Journal of
Applied Crystallography

ISSN 0021-8898

Received 22 April 2008
Accepted 8 October 2008

Stochastic fitting of specular X-ray reflectivity data using *StochFit*

Stephen M. Danauskas,[a] Dongxu Li,[b] Mati Meron,[b] Binhua Lin[b] and Ka Yee C. Lee[a]*

[a]Department of Chemistry, Institute for Biophysical Dynamics and James Franck Institute, The University of Chicago, Illinois 60637, USA, and [b]CARS, The University of Chicago, Illinois 60637, USA. Correspondence e-mail: kayeelee@uchicago.edu

Specular X-ray reflectivity data provide detailed information on the electron density distribution at an interface. Typical modeling methods involve choosing a generic electron density distribution based on an initial speculation of the electron density profile from the physical parameters of the experimental system. This can lead to a biased set of solutions. *StochFit* provides stochastic model-independent and model-dependent methods for analyzing X-ray reflectivities of thin films at an interface. *StochFit* divides an electron density profile into many small boxes and stochastically varies the electron density of these boxes to locate the best fit to a measured reflectivity. Additionally, it provides the ability to perform model-dependent fitting with a stochastic search of the parameter space in order to locate the best possible fit. While model-independent profile search algorithms have been described previously, they are difficult to implement because of the heavy computational requirements, and there is a dearth of software available to the general scientific public utilizing these techniques. Several cases that illustrate the usefulness of these techniques are presented, with a demonstration of how they can be used in tandem.

1. Introduction

Specular X-ray reflectivity (XR) can be used to determine the electron density distribution normal to the surface of thin films at interfaces and has broad applications in the fields of biology (Ege *et al.*, 2005; Miller *et al.*, 2004, 2006), materials science (Knapp *et al.*, 1996) and nanotechnology (Lin *et al.*, 2007; Schultz *et al.*, 2006; Sanyal *et al.*, 2008). In XR, the reflectivity from an interface is measured as a function of the angle, θ, of the incident X-ray beam, which is related to the z component of the wavevector transfer ($q_z = 2k\sin\theta$, where k is the wavevector). Internal reflection and refraction between layers in the film alter the intensity of the incident beam, and the measured reflectivity is given by the sum of these effects (see Fig. 1). Unfortunately, the phase information is lost during this measurement, and direct inversion of the reflectivity becomes impossible under most circumstances (Majkrzak & Berk, 1996). This loss of phase information leads to non-uniqueness in the solutions to the reflectivity, and introduces the need for modeling the measured reflectivity to determine the electron density profile (EDP) of the film (Politsch & Cevc, 2002).

Model-independent (MI) fitting for XR has been very successful in providing reasonable fits to layered structures (Chou *et al.*, 1997), Langmuir monolayers (Miller *et al.*, 2004; Pedersen, 1992), buried interfaces (Laub & Kuhl, 2006; Miller *et al.*, 2006) and even EDPs that do not adhere to a model-dependent (MD) formalism (Schultz *et al.*, 2006). MI algorithms have been implemented with widely varied techniques (Fenter & Zhang, 2005; Sanyal *et al.*, 1998; Bengu *et al.*, 2001; Choi *et al.*, 1998; Chou *et al.*, 1997; Dane *et al.*, 1998; Laub & Kuhl, 2006; Pedersen, 1992). Among the most common techniques that have been used are those that divide an EDP with no features into a small number of boxes. These boxes are then varied in either thickness or electron density (ED), and a smoothing procedure is often applied to generate a more physically relevant solution (Chou *et al.*, 1997). Reasonable boundary conditions gleaned from obvious physical constraints such as thickness, average film ED, subphase ED and superphase ED are also often applied to give the algorithm a reasonable starting point (van der Lee *et al.*, 2007). Additionally, model EDPs have been generated using spline functions (Pedersen, 1992; Politsch & Cevc, 2002) and Chebyshev polynomials (Laub & Kuhl, 2006). The dynamical Parratt (1954) recursion is then applied to generate a reflectivity from the EDP. This, unfortunately, leads to large computational requirements which have somewhat limited the effectiveness and widespread adoption of these techniques.

StochFit provides an easy to use graphical user interface (GUI) for fitting XR data. The program contains modules for both MI and MD fitting. In addition, it allows for a unique blending of the two. In this paper, we show the utility of *StochFit* in fitting XR data that cannot be modeled *via* standard MD techniques, and the accuracy with which synthetic data can be fitted. By employing a highly optimized reflectivity calculation, and splitting the calculation across multiple processors, accurate fits to data can be achieved in just several minutes. *StochFit* can also be used to search the MD parameter space stochastically for the global minimum and to organize all the models that have been identified for a given number of boxes by χ^2 value.

2. Methods

In *StochFit*, both MI and MD fitting routines depend on the calculation of the reflectivity based on a recursive method, first proposed by Parratt (1954). The electron density profile is divided into very small layers (< 1 Å), and the reflectivity at each layer is calculated

taking into account the layers beneath it, starting from the subphase and ending with the superphase. The final reflectivity at the interface is thus based on the contributions of each layer beneath it.

