Quantitative LEED analysis using a simultaneous optimization algorithm

2008 J. Phys.: Condens. Matter 20 304201

(http://iopscience.iop.org/0953-8984/20/30/304201)

J. Phys.: Condens. Matter **20** (2008) 304201 (9pp) doi:10.1088/0953-8984/20/30/304201

Quantitative LEED analysis using a simultaneous optimization algorithm

M Blanco-Rey[1,2], K Heinz[3] and P L de Andres[1]

[1] Instituto de Ciencia de Materiales (CSIC), Cantoblanco, 28049 Madrid, Spain
[2] Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK
[3] Lehrstuhl für Festkörperphysik, Universität Erlangen-Nürnberg, Staudtstrasse 7, 91058 Erlangen, Germany

Received 3 January 2008, in final form 13 February 2008
Published 8 July 2008
Online at stacks.iop.org/JPhysCM/20/304201

Abstract
The performance of a combinatorial simultaneous optimization (SO) algorithm is tested using experimental LEED $I(E)$ data from Cu(100) and $Fe_{0.57}Al_{0.47}$(100) surfaces. SO optimizes structures taking advantage of the experimental database at two levels: (i) commensurate subsets of the database with the number of unknown parameters are chosen to find local solutions using Broyden's method and (ii) these partial structural solutions are used to build a Markov chain over the whole database. This procedure is of global character, the same as simulated annealing or genetic algorithm methods, but displays a very competitive scaling law because after the first iteration candidates are not chosen by a blind/random pick; they are already solutions to the problem with a restricted experimental database.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Most properties of surfaces (e.g. their reactivity) are strongly influenced by their structure. A number of diffraction techniques, such as low-energy electron diffraction (LEED) [1–3], photoelectron diffraction (PD) [4], surface x-ray diffraction (SXRD), [5, 6], near-edge x-ray absorption fine structure (NEXAFS), [7] etc [1, 2, 8, 9] have been developed in surface science to find this structure with atomic accuracy. Among these techniques, LEED plays a prominent role in the field due to similar progress made in experimental and theoretical techniques, and nowadays is a technique that can be found in nearly every surface science laboratory. Experimental intensities for different Bragg beams diffracted from the surface are measured as a function of the energy of the primary beam of electrons; these carry the information on the geometrical position of atoms in the surface.

Diffraction amplitudes are complex-number functions of the parameters defining the structure; these functions may be multivalued and their inverse has to be found inside the physically appropriate Riemann sheet (the existence of several branches is usually referred to in the literature as 'multiple coincidences'). In principle nothing prohibits us from finding the inversion algorithm linking the structure and the measured intensities, which would provide us with the geometrical parameters in just one single step. In practice, several technical problems make this a difficult program; we shall only mention the following two difficulties: (1) the impossibility of measuring the diffracted wavefield phase makes the inversion difficult as a great deal of information is carried away with the phase itself and, (2) the strong interaction between the incident wavefield with the surface electrons, i.e. multiple scattering, makes the direct/inverse functions highly nonlinear, and quite difficult to localize the physically meaningful branches. Therefore, the standard approach over the years has been to minimize a cost function, the so-called R-factor, designed to quantify the agreement/discrepancy between experimental and calculated spectra. This in turn is not free of complications because the R-factor is a scalar function of a vector argument defining all the structural and non-structural parameters needed to define the system, leading to a non-polynomial scenario in the search for the global minimum that could only be solved with complete certainty in an infinite amount of searching time. We should mention here how Professor John Pendry and his collaborators have led the way in identifying and finding solutions to all the aforementioned fundamental problems for the LEED technique [1, 10–13].

