parameter with a randomly chosen value from within its allowed range. Once all of the parameter vectors have been initialized, the error function for each $\boldsymbol{p}_i$ is evaluated. The parameter vector with the lowest error is stored in the best-fit vector $\boldsymbol{b} = [b_1, b_2, \ldots, b_n]$. This vector is used to track the progress of the optimization and is updated whenever an equal or better solution than the best-so-far vector is found.

The crucial idea in DE is its simple scheme for creating new population members. Two randomly selected vectors, $\boldsymbol{p}_a$ and $\boldsymbol{p}_b$, are chosen from the current population. The difference vector $(\boldsymbol{p}_a - \boldsymbol{p}_b)$ is then used to mutate the best-so-far vector, $\boldsymbol{b}$, according to the relation

$$\boldsymbol{b}' = \boldsymbol{b} + k_{\mathrm{m}}(\boldsymbol{p}_a - \boldsymbol{p}_b), \tag{2.1}$$

where $k_{\mathrm{m}}$ denotes the mutation constant. The value of $k_{\mathrm{m}}$ must be empirically selected by the user to give fast convergence and in this work we have used $k_{\mathrm{m}} = 0.7$. As the evolving population vectors converge, the differences between them diminish and, hence, the difference vector remains scaled to an appropriate size.

With $\boldsymbol{b}'$ in hand, a trial vector $\boldsymbol{t} = [t_1, t_2, \ldots, t_n]$, which competes with the vector $\boldsymbol{p}_0$, is assembled. Starting with the randomly chosen $j$th parameter, the trial parameters $t_j$ are consecutively loaded (modulo $n$) from either $\boldsymbol{b}'$ or $\boldsymbol{p}_0$. A binomial distribution is used to decide which parameters come from $\boldsymbol{b}'$ and which come from $\boldsymbol{p}_0$. A random number chosen from a uniform $(0, 1)$ distribution is compared with a user-selected recombination constant $k_{\mathrm{r}}$. If the random number is less than or equal to $k_{\mathrm{r}}$, then $t_j$ is loaded with the $j$th parameter from $\boldsymbol{b}'$. If the random number is greater than $k_{\mathrm{r}}$, then the $j$th parameter of $\boldsymbol{t}$ is loaded from $\boldsymbol{p}_0$. In this work, we have used $k_{\mathrm{r}} = 0.5$. After $n - 1$ trials, $\boldsymbol{t}$ gets its final parameter from $\boldsymbol{b}'$, so that at least one parameter of $\boldsymbol{t}$ is different from $\boldsymbol{p}_0$. With the vector $\boldsymbol{t}$ assembled, any constraints

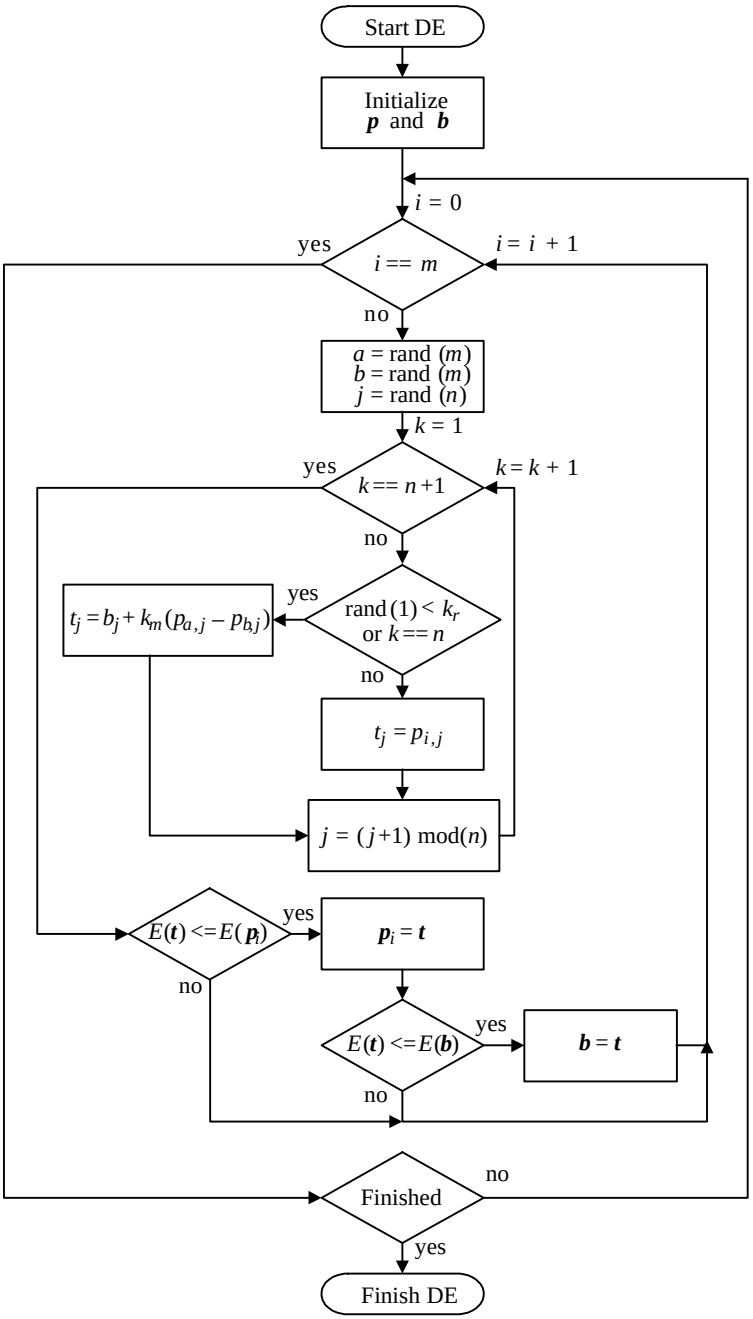


Figure 1. Flowchart for the Differential Evolution (DE) algorithm.

are then taken into account. If the value of the trial parameter $t_j$ falls outside the specified constraints, it is replaced by a randomly selected value according to the expression

$$t_j = p_j^{\min} + \mathrm{rand}(p_j^{\max} - p_j^{\min}), \tag{2.2}$$

where $p_j^{\min}$ and $p_j^{\max}$ denote the minimum and maximum permissible values of the $j$th parameter, respectively. If the vector $\boldsymbol{t}$ satisfies the inequality

$$E(\boldsymbol{t}) \leqslant E(\boldsymbol{p}_0), \tag{2.3}$$

then $\boldsymbol{t}$ is selected to replace $\boldsymbol{p}_0$, otherwise $\boldsymbol{p}_0$ propagates to the next generation. The procedure is then repeated for all remaining parameter vectors in the population $\boldsymbol{P}$, that is $\boldsymbol{p}_i$ with $i = 1, 2, \ldots, m-1$, with a new vector $\boldsymbol{b}'$ calculated each time. Finally, the algorithm is iterated over many generations until $E(\boldsymbol{b})$ fails to decrease, i.e. until the best-fit vector $\boldsymbol{b}$ has converged on the global minimum of the error function.

### (*b*) *The error function*

The choice of an appropriate error function is crucial for any data-fitting procedure regardless of the optimization method used. The DE algorithm gives us a great deal of flexibility in this choice since we need only choose a continuous function and do not require the function to have continuous derivatives. When fitting X-ray scattering data, the error function should have the following additional properties:

(1) a single deep global minimum;

(2) local minima that are much less deep than the global minimum;

(3) be fast and simple to calculate;

(4) have relative insensitivity to the absolute magnitude of the data, since X-ray scattering data often spans many orders of magnitude; and

(5) does not overemphasize outlying points in the experimental data, since we expect a Poisson distribution of statistical noise.