2.1. Reflectivity calculation

The refractive index, n, for X-ray radiation differs only slightly from unity and is given by equation (1):

$$n = 1 - \delta - i\beta, \tag{1}$$

where δ and β are given by

$$\delta = \lambda^2 \mathrm{SLD}/(2\pi) \quad \text{and} \quad \beta = \lambda\mu/(4\pi) \quad \text{or} \quad \beta = \lambda^2\varphi/(2\pi). \tag{2}$$

Here, λ is the wavelength of the X-ray beam and μ is the linear absorption coefficient. SLD is the scattering length density of the layer and φ is an absorption term, which are given by

$$\begin{aligned} \mathrm{SLD} &= \frac{r_\mathrm{e} N_\mathrm{A} \rho}{\mathrm{MW}} \sum_{i=1}^{N} f_{1i} \simeq \sum_{i=1}^{N} \frac{Z_i r_\mathrm{e}}{V_\mathrm{m}}, \\ \varphi &= \frac{r_\mathrm{e} N_\mathrm{A} \rho}{\mathrm{MW}} \sum_{i=1}^{N} f_{2i}, \end{aligned} \tag{3}$$

where, for N atoms, Z_i is the number of electrons of each atom i, r_e is the classical radius of the electron, V_m is the molecular volume, N_A is Avogadro's number, ρ is the bulk density, MW is the molecular weight, and f_{1i} and f_{2i} are the real and imaginary anomalous scattering factors for the ith atom, respectively. The approximation to SLD given above is commonly used and accurate in most cases. The previously mentioned EDP is defined as the value of δ as a function of the depth from the interface and is often reported in normalized ED units, where δ of the layer of interest is divided by δ of the subphase. The two different formulations of the absorption term, β, in (2) are used interchangeably in the literature (Parratt, 1954; Stoev & Sakurai, 1997). While the first equation for β is the original form proposed by Parratt, many computer programs favor the second formulation [*Motofit* (Nelson, 2006), *Parratt32* (Braun, 1997) *etc.*], and it is not always clear how this is implemented (van der Lee *et al.*, 2007). In this paper, we will be adhering to the second formulation, which does not take Compton scattering into account. For many systems, the interface is treated as a transparent medium, and the β term can be neglected without impacting the solution.

The measured reflectivity is then calculated by the dynamic method of Parratt, which is based on optical theory. The following brief summary of the derivation represents an exact solution to the wave equation for an s-polarized plane wave at an ideal interface (Stoev & Sakurai, 1997). For q values greater than the critical wavevector for total internal reflection,

$$q_\mathrm{c} = 4\left[\Pi(\mathrm{SLD}_\mathrm{sub} - \mathrm{SLD}_\mathrm{sup})\right]^{1/2}, \tag{4}$$

the reflectivity falls drastically. In (4), the sup and sub subscripts correspond to superphase and subphase, respectively. The normal component of the wavevector at the interface for a given q and a superphase X-ray refractive index of n_0 has a value of

$$k_0 = (2\pi/\lambda)\, n_0 \sin\theta_0, \quad \text{where} \quad \sin\theta_0 = q\lambda/(4\pi n_0), \tag{5}$$

and the normal component of the wavevector for layer i is given by

$$k_i = (2\pi/\lambda)\left[n_0^2 \sin^2\theta_0 - 2(\delta_i + i\beta_i - \delta_0 - i\beta_0)\right]^{1/2}, \tag{6}$$

where n_0, δ_0 and β_0 correspond to the respective parameters of the superphase from equation (1), and λ is the wavelength of the incoming X-ray beam. The Fresnel coefficient, f_i, for each layer is then

$$f_i = (k_i - k_{i+1})/(k_i + k_{i+1}) \tag{7}$$

and the phase factor for each layer is given by

$$a_i = \exp(-2ik_i t_i), \tag{8}$$

where the thickness of the layer, t_i, is taken into account. Assuming the boundary conditions of the subphase layer, R_nl, having no reflected wave ($R_\mathrm{nl} = 0$) and the first (superphase) and last (subphase) layers being infinite in thickness ($a_0 = a_\mathrm{nl} = 1$), the amplitude reflectivity for the ith layer is then

$$R_i = a_i(R_{i+1} + f_i)/(R_{i+1} f_i + 1). \tag{9}$$

After recursion, the total measurable reflectivity is given by the complex modulus of the amplitude reflectivity at the interface of the first layer and the superphase:

$$\mathbf{R} = |R_0^2|. \tag{10}$$

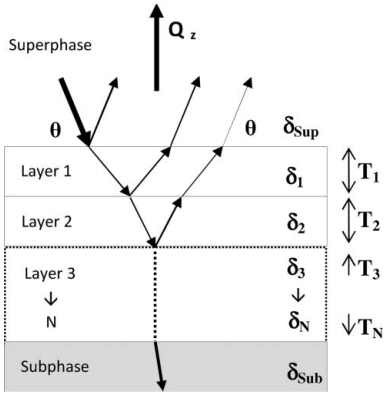


Figure 1
X-ray reflectivity geometry for a thin film at an interface. The film at the interface is divided into N layers to represent the difference in average electron density between portions of the film. The subphase layer and the superphase layer are considered infinite with respect to the thickness of the film. The arrows denote the direction of propagation, at incident angle θ, of the X-ray beam with secondary refraction and reflection ignored, T_i is the thickness of layer i, and δ_i is the dispersive component of the X-ray refractive index of layer i as defined in (2).

2.2. Model-independent fitting

For model-independent fitting, we perform an EDP search. An initial EDP is generated on the basis of the average SLD of the sample and the estimated sample thickness. Sample thickness can be estimated from physical measurements (AFM, ellipsometry *etc.*) or by taking the Fourier transform of the reflectivity (Kago *et al.*, 1998; Sakurai & Atsuo, 1992). The EDP is generated by selecting a number of boxes with a fixed thickness, typically of the order of ~0.5 Å per box, a smoothing parameter (σ) and a δ for each box (Fig. 1). The contribution of each box to the EDP is then smeared by the Gaussian error function (erf) given by

$$\delta(z) = \delta_0 + \sum_{k=0}^{\mathrm{NB}} \left(\frac{\delta_{k+1} - \delta_k}{2}\right)\left[1 + \mathrm{erf}\left(\frac{z - kB_\mathrm{t}}{2^{1/2}\sigma}\right)\right]. \tag{11}$$

Here, NB is the total number of boxes, and δ_k from $k = 1$ to NB is the δ value of each box k; for $k = 0$ and $k = \mathrm{NB} + 1$, δ_k is the index for the δ value of the superphase and subphase, respectively. B_t is the box thickness in ångströms, σ is the smoothing parameter in ångströms and z is a spatial parameter oriented perpendicular to the subphase.