In this paper we analyse a new route to overcome the intrinsic non-polynomial (NP) nature of traditional LEED analysis based on the use of an R-factor, in particular Pendry's

R-factor [11]. The method is applied to experimental data, and its performance is compared with state-of-the-art solutions proposed in the literature. More precisely, we want to solve the following problem. If we let N be the number of parameters defining the structure the R-factor is a scalar function defined in $\Re^N$. In general, this hypersurface displays a complex topography, showing several local minima. If we intend to find out its global minimum by a simple brute force search, i.e. evaluating the R-factor for every point of a grid in parameter space, we find that the number of trial model structures, t, scales exponentially with N. In the case of a grid where the number of values in each of the N dimensions is equal to n_p, $t = n_p^N$ structures should be tried. Because the time needed to find the structure scales exponentially with the number of unknown parameters, optimization theory classifies this as a non-polynomial or NP-complete problem [13]. In particular, if we carry out an exhaustive search on a continuous interval of length L for each parameter, making sure that the solution is at the worst case at a maximum distance ϵ from the true solution, we obtain:

$$t \simeq \left(\frac{L\sqrt{N}}{\epsilon}\right)^N. \tag{1}$$

For high N values, this scaling law, $t \sim N^{N/2}$, is even worse than the exponential law that appears in the grid case.

We will restrict ourselves to the determination of an ordered surface structure from quantitative LEED. For a single incidence direction, the database is formed by a set of $I(E)$ curves. These are intensity versus energy curves, one per reflection or 'beam'. For a given model surface, the accurate evaluation of $I(E)$ requires using a multiple scattering formalism, which is computationally expensive. Improvements in LEED $I(E)$ analysis efficiency made so far have been focused in the development of efficient intensity evaluation methods, such as Tensor-LEED (TLEED) [12], and in the development of algorithms that reduce the number of intensity evaluations during the optimization procedure. The present paper deals with the second subject.

In their pioneering contribution to develop direct methods for LEED, Pendry, Heinz and Oed introduced the idea of reducing the traditional optimization method to a multidimensional root finding. Starting from a linearized version of TLEED for the scattering amplitudes, they have proposed an iterative solution for the related nonlinear system of equations for intensities [13]. This class of methods fits all the datapoints simultaneously, since it involves the minimization of a vectorial function $f : \Re^N \to \Re^N$. Indeed direct methods based on the holographic approach [14], and the quasi-direct method proposed by the direct fitting of a subset of experimental intensities have been the two most promising routes to find solutions for structural problems that could easily get out of hand as the system becomes more and more complex.

Multidimensional function minimization techniques can be classified into local and global. Local methods (e.g. steepest descent making downhill moves in the R-factor hypersurface [15–17]) have the advantage of scaling like the number of unknowns to be found squared (N^2), but they have the disadvantage of exploring only the nearest single local minimum. Alternatively, global methods allow uphill moves on the hypersurface to explore several minima by crossing between 'valleys'. Most of these methods work by constructing Markov chains of structures and are based on the ergodic principle. Some examples are genetic algorithms (GA) [18], random sampling algorithms (RSA) [19] and simulated annealing (SA) [20]. All of them would find the correct global minimum if given infinite time; however, within a finite time interval the degree of success varies depending on the particular circumstances and has to be studied independently, often to optimize the parameters in the search. As an example of selection of different parameters, the simulated annealing algorithm is interpreted as a physical adiabatic cooling performed under quasi-equilibrium conditions, and the cooling scheme is chosen to achieve a robust performance [20, 21].

In a previous paper, we have proposed simultaneous optimization (SO) [22] based on a number of theoretical simulations, and we have shown it can be successful in two highly nonlinear scattering problems: single atom phase shift retrieval from backscattered intensities and surface structure optimization from LEED simulated data. In the present paper, the SO method is described in full detail and its performance is tested against real LEED experimental $I(E)$ data measured for Cu(100) and $Fe_{0.53}Al_{0.47}$(100) surfaces. These two systems have also been studied with RSA, allowing us to take a benchmark on performance against a different method known already to be efficient and robust. The SO algorithm consists of two shells: (i) an inner shell where structures are obtained by solving nonlinear systems of equations by applying Broyden's method [23] (at this point only a random subset of the experimental database is used) and (ii) an outer shell, where the structures obtained in the inner shell are validated against the whole experimental database by using Pendry's R-factor.