Point (4) suggests that a logarithmic function could be appropriate since it linearizes data spanning several orders of magnitude. Point (5) suggests that a robust error function (Press *et al*. 1989) will be one that is more suitable than the mean-squared error function commonly encountered in least-squares fitting. To confirm these conjectures, we have investigated a number of error functions that have been applied to fitting problems, as follows.
Mean-square error of the data (MSE),

$$E(\boldsymbol{p}) = \frac{1}{N-1}\sum_{j=1}^{N}[I_j - I(\theta_j;\boldsymbol{p})]^2. \tag{2.4}$$

Mean-absolute error of the data (MAE),

$$E(\boldsymbol{p}) = \frac{1}{N-1}\sum_{j=1}^{N}|I_j - I(\theta_j;\boldsymbol{p})|. \tag{2.5}$$

Mean-square error of the log transformed data ($\mathrm{MSE}_{\log}$).

$$E(\boldsymbol{p}) = \frac{1}{N-1}\sum_{j=1}^{N}[\log I_j - \log I(\theta_j;\boldsymbol{p})]^2. \tag{2.6}$$

Mean-absolute error of the log transformed data ($\mathrm{MAE}_{\log}$),

$$E(\boldsymbol{p}) = \frac{1}{N-1}\sum_{j=1}^{N} |\log I_j - \log I(\theta_j;\boldsymbol{p})|. \tag{2.7}$$

Tests of these functions with X-ray reflectivity and high-resolution X-ray diffraction data showed that the non-logarithmic error functions (2.4) and (2.5), as expected, did not effectively fit the data at low intensities. These occur at large scattering angles and contain information on the smallest length-scales present in the structure, which are often those at which the X-ray characterization is aimed. The error functions (2.6) and (2.7) both could cope adequately with such data, but equation (2.7) is preferred because of its lower sensitivity to outlying data points (due mainly to statistical noise in the experimental data). Clearly, we cannot assert that equation (2.7) is the best possible error function but it is very effective, and this is sufficient. We therefore now use the $\mathrm{MAE}_{\log}$ function for fitting all such data.

We note, however, that a non-logarithmic function is useful for fitting other types of X-ray scattering data such as those obtained in powder X-ray diffraction measurements in which the most important information is contained in the higher intensities. In such cases we prefer the MAE function as it is less sensitive to statistical noise than the MSE function.

### (*c*) *Estimation of convergence and model fitness*

We would like the error function to tell us how close the model is to the structure of the sample being characterized. However, what it gives us is a number that measures the difference between the experimental and simulated X-ray scattering data. We can therefore ask the following two questions.

(1) Is there a minimum value of the error function for (noisy) experimental data and a simulation based on the correct model?

(2) Can we predict it?

If we can answer these questions positively, then we can provide an automatic method for stopping the data-fitting procedure and can also assess the accuracy of the model. We have the advantage over many problems in that we know from more than 20 years of work by numerous researchers that simulating X-ray scattering data using dynamical theory is extremely accurate. Furthermore, we know that the statistical noise on X-ray generation and detection is accurately described by a Poisson distribution.

Let us assume that the structure of the sample is accurately described by the model, that there are no errors associated with the simulation of the X-ray scattering and that there are no systematic errors present in the experimental data. Then, in the absence of statistical noise, $E(\boldsymbol{p})$ is expected to equal zero, that is the experimental and simulated data are identical and represented by $I(\theta_j;\boldsymbol{p})$. Now, let us introduce noise into the simulation such that at any point

$$I'(\theta_j;\boldsymbol{p}) = I(\theta_j;\boldsymbol{p}) + \Delta I_j, \tag{2.8}$$

where $\Delta I_j$ is a random deviate selected from a Poisson distribution appropriate for a given incident intensity, background intensity and measurement time. By setting

$I_j = I'(\theta_j; \boldsymbol{p})$ in equation (2.7), we may calculate the value of $E(\boldsymbol{p})$ due to statistical noise in a single realization of $I'(\theta_j; \boldsymbol{p})$. Repeating this procedure, say, 100 times for different realizations of the noisy simulation $I'(\theta_j; \boldsymbol{p})$ allows us to estimate the mean, $E_{\min}$, and standard deviation, $\Delta E_{\min}$, of the minimum value of $E(\boldsymbol{p})$ expected in the presence of statistical noise. Looked at another way, $E_{\min}$ also gives the minimum error-function value that can be expected if the fitting procedure has converged to the correct structural model given noisy experimental data. We therefore stop the fitting procedure when the error function satisfies the inequality

$$E(\boldsymbol{p}) < E_{\min} + 3\Delta E_{\min}. \tag{2.9}$$

If the fitting procedure has converged, in the sense that $E(\boldsymbol{p})$ is no longer decreasing but does not satisfy equation (2.9), this indicates a deficiency in the model. If, however, the error function satisfies equation (2.9), there is no need to test separately for convergence; this is as good a model as the experimental data will sustain. There may of course be features in the structure that are not revealed by fitting the experimental and simulated data, and tests for this will now be discussed.

### (*d*) *Estimation of parameter errors*

Any data-fitting procedure will result in best-fit parameter values, which can be reported to as many decimal places as the programmer decides. However, it is essential to assess the accuracy and reliability of these values. It is a remarkably difficult and computationally intensive problem to calculate the errors in each parameter with respect to all other parameters for all but the simplest of problems (Press *et al.* 1989). We therefore compromise by calculating the pointwise errors of the parameters, taken around their best-fit values. The pointwise error is defined as the change $\Delta p_j$ of the parameter $p_j$ that increases the error function $E(\boldsymbol{p})$ by a specified amount, with all other parameters kept constant at their best-fit values. We typically specify this increase to be 5% of the error-function value after the fitting procedure has converged since it yields easily discernible changes in the simulated curve. The change $\Delta p_j$ is explicitly calculated for both signs of the deviation from $p_j$, since the error function can be asymmetric. It is then obvious which parameters are well determined by the fitting procedure, and which are poorly determined. Since in our programs a simulation may also be compared with a 'noisy' simulation (i.e. statistical noise included), we may also use this feature to design an experiment that will better determine the latter parameters.

An interesting consequence of the discussion on estimating the parameter errors is that it is possible to use the fitting procedure to help 'discover' new features in a structural model. If, for example, an additional layer is suspected, it can be inserted into the model and the fitting procedure repeated. Comparing the converged error-function value with equation (2.9) and calculating the pointwise parameter errors will help determine whether such 'discoveries' are appropriate.

### (*e*) *Smoothing*

A variety of methods for smoothing experimental data have been developed, such as FFT filtering, $N$-point averaging and Savitsky–Golay polynomial smoothing. If any of these methods are applied to experimental data prior to fitting, then a lower

error function naturally results after the fitting procedure converges. However, we do not recommend such a treatment for two reasons. First, any smoothing method will distort the data in some way and thus may change the best-fit parameter values. Second, smoothing the data voids use of the minimum error-function estimate, which is very useful in testing the accuracy of the model as well as the goodness-of-fit.

### (*f*) *Performance*

The performance of the data-fitting procedure is primarily affected by the following four factors.

(1) The quality and size of the experimental data. If the experimental data are noisy or contain a very large number of points, it will take longer to determine the best-fit parameter values.

(2) The quality of the initial estimates for the parameters. If the initial values for the parameters are grossly different from the optimized values, the fitting procedure will take longer to converge.