Each point in the EDP is treated as a layer, and the reflectivity is then calculated by iterating through each of the points of the EDP by the method described in §2.1. Using equation (11) leads to the EDP being defined as a function of increasing z, the opposite of the normal convention. In standard MD methods (*MotoFit*, *Parratt32*) $z = 0$ corresponds to a film/interface boundary, and the EDP is given as a function of decreasing z. However, this orientation is misleading in the MI algorithm, as we are unable to determine which point in the EDP corresponds to a traditional MD interface. It should be noted that the Parratt recursion is insensitive to the arbitrary orientation of z.

For absorbing films, the algorithm is not as straightforward. Since the absorption is typically at least an order of magnitude less than the reflectivity, a separate set of boxes to model the absorption can lead to a fit where the absorption is overemphasized with respect to the ED. Our solution to this problem is to link the absorption to a normalized EDP as shown below:

$$\begin{aligned}\beta(z) = \beta_{\rm sup} &+ \left(\frac{\beta_{\rm a}\delta_1/\delta_{\rm NB+1} - \beta_{\rm sup}}{2}\right)\left[1 + {\rm erf}\left(\frac{z}{2^{1/2}\sigma}\right)\right] \\ &+ \sum_{k=1}^{\rm NB-1} \frac{\beta_{\rm a}}{\delta_{\rm NB+1}}\left(\frac{\delta_{k+1} - \delta_k}{2}\right)\left[1 + {\rm erf}\left(\frac{z - kB_{\rm t}}{2^{1/2}\sigma}\right)\right] \\ &+ \left(\frac{\beta_{\rm sub} - \beta_{\rm a}\delta_{\rm NB}/\delta_{\rm NB+1}}{2}\right)\left[1 + {\rm erf}\left(\frac{z - {\rm NB}\,B_{\rm t}}{2^{1/2}\sigma}\right)\right], \quad (12)\end{aligned}$$

where $\beta_{\rm sup}$ and $\beta_{\rm sub}$ are the absorbances of the superphase and subphase, respectively, and $\beta_{\rm a}$ is a parameter that is allowed to vary. The remaining parameters correspond to those in equation (11). Here, we are linking the absorption profile to the EDP. The third term in equation (12) corresponds to the normalized EDP boxes multiplied by the varying absorption parameter. The second and fourth terms are necessary to ensure a smooth absorption profile as the film transitions to the superphase and subphase, respectively. Physically, the calculated value of $\beta_{\rm a}$ multiplied by the normalized EDP can be considered an overall absorption value for the film. Equation (12) allows the absorption to vary with changes in the ED, and maintains the absorption of the superphase and the subphase. To a first approximation, linking the absorption to the EDP is accurate for single-component systems, where the absorption does not significantly vary throughout the film. It will also be applicable to systems where the absorbing component is large and highly reflective with respect to other components. However, before using this technique, one should verify that the experimental system meets the above criteria.

During each iteration of the algorithm, either the δ of one box or the smoothing parameter σ is allowed to vary. For absorbing films, the parameter $\beta_{\rm a}$ is also allowed to vary. In this implementation, the thicknesses of the boxes are fixed. The EDP is allowed to vary up to 0.03 normalized ED units, the smoothing parameter is allowed to vary by up to 5% and $\beta_{\rm a}$ is allowed to vary by up to 3%. There are two algorithms available for minimizing XR. The default method of the program performs a 'greedy search'. If the resulting calculated reflectivity improves the fitness function, equations (13) or (14), the answer is accepted.

$$F = \sum_{i=1}^{N}\left[\ln(R_{{\rm calc}_i}) - (R_{{\rm exp}_i})\right]^2, \quad (13)$$

$$F = \sum_{i=1}^{N} x_i, \quad \text{where } x_i = \begin{cases} (1 - R_{{\rm calc}_i}/R_{{\rm exp}_i})^2 & \text{if } R_{{\rm calc}_i}/R_{{\rm exp}_i} \geq 1 \\ (1 - R_{{\rm exp}_i}/R_{{\rm calc}_i})^2 & \text{if } R_{{\rm calc}_i}/R_{{\rm exp}_i} < 1. \end{cases} \quad (14)$$

Here, $R_{\rm calc}$ is the calculated reflectivity and $R_{\rm exp}$ is the measured reflectivity.

Simulated annealing (SA) is available as an additional option in order to attempt to locate the global minimum of the fitness function. In this case, the current state of the EDP is accepted with the Boltzmann criteria over the previous state (Kirkpatrick *et al.*, 1983; Metropolis *et al.*, 1953). The SA algorithm will depend strongly on the cooling schedule and the initially selected temperature, and it is therefore more difficult to come up with a general set of rules (Černý, 1985). As both methods are stochastic, there is not a particular criterion for ending the search. However, MI fits using the 'greedy' algorithm typically do not yield significant improvement after three million iterations. An example of the GUI during the MI fitting process is shown in Fig. 2(*a*).

StochFit provides the option of choosing from several fitting functions, but in the absence of truly poor statistical errors, the fitness functions given by equations (13) and (14) give the best results. Both equations emphasize the high-q regime initially, but once all points are within the same order of magnitude as the measured reflectivity, all of the points are treated equally. Equation (14) in particular strongly penalizes calculated reflectivities that deviate significantly from the measured data. Empirically, we have observed that the χ^2 function

$$\chi^2 = \sum_{i=1}^{N}(R_{{\rm calc}_i} - R_{{\rm exp}_i})^2/\gamma_i^2, \quad (15)$$

(where γ^2 is the variance) employed by many gradient descent methods yields poor fits at high q for the MI algorithm. It should be noted that equations (13) and (14) do not take into account the statistical errors associated with $R_{\rm exp}$. Heavily weighting the fitness function with statistical errors tends to emphasize the low-q portion of the curve at the expense of the high-q portion, where much of the information is encoded. As long as most errors are relatively small with respect to the measured reflectivity, error correction terms are generally not necessary. However, several objective functions that take the statistical errors into account are included.