2. The algorithm

Several R-factor prescriptions can be considered for the structural problem [2]; the basic task consists in comparing values for calculated and experimental intensity curves, $I(E)$, but first-and/or second-order derivatives of the $I(E)$ curves might be used too.

Let us consider a simple and basic R-factor, related to a least-squares fit between experimental and theoretical intensities, that has been introduced in the literature by the name of R_2 and can be reduced to a χ^2 minimization problem for the case where all the experimental points have the same statistical variance [24]:

$$R \sim \sum_{\vec{\alpha}} \{q^{\text{th}}(\vec{x}, \vec{\alpha}) - q^{\text{exp}}(\vec{\alpha})\}^2 \tag{2}$$

where q is used to indicate that any convenient function of the intensities might be used for the comparison, $\vec{x} = (x_1, \ldots, x_N)$ are the N unknown structural parameters, and $\vec{\alpha} = (E_i; \vec{g})$ runs over energies and beams. The full experimental database is made of intensities measured for different energies, E_i, in every accessible beam, $\vec{g}$: $\mathcal{S} = \{I_{\vec{\alpha}}\}$, containing N_D numbers.

From quite general statistical arguments, we expect the R-factor hypersurface to have a number of shallow local minima, related to random correlations between theory and experiment, and a few deeper wells due to structural coincidences [20]. Thus, if a downhill method were to be used, an independent minimization would be needed at every local minimum. Since the number of minima grows exponentially with N, such a multiple launch strategy is an NP problem.

A most popular R-factor, R_P, has been defined by Pendry [11]. Intensities are affected by background contributions of temperature effects and electron damping, so R-factors that make use of $I(E)$ derivatives are, in general, more reliable because these contributions are largely cancelled out. In particular, R_P treats all maxima and minima at the $I(E)$ equally, since all of them contain useful information about constructive/destructive interference conditions closely related to the structure. In fact, Pendry's R-factor compares the so-called Y-functions, based on logarithmic derivatives, to highlight all the extrema found in the spectra, and R_P quantifies the agreement between experimental and calculated Y-functions.

Usually, the set of experimental data, $\mathcal{S}$, contains many more datapoints, N_D, than geometrical parameters in the model, N. As pointed out by Kleinle *et al*, R-factors using the information contained in their derivatives need to be calculated on a fine grid of energies where some datapoints are correlated. These authors suggested computing an R-factor that only depends on intensities on a coarse energy grid [25]. From a formal point of view, only a small subset of N independent datapoints $\mathcal{S}_N \in \mathcal{S}$ should be needed to uniquely determine the structure. In practice, however, multiple coincident solutions might be found for a reduced subset of the experimental database. A characteristic for the global solution is that it is independent with respect to different choices for the subset $\mathcal{S}_N$. This kind of procedure constitutes a list of structures converging to the 'global' one, where the starting points do not need to be chosen at random, but are already partial solutions to the global problem. The SO algorithm has been designed to evolve and validate solutions obtained with R_P from reduced database subsets.

We find that in most cases, these restricted subsets already contain the relevant topographic features of R_P, although they might also display features not related to the global hypersurface. Obviously, these solutions constitute far better trial structures than simple random moves on the R_P hypersurface. However, dealing with experimental or noisy data makes $R_P \neq 0$ at the global minimum. Therefore, the global solution is not necessarily a solution of every partial subset, and determining the global solution can be more difficult, in particular the algorithm could be trapped oscillating around different local minima. It is possible to damp and minimize these oscillations by using overdetermined systems, because by increasing the size of the system of equations the local topography becomes more and more similar to the global one. We have found that for overdetermined systems of the order of $N_D = 2N$ or $N_D = 3N$, the computing effort at the inner shell remains reasonable, while the efficiency increases considerably.