(3) The search range of the parameter values. If a large search range is specified, the fitting procedure may take longer to converge to the global minimum of the error function. However, because the DE algorithm is rather good at finding the global minimum, without becoming trapped in local minima, it tends not to falsely converge to incorrect values for the parameters.

(4) The number of adjustable parameters. The ability of the procedure to determine the optimum parameter values decreases as the number of parameters increases. In practice we find that up to ten parameters can be optimized in a matter of minutes, and that several tens of parameters can be optimized during an overnight run.

We have developed efficient programs for fitting X-ray reflectivity curves and high-resolution X-ray diffraction rocking-curves based on the fitting procedure presented in this section. The programs run on personal computers (PCs) under the Microsoft Windows 95, 98 and NT operating environments. All benchmarks reported in this work assume that the programs are run on a 300 MHz Pentium II-based PC fitted with 128 Mb of memory.

## 3. Experimental

X-ray reflectivity curves and high-resolution X-ray diffraction rocking-curves were measured using a Bede D1 diffractometer equipped with a Cu-anode sealed X-ray tube (2.2 kW). The radiation was collimated and monochromated using a single Si(022) asymmetric channel-cut crystal (CCC) and a $0.05 \times 10\,\mathrm{mm}^2$ slit placed *ca.* 50 mm after the CCC. This arrangement produced an intense beam (greater than $5 \times 10^6$ cps operating the X-ray tube at 40 kV and 35 mA) of CuK$\alpha_1$ radiation incident on the sample. A $0.5 \times 10\,\mathrm{mm}^2$ slit was placed immediately before the detector so as to reduce the background noise. The X-ray reflectivity curves were measured by performing $(\theta, 2\theta)$-scans in which the sample and detector axes are coupled in a ratio of 1:2. The X-ray diffraction rocking-curves were similarly measured by performing $(\theta, 2\theta)$-scans around the (004) Bragg reflection.

## 4. Application to X-ray reflectivity

X-ray reflectivity is a technique used to characterize the surface structure of materials irrespective of their crystalline perfection. Hence, the technique can be applied equally well to crystalline, polycrystalline and amorphous materials, and provides accurate information about the thickness, roughness and density in thin-film structures (Bowen & Wormington 1993). In this section we describe the structural model and simulation methods used in our X-ray reflectivity fitting program and give three examples of its application.

### (*a*) *The model and simulation*

The simulation method used is taken from the Bede REFS program (Wormington *et al*. 1992). We will consider a multilayer on a thick substrate in which the refractive index of each layer is assumed constant. For X-rays, the refractive index of a material is slightly less than unity and can be written as

$$n = 1 - r_{\mathrm{e}}\frac{\lambda^2}{2\pi}\sum_a (f_{\mathrm{a}} + f'_{\mathrm{a}} + f''_{\mathrm{a}})N_a, \tag{4.1}$$

where $r_{\mathrm{e}}$ is the classical electron radius and $\lambda$ is the X-ray wavelength. The atomic scattering factor is denoted by $f_{\mathrm{a}}$ and the real and imaginary parts of the dispersion correction are $f'_{\mathrm{a}}$ and $f''_{\mathrm{a}}$, respectively. Values for the scattering factor and its corrections are tabulated in the *International tables for crystallography* (Ibers & Hamilton 1974). The summation is taken over all constituent atoms, $a$, of number density $N_a$.

The amplitude ratio $X_j = E_{\mathrm{r},j}/E_{\mathrm{t},j}$ of the reflected and transmitted waves at the bottom of layer $j$ within the multilayer is obtained by solving Maxwell's equations and the appropriate boundary conditions. According to Parratt (1954), we may write

$$X_j = \frac{r_j + X_{j+1}\varphi_{j+1}^2}{1 + r_j X_{j+1}\varphi_{j+1}^2}, \tag{4.2}$$

where $r_j$ is the Fresnel coefficient for reflection from the interface between layers $j$ and $j+1$. For a sharp interface, $r_j$ is given by the expression

$$r_j = \frac{k_{z,j} - k_{z,j+1}}{k_{z,j} + k_{z,j+1}}, \tag{4.3}$$

where $k_{z,j} = 2\pi/\lambda(n_j^2 - \cos^2\theta)^{1/2}$ is the component of the wavevector in layer $j$ perpendicular to the surface of the multilayer (i.e. along the $z$-axis), $n_j$ is the refractive index of the layer and $\theta$ is the grazing angle of the incident plane wave. The complex phase factor for wave propagation through the layer thickness, $t_j$, is denoted by $\varphi_j = \exp(\mathrm{i}k_{z,j}t_j)$. To include the effects of grading (interdiffusion) and roughness within this formalism we need only modify the form of the Fresnel coefficient. From the work of Névot & Croce (1980) an appropriate modification is given by

$$r_j = \frac{k_{z,j} - k_{z,j+1}}{k_{z,j} + k_{z,j+1}} \exp[-2(k_{z,j}k_{z,j+1})^{1/2}\sigma_{j+1}], \tag{4.4}$$

where $\sigma_{j+1}$ denotes the width of the interface between layers $j$ and $j+1$ due to both grading and roughness.

To calculate the amplitude ratio at the top of the multilayer, $X_0$, equations (4.2) and (4.4) are applied recursively for all interfaces starting at the substrate (layer $N+1$), where $X_N = r_N$. The plane-wave reflectivity is then given by $R = |X_0|^2$ and is related to the reflected intensity through the correlation function

$$I(\theta; \boldsymbol{p}) = I_0 \int F(\theta') R(\theta' - \theta)\, \mathrm{d}\theta' + I_\mathrm{b}, \tag{4.5}$$

where $I_0$ and $I_\mathrm{b}$ denote the incident and background intensity, respectively. Here $F(\theta)$ denotes an instrument function and takes into account the finite divergence of the incident X-ray beam. We have included both this incident angle $\theta$ and the adjustable parameters $\boldsymbol{p}$ into our notation. Specifically $\boldsymbol{p}$ contains the following:

(1) the incident intensity $I_0$;

(2) the background intensity $I_\mathrm{b}$;

(3) the densities $\rho_j$ of the layers $j = 1, 2, \ldots, N$;

(4) the thicknesses $t_j$ of the layers $j = 1, 2, \ldots, N$; and

(5) the widths $\sigma_{j+1}$ of the interfaces between layers $j$ and $j+1$.

Finally, we note that X-ray reflectivity measurements cannot usually distinguish between layers of high atomic number $Z$ and low mass density, and those with low $Z$ and high density. We have therefore chosen to fit the density of layers in the structural model and assume their chemical composition.

### (*b*) *Example I.* Ta *layer on* $Al_2O_3$

In our first example, we consider a Ta (10 nm) layer deposited on an $Al_2O_3$ substrate. Figure 2*a* shows the measured and simulated X-ray reflectivity curves before fitting. Below the critical angle of the Ta layer, $\theta_\mathrm{c} \sim 0.5°$, the intensity is very high as a result of total external reflection. Above $\theta_\mathrm{c}$, the reflected intensity decreases rapidly and prominent oscillations (Kiessig fringes) are clearly visible due to the large difference in the refractive index of the Ta layer and that of the $Al_2O_3$ substrate. The period of the Kiessig fringes is related to the thickness of the Ta layer. Figure 2*b* shows the measured curve together with its best-fit simulation. The time for the fitting procedure to converge was less than two minutes, fitting a total of nine adjustable parameters. The best-fit parameter values and their uncertainties are given in table 1. It should be noted that a surface oxide layer (assumed to be $Ta_2O_5$) had to be included in the structural model to obtain close agreement of the measured and simulated curves.