2.3. Extracting a model-dependent fit from a model-independent fit

In order to extract physically meaningful parameters out of the MI fit, *StochFit* provides the means to apply an MD fit to the MI EDP, providing a bridge between methodologies. The MI fit is plotted against a generic MD fit, and the user provides a reasonable starting point for the model. In many cases it is evident from the MI EDP what constitutes a reasonable starting model. The base MD EDP is then fitted using a Levenberg–Marquardt nonlinear least-squares fit (LMLS; Marquardt, 1963). The EDP for the MD fit is minimized to the χ^2 fit function without the variance term. The interface for this module is shown in Fig. 2(*b*).

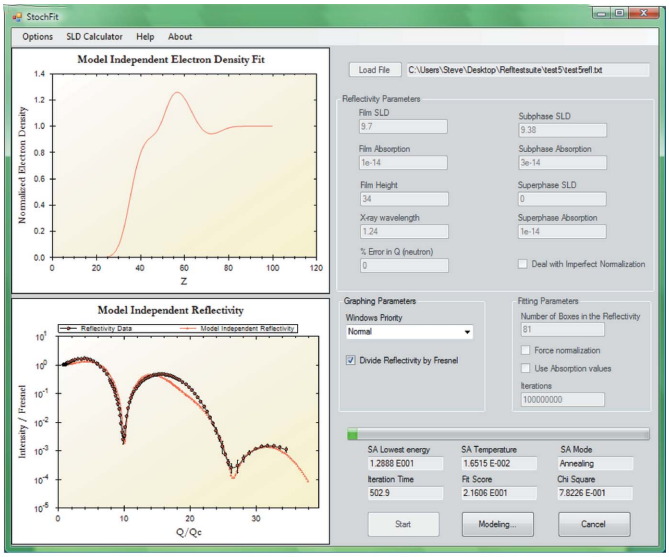
2.4. Stochastic box model fitting

In contrast to the transition layer model, the typical MD box model algorithm does not require generation of an EDP for the reflectivity calculation. The reflectivity is generated by providing an initial guess as to the numbers of layers of the sample that have a widely varying ED. Each of these portions is then considered a box. On the basis of the molecular structure and physical parameters, the structure is divided into a fixed number of boxes each with a δ, thickness and roughness at the interface of the boxes. The reflectivity is calculated by using the Nevot–Croce correction term to the Fresnel coefficients (Croce & Nevot, 1976; Hamley & Pedersen, 1994; Stoev & Sakurai, 1997):

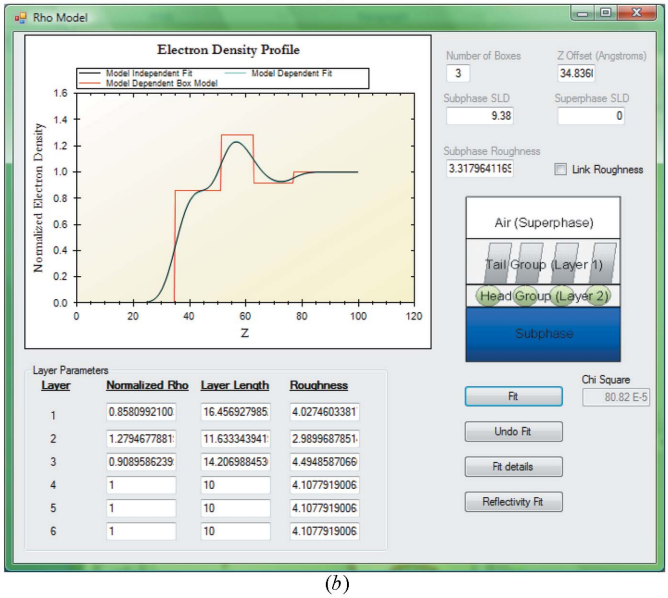
$$f_i = Q(k_i - k_{i+1})/(k_i + k_{i+1}), \quad \text{where } Q = \exp(-2\sigma_{i+1}^2 k_i k_{i+1}). \tag{16}$$

Utilizing this correction term, the number of layers in the calculation is reduced to the number of boxes plus one. For this calculation, the number of times one must calculate equations (6), (7), (8) and (9) decreases significantly and the calculation of the reflectivity is orders of magnitude faster. There are, however, certain caveats. The correction term begins to break down as the roughness for a layer approaches the layer thickness. In this case, the reflectivity must be calculated by the transition layer model of §2.2. Additionally, the model must conform to a Gaussian distribution of ED between layers.

The box model methodology has two major weaknesses. The first is that the number of boxes is fixed. For films of unknown composition, such as protein insertion (Ege *et al.*, 2005; van der Lee *et al.*, 2007), this can be a particularly difficult parameter to guess. The default strategy is to choose the smallest number of boxes that provides a good fit, but this does not necessarily have to be the correct EDP. Secondly, convergence of the model to a global minimum strongly depends on the starting parameters, which are unknown. The stochastic box model fitting varies each parameter for each box between user defined values. This allows one to form a collection of many models (up to 1000) for a given number of boxes within a fairly large range of values. A screenshot of this collection can be seen in Fig. 3.



(*a*)



(*b*)

Figure 2
(*a*) Graphical user interface for the model-independent fitting routine and (*b*) graphical user interface for the box model fitting to the model-independent fit. A model-independent fit in progress is shown in (*a*), and the model-dependent fit to the electron density profile obtained in (*a*) is shown in (*b*).