Having in mind that our goal is to find the global minimum in R_P, we first define a database subset, $\tilde{\mathcal{S}} \in \mathcal{S}$, such that it includes intensities where $\frac{\partial I(E;\vec{g})}{\partial E} \approx 0$ for at least one of the experimental curves (let us call the cardinal of this subset $\tilde{N}_D \geqslant N$). This choice of $\tilde{\mathcal{S}}$ is arbitrary, but justified by the fact that those are the points contributing more to Pendry's R_P.

A system of N_{eq} equations and N unknowns is constructed with N_{eq} intensities chosen at random inside $\tilde{\mathcal{S}}$. We denote this subset by: $\mathcal{S}_{N_{\mathrm{eq}}} \in \tilde{\mathcal{S}}$ ($\tilde{N}_D \geqslant N_{\mathrm{eq}} \geqslant N$). Thus, $\binom{\tilde{N}_D}{N_{\mathrm{eq}}}$ different choices of systems of equations are available. We define the following function:

$$f_i(\vec{x}; E_i) = \sum_{\vec{g}} |Y^{\mathrm{th}}_{\vec{g}}(E_i; \vec{x}) - Y^{\mathrm{exp}}_{\vec{g}}(E_i)| = 0; \quad i = 1, \ldots, N_{\mathrm{eq}} \tag{3}$$

where $(E_1, \ldots, E_{N_{\mathrm{eq}}})$ is a set of N_{eq} different experimental points taken from $\tilde{\mathcal{S}}$ at random. Summation over beams is performed to gather as much information as possible and because it results in a smoother function than the one corresponding to the $\vec{g}$ value associated with the extreme in the curve. This is relevant, since derivatives need to be calculated in the inner shell (see section 2.2).

2.1. The outer shell: the Markov chain

The global search starts with a structure chosen at random inside the physically accessible region. This will be validated in the outer shell by an iterative process. The kth iteration in the outer shell starts by calling the inner shell with a structural candidate, $\vec{x}^{(k)}$, and a subset of experimental datapoints, $\mathcal{S}_{N_{\mathrm{eq}}}$, chosen at random in $\tilde{\mathcal{S}}$. The inner shell returns as output of a structure, $\vec{x}'^{k}$. Before computing R_P for the global database, which is more expensive from a computational point of view than invoking the inner shell, it is worth checking that (i) the new structure, $\vec{x}'^{k}$, is also a solution for a different choice of $\mathcal{S}_{N_{\mathrm{eq}}}$ and (ii) whether the new solution stays within a given convergence radius from the old one. $N_{\mathrm{PC}} \sim N$ inner shell calls are made inside the kth iteration on the outer shell to reduce the number of times R_P needs to be computed.

The $(k+1)$th iteration begins with a structure, $\vec{x}^{(k+1)}$, incorporating small random modifications over the solution in the kth iteration. Unlike SA, the modifications are always accepted, which is known in the literature as blind random search iterative improvement, and it is a global search method [26]. Following the RSA idea [19], $\vec{x}^{(k+1)}$ is obtained by considering a Gaussian probability distribution centred at the solution in the kth iteration, where the width of the Gaussian, σ, depends on the value of $R_P(\vec{x}'^{k})$, computed at the end of the kth iteration. Therefore, uphill moves are guaranteed and the search space size is reduced for small values of R_P, where long jumps are not necessary.

The partial problem tackled in the inner shell is highly nonlinear; it might happen that the system solver does not find the right solution. However, wrong trial structures coming from the inner shell usually already have a few useful components, making better candidates than a pure random blind choice made by conventional algorithms based in the ergodic theorem. Non-Gaussian probability distributions, such

as Cauchy–Lorentz, have been used in the literature to generate a Markov chain of structures [21]. In our case, we have seen that the distribution shape is not crucial since the system solver method in the inner shell is already quasi-global, and it overcomes R_P hypersurface simple barriers.