The parameter values for the simulated curve shown in figure 2*a* were chosen to be far from the anticipated best-fit parameter values. This was a deliberate choice in order to demonstrate that the fitting procedure rapidly converges to the global minimum in the error function without getting trapped in local minima. The progress of the fitting procedure is illustrated in figure 3, which shows the value of the error function versus the number of generations (iterations of the DE algorithm). Horizontal sections are times during which the fitting procedure is temporarily in a local minimum. The fitting procedure is seen to have converged to the global minimum

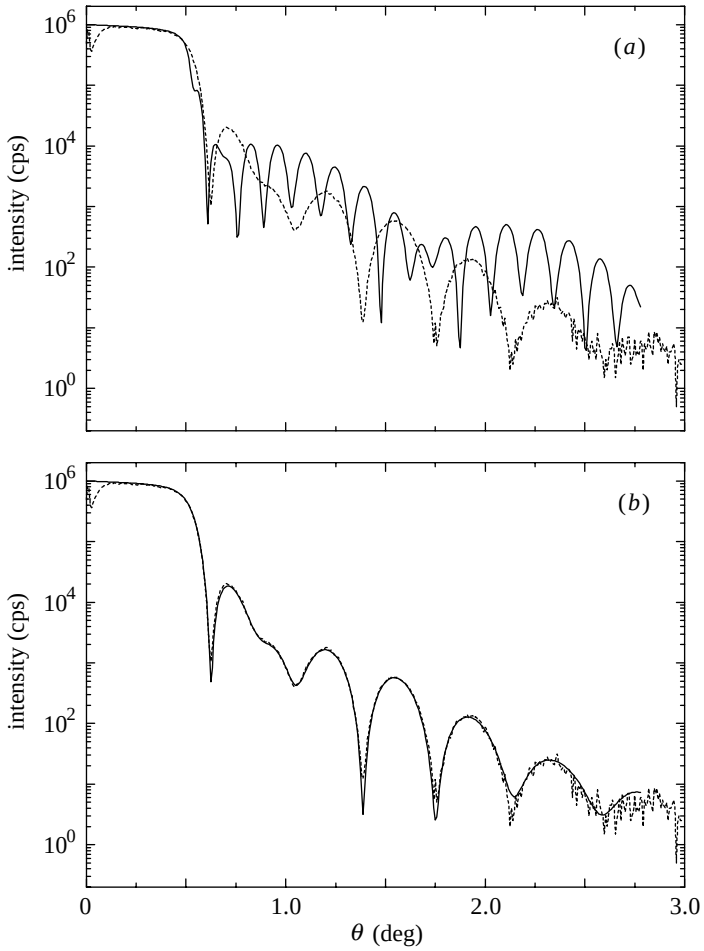


Figure 2. Comparison of experimental and simulated X-ray reflectivity curves for a Ta layer on $Al_2O_3$, (*a*) before and (*b*) after the fitting procedure has converged. The dashed lines represent the measurements and the solid lines are the simulations.

Table 1. *Best-fit parameter values for the* Ta *layer on* $Al_2O_3$

| layer | material | $t$ (nm) | $\sigma$ (nm) | $\rho$ (g cm$^{-3}$) |
|---|---|---|---|---|
| 2 | $Ta_2O_5$ | $2.70 \pm 0.05$ | $0.71 \pm 0.03$ | $8.6 \pm 0.2$ |
| 1 | Ta | $10.49 \pm 0.02$ | $0.45 \pm 0.02$ | $16.1 \pm 0.2$ |
| substrate | $Al_2O_3$ | $\infty$ | $0.38 \pm 0.02$ | 3.99 |

after only 1000 generations. Figure 4 shows the value of the error function versus the thickness of the Ta layer, with all other parameters held at their best-fit values. We note that the error function has a single deep global minimum and many local minima. Harmonic minima, which occur at half and twice the best-fit Ta layer thickness, are the deepest of the local minima. The global minimum is 'shielded' by fairly large maxima on either side. This is a typical feature in such curves when the thickness is

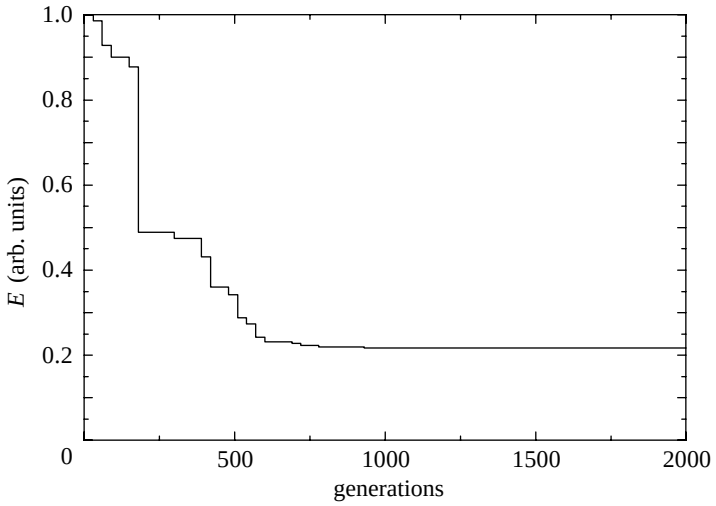


Figure 3. Variation of the error function, $E$, with the number of DE generations. The fitting procedure has converged after approximately 1000 generations.

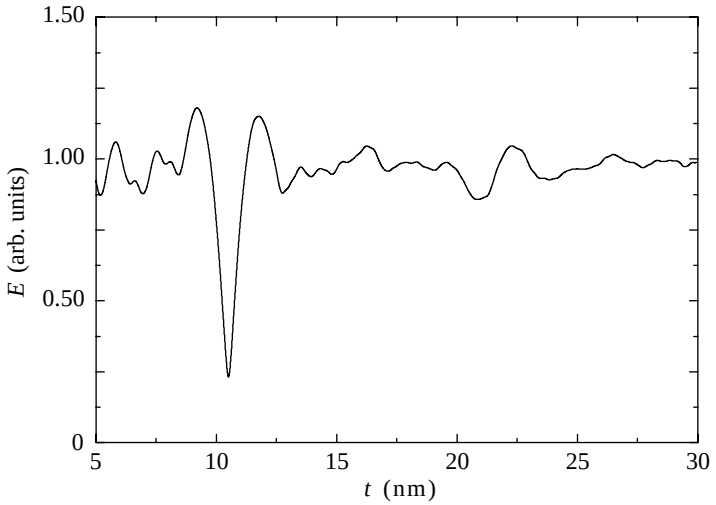


Figure 4. Variation of the error function, $E$, with the Ta layer thickness. All other adjustable parameters in the model are held constant at their best-fit values.

varied, and turns out to be caused by the beating of two sets of oscillations (Kiessig fringes) in which one period is fixed and the other is variable. This characteristic shape is very useful for recognizing whether the global minimum is in fact within the range specified for the thickness parameters in question.