3. Data analysis

3.1. Gold nanoparticles at an air/water interface

For systems that do not adhere to the 'box model' criteria (*e.g.* when roughness is large with respect to layer thickness, or the EDP has a non-Gaussian distribution over large layer thicknesses) the options for fitting XR data are drastically limited. Here, we examine the reflectivity of a thiolated gold Langmuir monolayer at the air/water interface. Further information regarding the sample preparation and instrumental setup can be found in detail elsewhere (Lin *et al.*, 2003, 2007; Schultz *et al.*, 2006). The only parameters needed were subphase SLD (9.38×10^{-6} Å^{-2}), subphase absorption (3×10^{-8} Å^{-1}), layer thickness (60 Å) and an initial guess for the SLD of the film (56×10^{-6} Å^{-2}). The thickness was determined using the Fourier transform of the data from the *Motofit* analysis package. Equation (14) was used as the fitness function. The number of small boxes to vary was set at 120. The search for the smoothing parameter σ was limited to 15% of the iterations, the search for absorption was limited to 15% of the iterations and normalization of the generated reflectivity was forced. Empirically, we have observed that it is necessary to reserve a relatively high percentage of the total iterations for a search of the σ space. In physical terms, σ determines the degree to which the reflectivity falls off with increasing q, the ED distribution at the air/film interface and the degree to which features are blended within the EDP.

The fit to the data is shown in Fig. 4. The MI fit is virtually indistinguishable from the experimental data. The χ^2 value of 1.12 was achieved over approximately 25×10^6 iterations. The high number of iterations was necessary to fully resolve the high-q region. However, the basic shape of the modeled EDP was set after approximately 3×10^6 iterations. The EDP shows a very reasonable shape based on the

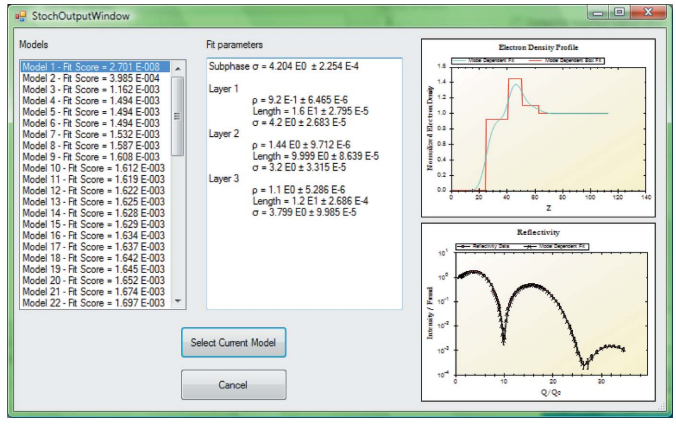


Figure 3
Graphical user interface for the stochastic box model module. The text box on the left contains all models found based on selected criteria, while the graphs on the right are updated on the basis of the model selected.

Table 1
Parameters from generated reflectivities, and parameters for the model-dependent fit to the model-independent fit.

All ρ values are reported as normalized electron density; thicknesses (T) and roughnesses are reported in ångströms.

Model	Parameter	Value	MI Fit	% error
Model 1				
$\chi^2 = 1.27 \times 10^{-4}$	$\rho 1$	0.95	0.947	0.32
	$\rho 2$	1.44	1.436	0.28
	$T1$	16	15.98	0.13
	$T2$	10	10.03	0.30
	σ	3.15	3.135	0.48
Model 2				
$\chi^2 = 6.48 \times 10^{-115}$	$\rho 1$	0.95	0.947	0.32
	$\rho 2$	1.44	1.437	0.21
	$T1$	16	15.99	0.06
	$T2$	10	10.03	0.30
	σ_{sub}	4.2	4.175	0.60
	σ_1	3.2	3.185	0.47
	σ_2	3.8	3.81	0.26
Model 3				
$\chi^2 = 9.48 \times 10^{-17}$	$\rho 1$	0.95	0.947	0.32
	$\rho 2$	1.44	1.437	0.21
	$\rho 3$	1.2	1.200	0
	$T1$	16	15.95	0.31
	$T2$	10	10.09	0.9
	$T3$	12	11.96	0.33
	σ	3.15	3.136	0.44
Model 4				
$\chi^2 = 3.69 \times 10^{-4}$	$\rho 1$	0.92	0.878	4.57
	$\rho 2$	1.44	1.476	2.50
	$\rho 3$	1.1	1.095	0.45
	$T1$	16	16.02	0.13
	$T2$	10	9.65	3.50
	$T3$	12	12.64	5.33
	σ_{sub}	4.2	4.19	0.23
	σ_1	4.2	3.87	7.86
	σ_2	3.2	3.65	14.1
	σ_3	3.8	4.48	17.9

known profile of the gold nanoparticles being roughly spherical (Lin *et al.*, 2007). The overall roughness value, σ, was 5.74 Å and the β term in the X-ray refractive index was calculated to be 2.1×10^{-7}. After an exhaustive search, we concluded that there was no 'box' model that would provide a fit to the data.

3.2. Fitting arbitrary profiles of Langmuir monolayers at the air/water interface

In order to truly judge the efficacy of a fitting routine, it must be tested against known parameters. To that end, we have generated several test reflectivities based on a box model Gaussian distribution of electron density. A sampling of the generated reflectivities is shown in Table 1. The models vary between two to three boxes, and between a single roughness for all interfaces to multiple roughnesses for each interface. These represent reasonable profiles of phospholipids or other thin films at the air/water interface (Ege *et al.*, 2005). For easy comparison, each model was fitted with the same set of parameters, with the exception of thickness, and each model was treated as fully transparent [β of equation (1) is ignored]. All models were fitted with 80 boxes. The search for the smoothing parameter σ was limited to 20% of the iterations.

Table 1 illustrates that *StochFit* is highly effective at reproducing the original reflectivity. In most cases, the original EDP is indistinguishable from the MI EDP [Figs. 5(*a*) and 5(*b*) illustrate this]. The third column shows the values used to generate the EDP and the fourth column is the MD LMLS fit to the MI data. All electron densities are reported in units of normalized electron density; thicknesses and roughnesses (σ) are in ångströms. The errors for the parameters of the first three models are well below 1%, and all χ^2 scores are excellent. The fourth model is a three-box model with multiple roughnesses which becomes trapped in a very deep local minimum. While the percent errors for the parameters are relatively high, they are highest in the roughness parameters. Fortunately, the roughness parameters have the least detrimental effect on the shape of the curve as long as they are still within reasonable values. Figs. 5(*c*) and 5(*d*) show that the general shape of the original EDP is recovered, if not the exact values. Simulated annealing, despite its reputation for overcoming local minima, was also unable to recover the initial value.