2.2. The inner shell: the system solver

To deal with the problem of solving the highly nonlinear system $\vec{f} = 0$, a globally convergent algorithm that combines a multidimensional secant method with a descent strategy is used [24]. The only requirement is that $\vec{f}$ is a continuous and differentiable function with respect to $\vec{x}$. The ith iteration moves in a conventional multidimensional secant method, $\delta\vec{x}^{(i)} = \vec{x}^{(i+1)} - \vec{x}^{(i)}$, and is given by:

$$\begin{aligned} \tilde{J}^{(i)} \cdot \delta\vec{x}^{(i)} &= -\vec{f}^{(i)} \\ \tilde{J}^{(i+1)} \cdot \delta\vec{x}^{(i)} &= \delta\vec{f}^{(i)} \end{aligned} \tag{4}$$

where $\delta\vec{f}^{(i)} = \vec{f}^{(i+1)} - \vec{f}^{(i)}$ and $J^{(i)}_{jk} = \frac{\partial f^{(i)}_j}{\partial x^{(i)}_k}$ are the Jacobian matrix elements. In order to save derivative evaluations, we substitute the exact Jacobian by Broyden's approximation [23],

$$\tilde{J}^{(i+1)}_B = \tilde{J}^{(i)}_B + \frac{(\delta\vec{f}^{(i)} - \tilde{J}^{(i)}_B \cdot \delta\vec{x}^{(i)}) \otimes \delta\vec{x}^{(i)}}{\delta\vec{f}^{(i)} \cdot \delta\vec{f}^{(i)}} \tag{5}$$

rather than the exact Jacobian $\tilde{J}^{(i+1)}$. The procedure is initialized with the exact Jacobian. When the system $\vec{f} = 0$ is overdetermined, the linear system of equations (equation (4)) is solved in a least-squares sense.

However, this method tends to wander around the parameter space if the starting point is not close enough to the solution. This can be avoided by embedding the procedure in a globally convergent strategy that tries to minimize the scalar function $|\vec{f}|^2$. The latter defines a hypersurface that can be seen as a projection of the global R_P onto the data subspace $\mathcal{S}'_{N_{\rm eq}}$. If we use the exact Jacobian $\tilde{J}$, $\delta\vec{x}^{(i)}$ yields already a descent direction for $|\vec{f}|^2$, but the whole step may not make $|\vec{f}|^2$ smaller. To ensure it does, a line search is performed along this vector,

$$\vec{x}^{(i+1)} = \vec{x}^{(i)} + \lambda\delta\vec{x}^{(i)} \tag{6}$$

taking the λ value that yields the smallest value of $|\vec{f}|^2$. Sophisticated line searches could be made at this stage [24], but an equispaced sampling along the line is found to be enough for our purposes. Finally, at the end of each iteration, the exiting criterion is checked. We use convergence in the modulus of $\vec{x}$ as the exiting condition:

$$|\vec{x}^{(i+1)} - \vec{x}^{(i)}| < x_{\rm min}. \tag{7}$$

This method is non-local, because, even if the line search takes place along a descent direction of $|\vec{f}|^2$, a longer step can overcome a hypersurface barrier. If $\tilde{J}^{(i+1)}_B$ deviates from the descent direction, the procedure can be restarted with the exact Jacobian.

Every solution to equation (3) minimizes $|\vec{f}|^2$. However, equation (3) may eventually have multiple solutions, and only one of them coincides with the global minimum of the R-factor. The others correspond to local minima of the R-factor. In fact, the $|\vec{f}|^2$ hypersurface shares topographic features with the R-factor.