### (*c*) *Example II.* $GaAs/Al_{0.3}Ga_{0.7}As$ *layers on* GaAs

For our next example we consider an $Al_{0.3}Ga_{0.7}As$ (50 nm) layer capped with a GaAs (50 nm) layer grown on a GaAs substrate. Figure 5*a* shows the measured and simulated X-ray reflectivity curves prior to fitting. We see that the Kiessig fringes are far less prominent than in the previous example because the refractive index

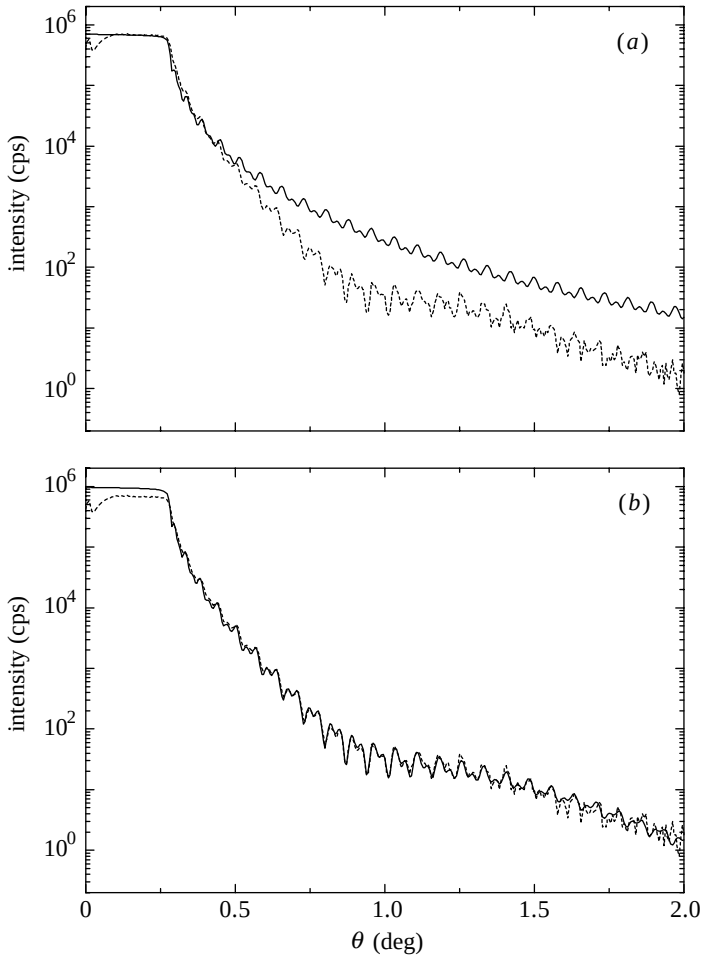


Figure 5. Experimental and simulated X-ray reflectivity curves for an $Al_xGa_{1-x}As$ layer capped with GaAs on a GaAs substrate, (*a*) before and (*b*) after fitting. The dashed lines are the measurements and the solid lines represent the simulations. Experimental data courtesy of Professor B. K. Tanner (University of Durham, UK).

of $Al_{0.3}Ga_{0.7}As$ is similar to that of GaAs. Furthermore, the periods of the Kiessig fringes are much smaller than in the previous example because of the thicker layers considered. The measured curve and best-fit simulation are shown in figure 5*b*. While more iterations are required than in the previous example, the fitting procedure still took less than three minutes to converge and fit 10 adjustable parameters. The best-fit parameters and their respective uncertainties are given in table 2.

We included an additional layer of GaAs in the structural model and allowed its density to be fitted within the range 2.66–5.32 g cm$^{-3}$ (i.e. 50–100% of its bulk value) to test for the presence of a surface oxide layer. If no surface layer were present, the density of the top layer would naturally converge to 5.32 g cm$^{-3}$. However, the density converged to 3.19 g cm$^{-3}$, indicating the presence (most likely) of an oxide. This conjecture is further supported by testing the value of the error function after convergence. The theoretical minimum error-function value for this example

Table 2. *Best-fit parameter values for the* $GaAs/Al_{0.3}Ga_{0.7}As$ *layers on* GaAs

| layer | material | $t$ (nm) | $\sigma$ (nm) | $\rho$ (g cm$^{-3}$) |
|---|---|---|---|---|
| 3 | GaAs | $2.28 \pm 0.02$ | $0.57 \pm 0.02$ | $3.19 \pm 0.05$ |
| 2 | GaAs | $50.97 \pm 0.01$ | $0.64 \pm 0.03$ | 5.32 |
| 1 | $Al_{0.3}Ga_{0.7}As$ | $50.40 \pm 0.02$ | $0.5 \pm 0.1$ | 4.87 |
| substrate | GaAs | $\infty$ | $0.7 \pm 0.1$ | 5.32 |

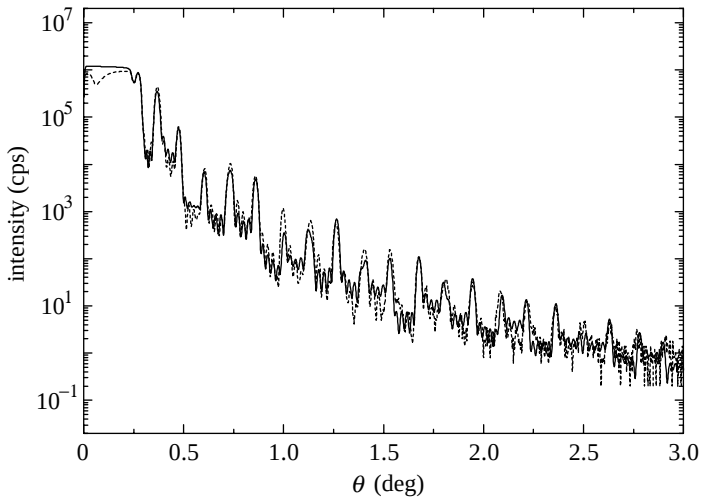


Figure 6. X-ray reflectivity curves for an $Si_{1-x}Ge_x/Si$ superlattice on an Si substrate. The dashed line represents the measurements and the solid line is the best-fit simulation.

is $E_{\min} = 0.039 \pm 0.001$ and the converged value is $E = 0.043$, satisfactorily close. However, if we omit the surface layer, the error-function value never decreases below $E = 0.2$. This example helps to show how the fitting procedure developed in this work can be used to 'discover' layers that have not been intentionally grown.

Finally, we note that detector saturation is evident in the measured data (see figure 5), but this has clearly not prevented the fitting procedure from converging. The reason for this is that we treated the incident intensity as an adjustable parameter and ignored most of the measured data in the region of total external reflection. The incident intensity is, in effect, determined from the initial slope of the reflectivity curve. However, for the most accurate characterization of a sample it is important to reduce this effect experimentally, for example by using an Al absorber to attenuate the reflected beam at very low incident angles, or a high dynamic-range detector.

### (*d*) *Example III. A* $Si_{1-x}Ge_x/Si$ *superlattice on* Si

In our last X-ray reflectivity example, we have applied our fitting procedure to a superlattice with nominal structure $[Si_{0.5}Ge_{0.5}(10\,\text{nm})/Si(22\,\text{nm})]_5$ capped with Si(10 nm), grown by molecular beam epitaxy (MBE) on a Si(001) substrate. The measured curve and its best-fit simulation are shown in figure 6. The reflectivity curves contain much fine detail; Bragg peaks and Kiessig fringes between them are clearly visible. The angular separation of adjacent Bragg peaks is related to the

Table 3. *Best-fit parameter values for the* $Si_{1-x}Ge_x$/Si *superlattice on* Si

| layer | material | $t$ (nm) | $\sigma$ (nm) |
|---|---|---|---|
| 12 | $SiO_2$ | $1.18 \pm 0.02$ | $0.19 \pm 0.01$ |
| 11 | Si | $9.31 \pm 0.02$ | $1.19 \pm 0.01$ |
| ...10 | $Si_{0.43}Ge_{0.57}$ | $9.08 \pm 0.01$ | $0.97 \pm 0.01$ |
| 1... | Si | $22.89 \pm 0.01$ | $0.43 \pm 0.01$ |
| substrate | Si | $\infty$ | $0.3 \pm 0.1$ |

superlattice period while the period of the Kiessig fringes is related to the total thickness of the structure including the Si capping layer. Despite the reasonably large number of adjustable parameters involved in this example (a total of 11), the fitting procedure still managed to converge in *ca.* 15 min. The best-fit parameters and their uncertainties are listed in table 3. With the exception of the surface oxide layer (assumed to be $SiO_2$), which had a fitted density of $1.5 \pm 0.1$ g cm$^{-3}$, all density values were fixed at their bulk values during the fitting. The Ge concentration was determined precisely by high-resolution X-ray diffraction to be $x = 57 \pm 5\%$ and also remained fixed during the fitting.