The fourth model proved highly resistant to finding the global minimum. The local minimum was eventually found after an exhaustive search of the parameters space by the methodology of §2.4. LMLS minimization methods are very sensitive to the starting point, hence the need for a very good guess for the original parameter set when modeling the data. We performed an MD LMLS fit to the MI fit to determine a reasonable starting parameter set. We then ran

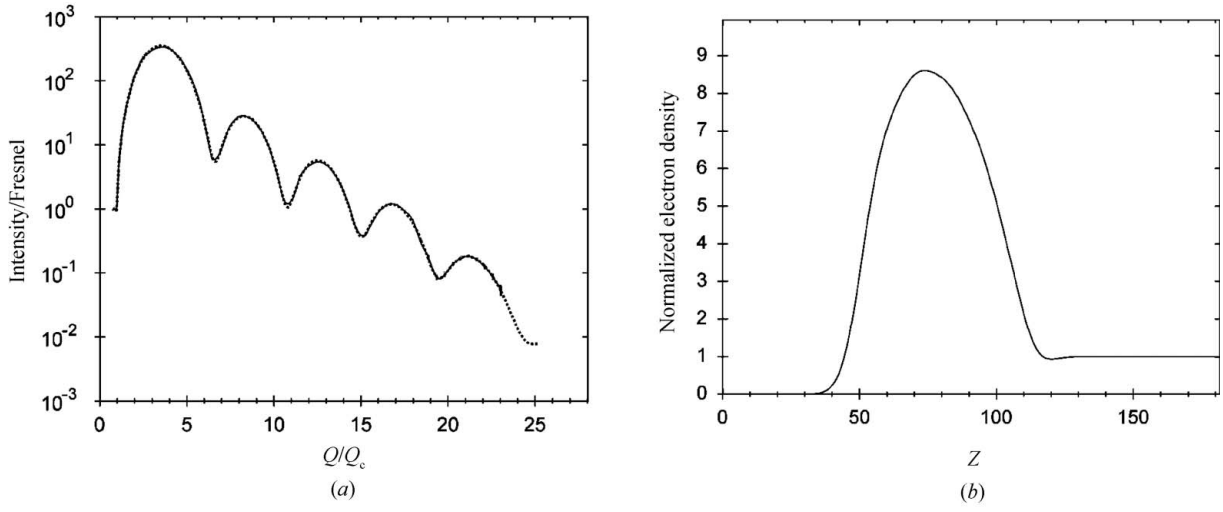


Figure 4
Model-independent fit of a gold monolayer at the air/water interface. (*a*) The solid line is the measured reflectivity, and the dotted line is the model-independent fit. The reflectivity is divided by the Fresnel reflectivity to enhance features. (*b*) The normalized electron density profile of the model-independent fit.

100 000 individual MD LMLS runs with the values of the parameters being varied by as much as 100% of the original presupposed values. This finds the global minimum, in addition to hundreds of other local minima. A maximum of 1000 unique minima with the lowest χ^2 values are reported. From this algorithm, we can see that, while the global minimum is difficult to find, almost all of the local minima with reasonable parameters share the same general shape. The solution found by the MI algorithm is actually the most common solution found by the LMLS fitting procedure as well, with the only deviation occurring at very high q. The provision of parameters very close to those of the original (generated) EDP to the LMLS algorithm for model 4 does result in the original curve being obtained. However, if the parameters differ from the original EDP by more than 10%, the resulting curve deviates to the deep local minimum found by the MI algorithm. This level of error would be common when providing the best guess of parameters to an MD algorithm. The stochastic search of the parameter space for model 4 thus demonstrates that this algorithm allows for one to locate a nearly inaccessible global minimum.

4. Benchmarks and reflectivity comparison

The length of time necessary to perform a fit depends on several parameters. The number of small boxes, the number of points per ångström in the EDP, resolution smearing and the total thickness of the EDP all impact on the speed of the algorithm. For a film with a total thickness of 111 Å, 80 small boxes and 3 points Å^{-1} (model 1 from §3.2) on an Intel Core2 Quad 6600 CPU overclocked to 3.0 GHz, *StochFit* performs 1140 iterations per second. An excellent fit to the data is seen after only ∼6000 iterations (5 s), and no further improvement in the fit is seen after $\sim 3 \times 10^6$ iterations (44 min). In general, doubling the total number of points (points per ångström × thickness) halves the performance of the algorithm. Resolution smearing, where the reflectivity is convoluted with a function describing the angular divergence of the beam (Stoev & Sakurai, 1997; Tolan, 1999), also significantly slows down the algorithm. Allowing for resolution smearing decreases the speed of the algorithm by a factor of ∼13, and a trade-off between computational time and the needed accuracy of the reflectivity must be reached. For samples that are illuminated by highly coherent X-rays, resolution smearing is generally unnecessary.

To verify the correctness of the reflectivities generated by *StochFit*, we generated MD models based on Table 1 model 1 using *StochFit*, *MotoFit* and *Parratt32*. The reflectivities for this model were very similar for all three programs. The maximum percent error between *MotoFit*'s and *StochFit*'s MD routines was 0.025%, with an average percent error of 0.002%. For comparison, the maximum percent error between *MotoFit*'s output and *Parratt32*'s output was 0.04%, with an