However, that is true only for the ideal case of perfect agreement between theory and experiment at the global minimum, where $R_P = 0$. In a realistic scenario, the global minimum has $R_P > 0$. This implies that there might exist partial problems that do not possess a solution at the global minimum. Adding more datapoints to the partial problem should filter out oscillations near the R-factor minima and reconcile the partial problem topographies. We define the oversampling factor $N_{\rm ov}$ for the partial problem such that the systems contain $N_{\rm eq} = N_{\rm ov} \times N$ equations and N unknowns, and solve linear equations (4) in a least-squares sense. Although $N_{\rm ov}$ should be chosen for each specific problem according to the fit quality, the condition $N_D \gg N_{\rm eq}$ must hold in order to keep the partial problem tractable at a low computational cost.

2.3. Scaling law estimate

By construction, each Broyden approximate Jacobian evaluation requires $2N$ intensity evaluations, while the exact Jacobian would need N^2. The line search makes a constant number of evaluations, and R_P is evaluated every $N_{\rm PC} \sim N$ partial problem calls. Thus, the total scaling exponent is $\beta \approx 2$ before entering the Markov chain. On the other hand, we can expect the outer shell to have a maximum scaling similar to the RSA algorithm, which has an exponent $\beta \approx 2.5$ [19]. Notice that the present estimate considers the number of intensity evaluations only, and not other N-dependent operations present in the algorithm that affect the total computation time, such as matrix inversions. This analysis gives an upper boundary value for the exponent, $\beta \lesssim 4.5$. SO yields efficient scaling laws because the structures entering the Markov chain are by construction partial solutions to the global problem, i.e. the outer shell is similar to an importance sampling, rather than to a standard random sampling. In fact, we have already seen in the statistical results from noiseless simulated LEED $I(E)$ data of the Ir(110)-p(2×1) surface, that SO yields an exponent $\beta = 4.1 \pm 0.1$ [22]. A comparable value, $\beta = 3.5 \pm 0.2$, was found in the atomic phase shift retrieval from noiseless backscattered intensities from a single atom [22].

From the arguments above, it is expected that the use of overdetermined nonlinear systems will affect the scaling exponent only slightly, though the total computational effort will increase. This has been observed in the problem of atomic phase shift retrieval from single atom noisy backscattered intensities [22]. In that example, a convincing success rate could be obtained at low values of $N_{\rm ov}$. Adding 5% of Gaussian noise in the intensities and applying oversampled SO resulted in a scaling law exponent $\beta = 4.4 \pm 0.3$, this value slightly increased to $\beta = 4.8 \pm 0.4$ when the noise was 10%. For both noise values, the success rate above 90% could be routinely obtained using $N_{\rm ov} = 2$ when searching for $l_{\rm max} < 5$ phase shifts, and $N_{\rm ov} = 4$ for $l_{\rm max} \geqslant 5$.

The main quantity used to describe the efficiency of SO is the computing effort, defined as the expected value

$$\langle t \rangle = \frac{\langle N_{\text{call}} \rangle}{N_e} \tag{8}$$

where N_{call} is the number of individual intensity evaluations and N_e is the number of energy points per $I(E)$ curve. In those previously published theory–theory benchmarks [22], it was found that for each value of N the probability distribution function (PDF) of a number N_{stat} of independent searches shows a peak at low t values and is exponentially decreasing. Interestingly, it can be fitted to a one-event Poisson-like distribution:

$$p(t) = w^2 t \mathrm{e}^{-wt}. \tag{9}$$

This is the PDF of a single random event taking place in a predetermined time interval between t and $t + \mathrm{d}t$, the event being that SO finds the global minimum. The area under $p(t)$,

$$P(t) = \int_0^t p(t')\,\mathrm{d}t' = 1 - (1 + wt)\mathrm{e}^{-wt} \tag{10}$$

gives the success rate of a search that uses a computing effort t.