It is interesting to note that the fitting procedure automatically found an asymmetry between the SiGe-on-Si and Si-on-SiGe interfaces. The width of the latter interfaces was almost twice that of the former. This asymmetry was previously reported by Powell *et al.* (1992) and was found by transmission electron microscope (TEM) images to be due to a 'wavy' morphology at the Si-on-SiGe interfaces. It is now known to be a characteristic growth morphology in compressively strained layers, such as SiGe grown on Si.

## 5. Application to X-ray diffraction

High-resolution X-ray diffraction is widely employed to characterize the structure of semiconductor thin-film structures. The technique provides accurate information about the composition, quality, strain, thickness and tilt of crystalline layers.

### (*a*) *The model and simulation*

The simulation method employed in this work is taken from the Bede RADS program by Bowen *et al.* (1991) and subsequent developments by Wormington. The program is based on the Takagi–Taupin dynamical theory of X-ray diffraction. In the two-beam approximation only the incident and diffracted waves have appreciable amplitudes and the dynamical equations can be solved analytically for a uniform layer of specified composition, thickness and strain (Halliwell *et al.* 1984; Bartels *et al.* 1986). In writing the solution, we make use of the complex deviation parameter, $\eta$, defined as

$$\eta = \frac{\alpha_H + \chi_0(1-b)}{2|b|^{1/2}C(\chi_H\chi_{\bar{H}})^{1/2}}. \tag{5.1}$$

Here, $\alpha_H$ represents the angular deviation parameter, $b = \gamma_0/\gamma_H$ is the asymmetry factor, $\gamma_0$ and $\gamma_H$ are the direction cosines of the incident and diffracted waves,

respectively, $\chi_0$, $\chi_H$ and $\chi_{\bar{H}}$ represent the electric susceptibilities where $\bar{H} \equiv -H$, and $C$ is the polarization factor, which has a value of 1 and $|\cos\theta|$ for $\sigma$- and $\pi$-polarized waves with $\theta$ being the angle of incidence on the reflecting planes. The angular deviation parameter, $\alpha_H$, is a function of $\theta$ and is given by

$$\alpha_H = -4(\sin\theta - \sin\theta_{\mathrm{B}})\sin\theta_{\mathrm{B}} \approx -2\Delta\theta\sin 2\theta_{\mathrm{B}} \tag{5.2}$$

where $\Delta\theta = \theta - \theta_{\mathrm{B}} + \varphi$ is the angular deviation from the kinematic Bragg angle, $\theta_{\mathrm{B}}$, taking into account the angle $\varphi$ between the reflecting planes and the surface of the crystal. In order to calculate these quantities, the tetragonal distortion of the unit cell induced by the epitaxy, and modified by any relaxation, must be taken into account in each layer of the structure.

The electric susceptibilities of a crystalline layer have the following form:

$$\chi_H = -r_{\mathrm{e}}\frac{\lambda^2}{\pi V}F_H, \tag{5.3}$$

where $r_{\mathrm{e}}$ is the classical electron radius, $V$ is the volume of the unit cell and $F_H$ is the structure factor of the unit cell, given by the expression

$$F_H = \sum_a (f_{\mathrm{a}} + f'_{\mathrm{a}} + \mathrm{i}f''_{\mathrm{a}})\mathrm{e}^{-2\pi\mathrm{i}(hx_a + ky_a + lz_a)}. \tag{5.4}$$

Here $f_{\mathrm{a}}$ is the scattering factor away from any absorption edges and $f'_{\mathrm{a}}$, $\mathrm{i}f''_{\mathrm{a}}$ are the real and imaginary corrections of the scattering factor, respectively. The atomic positions of the atoms within the unit cell are denoted $(x_a, y_a, z_a)$, and $(hkl)$ represents the Miller indices of the reflecting planes. The summation is taken over all constituent atoms, $a$, within the unit cell.

The amplitude ratio $X_t = D_0/D_H$, where $D_0$ and $D_H$ denote the amplitudes of the incident and diffracted waves, at the top of a crystalline layer is related to the amplitude ratio at the bottom of the layer according to

$$X_t = \eta + (\eta^2 - 1)^{1/2}[(S_1 + S_2)/(S_1 - S_2)], \tag{5.5}$$

where

$$S_{1,2} = [X_0 - \eta \pm (\eta^2 - 1)^{1/2}]\exp[\mp\mathrm{i}T(\eta^2 - 1)^{1/2}] \tag{5.6}$$

and

$$T = \pi C(\chi_H\chi_{\bar{h}})^{1/2}t/(\lambda|\gamma_0\gamma_H|^{1/2}), \tag{5.7}$$

with $t$ being the thickness of the layer. The amplitude ratio at the top of the substrate is given by the Darwin–Prins formula, namely

$$X_\infty = \eta - \mathrm{sgn}[\mathrm{Re}(\eta)](\eta^2 - 1). \tag{5.8}$$

With this result in hand, the amplitude at the top of the first layer is calculated using equation (5.5). The calculation is then repeated recursively for all of the layers within the structure until the amplitude ratio at the top of the multilayer, $X$, is obtained. The plane-wave reflectivity for the multilayer is given by $R = |X|^2/b$ and for a randomly polarized incident wave is obtained by the arithmetic average of the $\sigma$- and $\pi$-reflectivities. The diffracted intensity is then calculated using equation (4.5), with the vector $\boldsymbol{p}$ containing the following adjustable parameters.

Table 4. *Nominal and best-fit parameter values for the pHEMT structure on* GaAs(001)

| | | nominal | | best-fit | |
|---|---|---|---|---|---|
| layer | material | $x$ (%) | $t$ (nm) | $x$ (%) | $t$ (nm) |
| 3 | GaAs | — | 35.0 | — | $38.9 \pm 0.3$ |
| 2 | $Al_xGa_{1-x}As$ | 24 | 21.5 | $35 \pm 8$ | $25.3 \pm 0.3$ |
| 1 | $In_xGa_{1-x}As$ | 22 | 9.8 | $21.4 \pm 0.1$ | $11.8 \pm 0.2$ |
| substrate | GaAs | — | $\infty$ | — | $\infty$ |

(1) The incident intensity $I_0$.

(2) The background intensity $I_b$.

(3) The composition parameters $x_j$ and $y_j$ of the layers $j = 1, 2, \ldots, N$. In the case of quaternary materials, for example $In_xGa_{1-x}As_yP_{1-y}$, a measurement of the material's bandgap by a complementary technique such as photoluminescence is required to determine the composition of the material.

(4) The thicknesses $t_j$ of the layers $j = 1, 2, \ldots, N$.