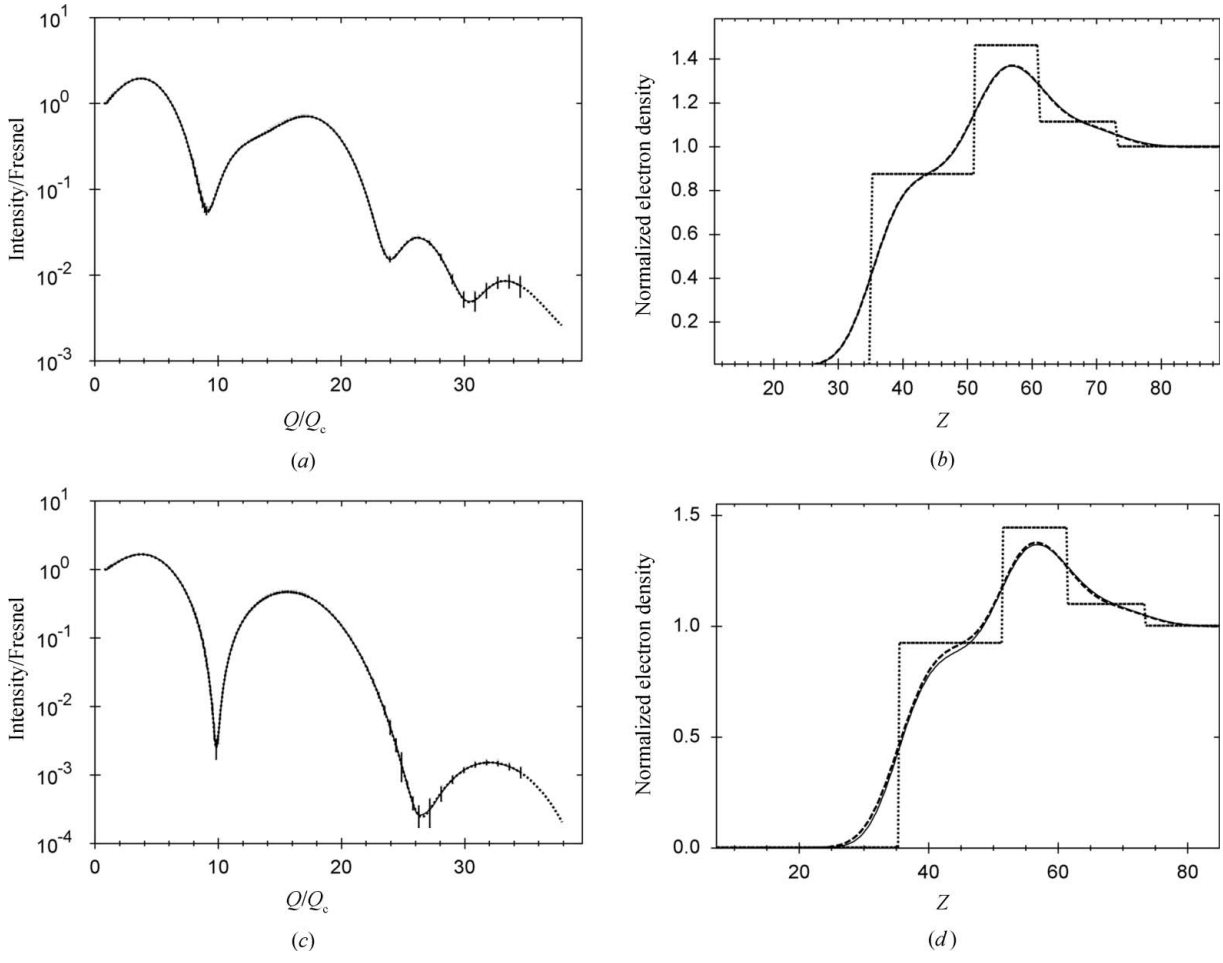


Figure 5
(*a*) XR and (*b*) EDP of model 3 in Table 1. (*a*) The solid line is the generated XR data and the dashed line is the MI fit. (*b*) The solid line is the MI-generated EDP and the dashed line is the MD fit to the MI data. Here, the overlap is nearly complete, and one cannot distinguish between the fits. (*c*) XR and (*d*) EDP of model 4 in Table 1. (*c*) The solid line is the generated data and the dashed line is the MI fit. (*d*) The solid line is the original EDP used to generate the reflectivity. The dotted line in both (*b*) and (*d*) is the box model fit associated with the corresponding MD profile.

average percent error of 0.008%. *StochFit*'s MI algorithm gives a larger error with respect to *MotoFit*'s MD routines; however, this is not unexpected as the algorithms differ and the MD algorithm is an approximation. The maximum percent error between *MotoFit*'s MD routine and *StochFit*'s MI routine was 0.801%, with an average error of 0.108%. In general, the deviation between the MD and MI routines is seen to be small. As illustrated in §3.2, the effect on the fit is negligible.

5. Features, download and redistribution details

StochFit is distributed as a self-contained installer for Microsoft Windows (2000, XP and Vista) under the GNU General Public License version 2. This ensures that the code is available for verification and modification. The source code and the executable are hosted at http://stochfit.sourceforge.net/. The GUI is developed in C# and the core numerical routines are written in C++ as dynamic link libraries. It is threaded using OpenMP (http://www.openmp.org/) in order to split the calculation across multiple processors. The only prerequisites are the .NET Frameworks 1 and 2, which are downloadable from Microsoft. The following open source components are also integral to the scientific operation of the program:

(1) levmar (http://www.ics.forth.gr/~lourakis/levmar): Levenberg–Marquardt nonlinear regression

(2) ZedGraph (http://sourceforge.net/projects/zedgraph): graphing library

StochFit has several features that have not been directly touched upon in the previous sections.