3. Results with experimental data

3.1. Clean Cu(100) surface

SO performance has been benchmarked using experimental normal incidence LEED $I(E)$ curves from a clean Cu(100) surface. Pendry's R-factor, R_P, is used to quantify the agreement between experimental and calculated spectra. The experimental data were taken at a low temperature (90 K) in order to reduce thermal diffuse scattering (details can be found elsewhere [27]). The sample quality was good (common impurities, such as C, O and S, were below the Auger detection limit) and the $I(E)$ curves were reproducible with $R_P < 0.02$. R_P values among symmetry related beams lying below 0.04, ensuring a correct sample alignment. Previous dynamical analyses of this surface have been reported in [27] and [28]. It was found that it is necessary to introduce an energy dependent inner potential ($V_{\text{or}}(E)$) in order to accurately reproduce the correct lattice parameter, $a_0 = 2.55$ Å [28]. A theoretical model for $V_{\text{or}}(E)$ can be parametrized from first principles [29]. When a constant V_{or} is used, the $I(E)$ fit yields a contracted in-plane lattice parameter value of $a_0 = 2.53$ Å [27]. This deviation of hundredths of Å lies beyond the R_P variance limit due to the systematic error caused by neglecting $V_{\text{or}}(E)$. To discuss the efficiency of the SO algorithm when fitting geometrical data, for simplicity, we will restrict ourselves to a constant V_{or}, use the smaller $a_0 = 2.53$ Å and keep Debye parameters fixed at their optimum values. The database used in this work consists of four beams, namely (10), (11), (20) and (21), measured in the energy range 55–410 eV and producing a total data base width of $\Delta E = 940$ eV. The maximum energy has been chosen to be smaller than the 500 eV used in the original calculations [27], so as to make the $V_{\text{or}}(E)$ dependence less relevant for the fit. A maximum angular momentum number $l_{\text{max}} = 11$ achieves convergence such that differences in R_P are smaller than 0.001.

An energy-averaged value of V_{or} is also to be fitted. Conventional directed algorithms act upon the whole $I(E)$ curve, so that it is usually enough to apply a rigid shift to the $I(E)$ curve to make peaks coincide and obtain a good estimate of V_{or}. The straight-forward way of introducing V_{or} in a SO search would be to use it only in the outer shell of the algorithm. In the first iterations, when calculated and experimental $I(E)$ curves are poorly correlated, the estimated value of V_{or} will be, in general, in disagreement with the actual value. If this incorrect V_{or} is kept fixed during inner shell iterations, it will prevent the algorithm from approaching the solution. A better strategy consists in considering V_{or} as an extra unknown in the inner shell, thus adding another equation to the system. We have used this second procedure in subsequent calculations.

As a reference, we will use the structure determined full-dynamically by Müller *et al* [27], with $a_0 = 2.53 \pm 0.01$ Å for the lattice parameter, and the following interlayer distances (ordered from vacuum to bulk): $d_1 = 1.765 \pm 0.005$ Å, $d_2 = 1.805 \pm 0.010$ Å, $d_3 = 1.80 \pm 0.01$ Å, $d_4 = 1.79 \pm 0.02$ Å, $d_5 = 1.80 \pm 0.03$ Å, $d_6 = 1.79 \pm 0.04$ Å, and bulk interlayer distance $d_{\text{b}} = 1.79 \pm 0.07$ Å. Deeper vertical distances have larger error bars because of electron attenuation. Other non-structural parameters are $V_{\text{oi}} = 4.68$ eV, and isotropic thermal vibration amplitudes $v_1 = 0.12$ Å for the topmost layer atoms and $v_{\text{b}} = 0.07$ Å for subsurface atoms, determined within a Debye–Waller-like approach. These values yield a minimum $R_{P_{\text{min}}} = 0.085 \pm 0.013$ for a database of size $\Delta E = 1600$ eV, where the R-factor variance is $\text{var}(R_{P,\text{min}}) = R_{P,\text{min}}\sqrt{\frac{8V_{\text{oi}}}{\Delta E}}$ [11]. If the reduced database ($\Delta E = 940$ eV) is used, the same structure yields $R_{P,\text{min}} = 0.15 \pm 0.03$. We shall use this value to define the lowest boundary for exiting conditions of the statistical searches as $R_P < R_{P,\text{min}} + \text{var}(R_{P,\text{min}})$. Thus, for $N = 6$ interlayer distances, the search is finished when $R_P < 0.18$. For smaller values of N, best fit structures yield larger $R_{P,\text{min}}$ values: $R_{P,\text{min}}(N = 2) = 0.18$, $R_{P,\text{min}}(N = 3) = 0.17$, $R_{P,\text{min}}(N = 4) = 0.16$ and $R_{P,\text{min}}(N = 5) = 0.16$. The corresponding variance is ∼0.03. Therefore, the following search exiting R_P values are used: 0.21 for $N = 2$, 0.20 for $N = 3$ and 0.19 for $N = 4, 5$.