Finally, we note that one difficulty presented when fitting high-resolution X-ray diffraction rocking-curves is the dominance of the substrate peak; this may be four or more orders of magnitude greater than the next significant features to be fitted. It is therefore very important to avoid small errors in the fit around the substrate peak from dominating the error function. We achieve this in three ways: by using the $MAE_{log}$ function; by carefully matching the position and intensity of the substrate peak in all simulations; and by including an appropriate instrument function when evaluating equation (4.5).

### (*b*) *Example I.* $Al_xGa_{1-x}As$ *layer on* GaAs(001)

In our first X-ray diffraction example, we consider a layer of $Al_xGa_{1-x}As$ (500 nm) with nominal $x = 70\%$ grown on a GaAs(001) substrate. Figure 7*a* shows the measured and simulated rocking-curves before fitting. The angular scale $\Delta\theta = \theta - \theta_B$ used in this, and all remaining figures, denotes the offset between the incident angle and the (004) Bragg angle of the substrate ($\theta_B = 33.025°$ for GaAs and CuK$\alpha_1$ radiation). The peak to the left of the dominant substrate peak is due to the $Al_xGa_{1-x}As$ layer. Also visible are clear oscillations (Pendellösung fringes), the period of which is related to the thickness of the $Al_xGa_{1-x}As$ layer. The presence of these fringes indicates that the layer is pseudomorphic and has very good crystalline quality. Figure 7*b* shows the measured rocking-curve together with its best-fit simulation. Despite being started far from the expected structure of the sample, the agreement between the measured rocking-curve and its best-fit simulation is excellent (convergence time was less than 3 min). The best-fit parameters yield an $Al_xGa_{1-x}As$ layer of thickness $t = 487 \pm 5$ nm and composition $x = 72.7 \pm 0.2\%$, which is acceptably close to the nominal structure. In this very simple case, the composition can be checked by simple measurement of the angular separation between the layer and substrate peak, confirming the accuracy of the simulation.

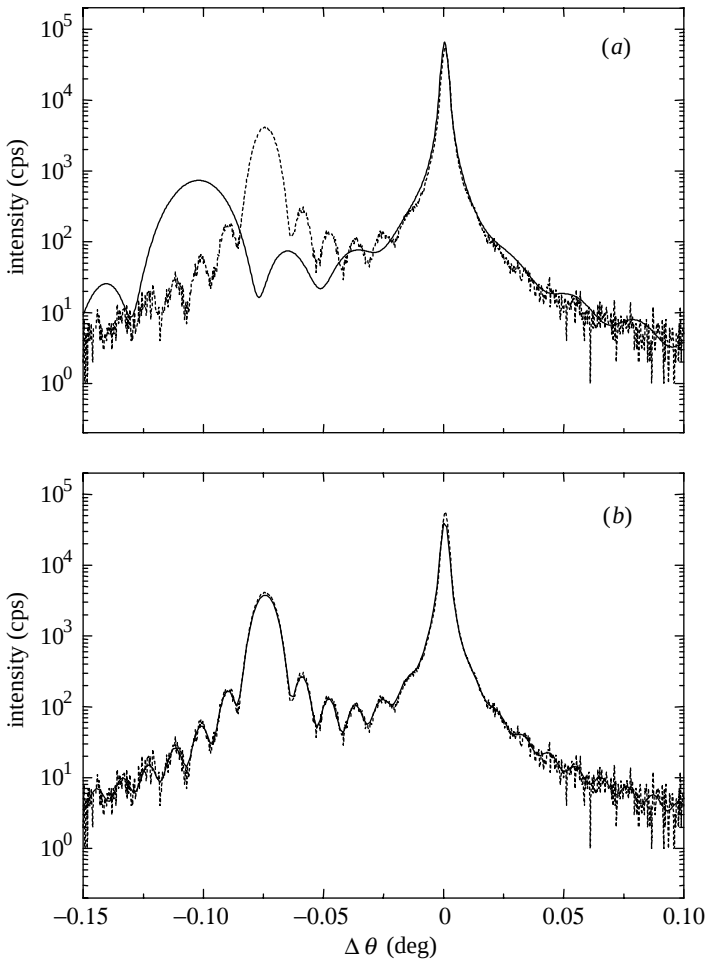


Figure 7. Comparison of experimental and simulated X-ray diffraction rocking-curves for an $Al_xGa_{1-x}As$ layer on a GaAs(001) substrate, (*a*) before and (*b*) after fitting. The dashed lines are the measurements and the solid lines represent the simulations.

### (*c*) *Example II. A pHEMT structure on* GaAs(001)

As a second example we have fitted the X-ray diffraction rocking-curve from a pseudomorphic high-electron-mobility transistor (pHEMT) structure grown on a GaAs(001) substrate. The structure is quite complicated (see table 4) but is typical of device structures that must be routinely characterized in the compound semiconductor industry. The measured rocking-curve, together with simulations before and after the fitting procedure has converged, is shown in figure 8*a*, *b*, respectively. The substrate peak is clearly visible together with a broad, far less intense, peak to the left ($\Delta\theta \sim -1.0^\circ$). The latter peak is due to the thin $In_xGa_{1-x}As$ layer in the structure and its angular position provides an accurate determination of the In concentration. Also visible are Pendellösung fringes: the short period fringes are related to the total thickness of the layers within the structure.

The fitting procedure has once again given a very good match between the mea-

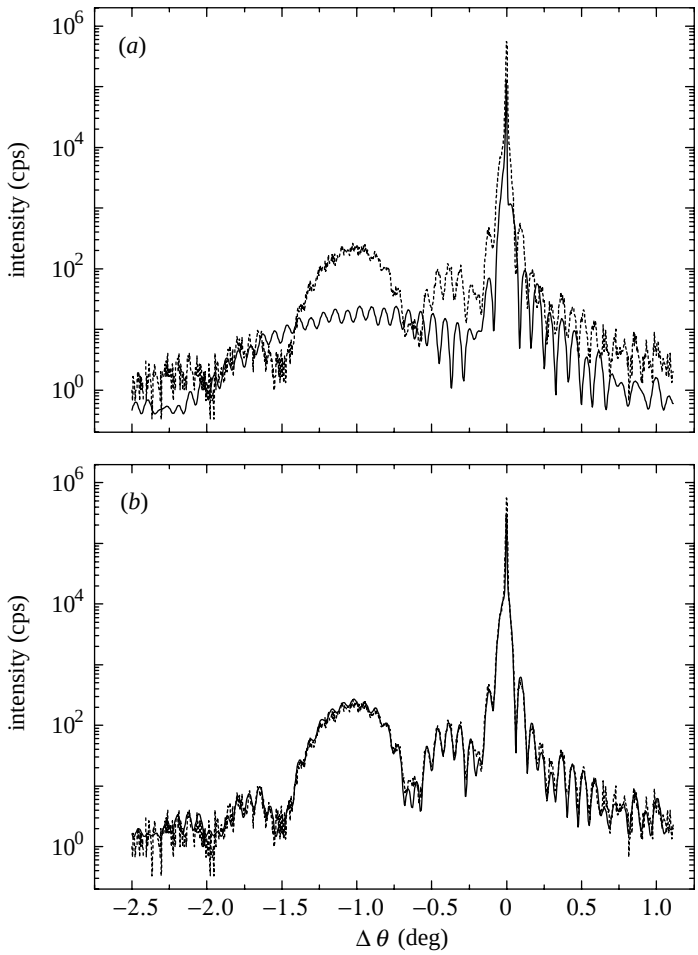


Figure 8. Comparison of experimental and simulated X-ray diffraction rocking-curves for a pHEMT structure on GaAs(001), (*a*) before and (*b*) after fitting. The dashed lines represent the measurements and the solid lines are the simulations.

sured rocking-curve and the best-fit simulation (the convergence time was *ca.* 15 min with a total of seven parameters fitted) despite the complicated structure of the sample. We note that while most of the parameter values are determined accurately there is a rather large uncertainty in the composition of the thin $Al_xGa_{1-x}As$ layer. This is not a deficiency in the fitting method, but a valuable indicator that the experimental measurement was relatively insensitive to the value of this parameter. In order to better determine the composition of the $Al_xGa_{1-x}As$ layer additional experiments and analysis are necessary.