(1) Report output of all parameters and graphs in the Adobe PDF document format

(2) Ability to alter the number of points in the ED and the total thickness of the ED for the MI module to increase calculation speed

(3) Export of all graphs as vector graphics for import into documents, as well as bitmaps and jpegs; all graphs in this paper were generated by *StochFit*

(4) The superphase SLD can be non-zero to account for buried interfaces

(5) Gaussian smearing of the resolution based on an approximated instrumental error in q or actual error in q

(6) Correction for imperfect normalization

(7) Constraints for model-dependent fits

(8) Correction for imperfect normalization of highly absorbing films

(9) X-ray and neutron SLD calculator

(10) Text file output for all data fits

6. Conclusions

StochFit implements a highly effective algorithm for analyzing X-ray reflectivity data using a model-independent algorithm. For systems that adhere to 'box-model' criteria, a model-dependent fit of the model-independent fit can be performed to extract physically meaningful values. During the test of calculated reflectivities from films with known electron density profiles, *StochFit*, in many cases, found the global minimum for the solution, and in the worst case was equivalent to the most common model-dependent fit. Additionally, *StochFit* implements a fully featured model-dependent fitting module that allows one to sample the parameters space stochastically for a given number of boxes, both with and without constraints. By utilizing an intuitive graphical user interface and high-quality graphical output, *StochFit* allows the user to generate fits to measured reflectivities with little more than knowledge of the composition of the superphase and the subphase, and a reasonable speculation of the film thickness and average electron density.

XR for the Au Langmuir monolayers was performed at the Advanced Photon Source at Argonne National Laboratory. The initial MI algorithm for *StochFit* was developed by DL. SMD would like to thank Andrew Nelson for advice on generating reflectivities and the finer points of XR analysis. SMD would also like to thank Jarek Majewski for many helpful discussions related to XR and neutron data analysis. ChemMatCARS Sector 15 is supported by the National Science Foundation/Department of Energy under grant No. CHE-0535644. Use of the Advanced Photon Source was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, under contract No. DE-AC02-06CH11357. SMD was supported in part by the University of Chicago MRSEC program of the NSF DMR-0213748 and the GAANN Department of Education award No. P200A000222. This work was partially supported by the Packard Foundation (99-1465) and the NSF (MCB-0616249).

References

Bengu, E., Salud, M. & Marks, L. D. (2001). *Phys. Rev. B*, **6319**, 195414.
Braun, C. (1997). *Parratt32*. Version 1.5. HMI Berlin, Germany.
Černý, V. (1985). *J. Optim. Theory Appl.* **45**, 41–51.
Choi, J. W., Cho, K. S., Rhee, H. W., Lee, W. H. & Lee, H. S. (1998). *Bull. Korean Chem. Soc.* **19**, 549–553.
Chou, C. H., Regan, M. J., Pershan, P. S. & Zhou, X. L. (1997). *Phys. Rev. E*, **55**, 7212–7216.
Croce, P. & Nevot, L. (1976). *Rev. Phys. Appl.* **11**, 12.
Dane, A. D., Veldhuis, A., de Boer, D. K. G., Leenaers, A. J. G. & Buydens, L. M. C. (1998). *Physica B*, **253**, 254–268.
Ege, C., Majewski, J., Wu, G., Kjaer, K. & Lee, K. Y. C. (2005). *Chem. Phys. Chem.* **6**, 226–229.
Fenter, P. & Zhang, Z. (2005). *Phys. Rev. B*, **72**, 4.
Hamley, I. W. & Pedersen, J. S. (1994). *J. Appl. Cryst.* **27**, 29–35.
Kago, K., Endo, H., Matsuoka, H., Yamaoka, H., Hamaya, N., Tanaka, M. & Mori, T. (1998). *J. Synchrotron Rad.* **5**, 1304–1308.
Kirkpatrick, S., Gelatt, C. D. & Vecchi, M. P. (1983). *Science*, **220**, 671–680.
Knapp, G. S., Beno, M. A. & You, H. (1996). *Annu. Rev. Mater. Sci.* **26**, 693–725.
Laub, C. F. & Kuhl, T. L. (2006). *J. Chem. Phys.* **125**, 244702.
Lee, A. van der, Salah, F. & Harzallah, B. (2007). *J. Appl. Cryst.* **40**, 820–833.
Lin, B., Meron, M., Gebhardt, J., Graber, T., Schlossman, M. & Viccaro, P. J. (2003). *Physica B*, **336**, 75–80.
Lin, B. H., Schultz, D. G., Lin, X. M., Li, D. X., Gebhardt, J., Meron, M. & Viccaro, P. J. (2007). *Thin Solid Films*, **515**, 5669–5673.
Majkrzak, C. F. & Berk, N. F. (1996). *Physica B*, **221**, 520–523.
Marquardt, D. W. (1963). *J. Soc. Ind. Appl. Math.* **11**, 431–441.
Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. (1953). *J. Chem. Phys.* **21**, 16.
Miller, C. E., Majewski, J., Faller, R., Satija, S. & Kuhl, T. L. (2004). *Biophys. J.* **86**, 3700–3708.
Miller, C. E., Majewski, J. & Kuhl, T. L. (2006). *Colloids Surf. A Physicochem. Eng. Asp.* **284–285**, 434–439.
Nelson, A. (2006). *J. Appl. Cryst.* **39**, 273–276.
Parratt, L. G. (1954). *Phys. Rev.* **95**, 359–369.
Pedersen, J. S. (1992). *J. Appl. Cryst.* **25**, 129–145.
Politsch, E. & Cevc, G. (2002). *J. Appl. Cryst.* **35**, 347–355.
Sakurai, K. & Atsuo, I. (1992). *Jpn. J. Appl. Phys.* **31**, L113–L115.
Sanyal, M. K., Agrawal, V. V., Bera, M. K., Kalyanikutty, K. P., Daillant, J., Blot, C., Kubowicz, S., Konovalov, O. & Rao, C. N. R. (2008). *J. Phys. Chem. C*, **112**, 1739–1743.
Sanyal, M. K., Hazra, S., Basu, J. K. & Datta, A. (1998). *Phys. Rev. B*, **58**, R4258.
Schultz, D. G., Lin, X. M., Li, D. X., Gebhardt, J., Meron, M., Viccaro, P. J. & Lin, B. H. (2006). *J. Phys. Chem. B*, **110**, 24522–24529.
Stoev, K. & Sakurai, K. (1997). *Rigaku J.* **14**, 22–37.
Tolan, M. (1999). *X-ray Scattering with Coherent Radiation, X-ray Scattering from Soft-Matter Thin Films*, edited by G. Hohler, pp. 151–168. Berlin: Springer.