### (*d*) *Example III.* An AlAs/GaAs *superlattice on* GaAs(001)

In the final high-resolution X-ray diffraction example, we have applied our fitting strategy to a superlattice with nominal structure $[AlAs(90\ nm)/GaAs(80\ nm)]_{15}$ on a GaAs(001) substrate. The measured rocking-curve, together with its best-fit simulation, is shown in figure 9. The data are visually quite striking and we

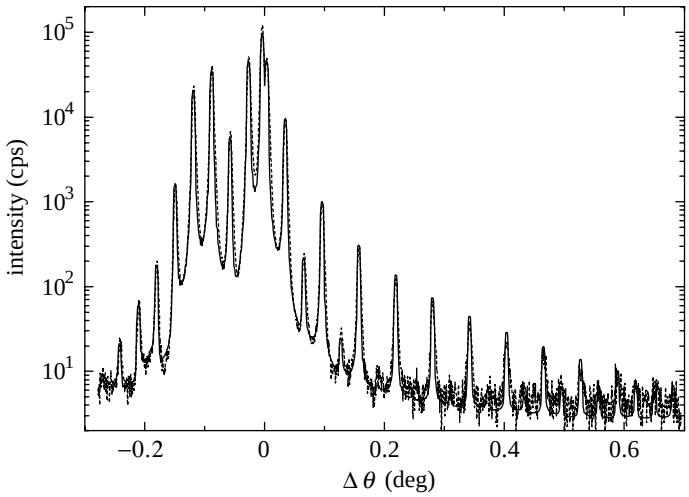


Figure 9. X-ray diffraction rocking-curves for an AlAs/GaAs superlattice on a GaAs(001) substrate. The dashed line represents the measurements and the solid line is the best-fit simulation. Sample courtesy of Dr D. Braddock (Ovation Semiconductor).

clearly observe satellite peaks associated with the superlattice on either side of the substrate peak. We find excellent agreement between the measured rocking-curve and the best-fit simulation both in terms of the positions and the intensities of the satellite peaks. The best-fit parameters yield a superlattice structure of $[\mathrm{AlAs}(88.7 \pm 0.1\ \mathrm{nm})/\mathrm{GaAs}(82.9 \pm 0.1\ \mathrm{nm})]_{15}$. Since the fitted structure is in close agreement with the nominal structure, we conclude that the growth of this sample was extremely well controlled.

## 6. Conclusions

We have demonstrated successful data-fitting and parameter optimization in previously intractable X-ray scattering problems, using a combination of the DE algorithm, a thoughtful consideration of the error function and a treatment of convergence and parameter errors. The procedure is robust against nonlinearity, local minima in the error function, data that span many orders of magnitude and the choice of initial parameter values. The method is conceptually simple, easy to implement and rapid in execution. A method is given for estimating errors in the inferred structure, which provides a way of specifying tolerances in manufactured components such as thin-film structures. We also provide a method of testing the validity of a particular structural model within the intrinsic limitations of the data available.

While we have focused on the general area of X-ray scattering, and in particular X-ray reflectivity and diffraction, the data-fitting method developed in this paper is general and could be applied to many inverse problems in the physical and engineering sciences. Some examples are ellipsometry, neutron reflectivity and X-ray fluorescence—indeed, any spectroscopic method for which there is no direct inverse transform. Certain imaging methods may also be tractable using DE, such as fringe analysis in optical or X-ray metrology. We believe that the DE algorithm has the potential to transform many data-fitting and optimization processes in both science and industry. In the former it can provide more accurate and rapid information;

in the latter, it can provide the ability to apply powerful and accurate scientific techniques that had previously been thought too inefficient.

We thank K. V. Price and R. Storn, the originators of the DE algorithm, for their valuable suggestions during the early stages of this work. We are also grateful to B. K. Tanner for his assistance in testing our new fitting strategy against conventional methods of fitting X-ray scattering data, and to the directors of Bede Scientific Instruments Ltd for giving permission to publish this paper. The use of EAs in software programs for the interpretation of X-ray scattering data is the subject of a pending US patent application.

## References

Bartels, W. J., Hornstra, J. & Lobeek, D. J. W. 1986 X-ray diffraction of multilayers and superlattices. *Acta. Crystallogr.* A **42**, 539–545.

Bevington, P. R. 1969 *Data reduction and error analysis for the physical sciences*. McGraw-Hill.

Bowen, D. K. & Wormington, M. 1993 Characterization of materials by grazing-incidence X-ray scattering. *Adv. X-ray Analysis* **36**, 171–184.

Bowen, D. K., Loxley, L., Tanner, B. K., Cooke, M. L. & Capano, M. A. 1991 Principles and performance of a PC-based program for simulation of double-axis X-ray rocking curves of thin epitaxial films. *Mater. Res. Soc. Symp. Proc.* **208**, 113–118.

Cook, S. 1971 The complexity of theorem-proving procedures. In *Proc. 3rd ACM Symp. Theory of Computing*, pp. 151–158.

Halliwell, M. A. G., Lyons, M. H. & Hill, M. J. 1984 The interpretation of X-ray rocking curves from III–V semiconductor device structures. *J. Cryst. Growth* **68**, 523–531.

Holland, J. H. 1975 *Adaptation in natural and artificial systems, an introductory analysis with application to biology, control, and artificial intelligence*. Ann Arbor, MI: University of Michigan Press.

Ibers, J. A. & Hamilton, W. C. (eds) 1974 *International tables for crystallography*, vol. IV. Birmingham: Kynoch.

Névot, L. & Croce, P. 1980 Caractérisation des surfaces par reflexion rasante de rayons X. Application à l'étude du polissage de quelques verres silicates. *Rev. Phys. Appl.* **15**, 761–779.

Parratt, L. G. 1954 Surface studies of solids by total reflection of X-rays. *Phys. Rev.* **95**, 359–369.

Powell, A. R., Bowen, D. K., Wormington, M., Kubiak, R. A., Parker, E. H. C., Hudson, J. M. & Augustus, P. D. 1992 X-ray diffraction and reflectivity characterisation of SiGe superlattice structures. *Semicond. Sci. Tech.* **7**, 627–631.

Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. 1989 *Numerical recipes in Pascal—the art of scientific computing*, ch. 10 and 14. Cambridge University Press.

Price, K. V. & Storn, R. 1997 Differential evolution. *Dr Dobb's J.* April, pp. 18–24.

Storn, R. & Price, K. V. 1995 Differential evolution—a simple and efficient scheme for global optimization over continuous spaces. Technical report TR-95-012, ICSI (URL: ftp://ftp.icsi.berkeley.edu/pub/techreports/1995/tr-95-012.ps.gz).

Wormington, M., Bowen, D. K. & Tanner, B. K. 1992 Principles and performance of a PC-based program for simulation of grazing incidence X-ray reflectivity profiles. *Mater. Res. Soc. Symp. Proc.* **238**, 119–124.