In order to examine the behaviour of SO when handling real experimental data, we perform N_{stat} equivalent statistical searches using random starting points inside the search space, which will be used to evaluate averages. The search space consists of the topmost N interlayer spacings of the Cu(100) surface, plus a constant value of V_{or} in the range 2–10 eV. Intensities are calculated in the energy range 55–410 eV with a step of 3 eV. Thus, the nonlinear systems of equations in the inner shell contain $N + 1$ unknowns. Size effects have been modelled using two different search spaces of hypervolume 1 Å^N and (0.4 Å)^N. The latter corresponds to the usual tensor-LEED applicability range.

Figure 1 shows the R_P map for $N = 2$ in the 1 Å^2 space. The global minimum is located $(d_1, d_2) = (1.75, 1.81)$ Å with $R_P = 0.18$, and also a few secondary deep wells with $R_P \lesssim 0.5$ are found. The surface shows a strong corrugation, and several local shallow minima with $R_P \sim 0.8$. As pointed out by Rous *et al* [16], this topological feature is independent

Table 1. Average t values corresponding to histograms of figure 1. $\langle t\rangle$ is the histogram average, $\langle t\rangle_E = 2/w$ is the expected value obtained after a least-squares fit of the histograms to a function $p(t) = w^2 t e^{-wt}$ and $\langle t\rangle_P$ ($P = 0.90, 0.9998$) are the average t values needed to get a successful search with P probability. The values marked by (*) have been obtained from a fit to $p(t) = \lambda e^{-\lambda t}$.

	0.4 Å^3				1 Å^3			
N_{ov}	$\langle t\rangle$	$\langle t\rangle_E$	$\langle t\rangle_{0.90}$	$\langle t\rangle_{0.9998}$	$\langle t\rangle$	$\langle t\rangle_E$	$\langle t\rangle_{0.90}$	$\langle t\rangle_{0.9998}$
1	160 ± 10	146 ± 6	284	805	1620 ± 150 (*)	1270 ± 160	3729 (*)	13 793 (*)
2	100 ± 6	88 ± 4	170	482	1050 ± 120	960 ± 100	1873	5 297
3	110 ± 7	97 ± 5	188	532	1550 ± 150 (*)	1600 ± 300	3578 (*)	13 235 (*)

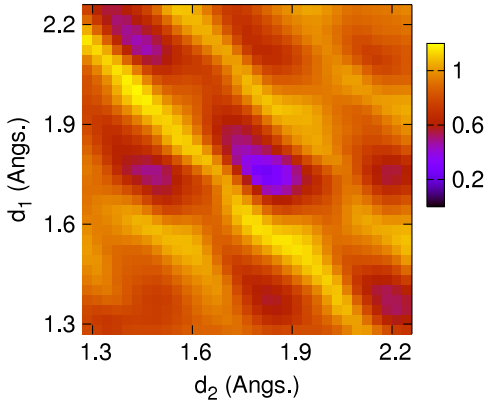

Figure 1. R_P map for Cu(100) data in a search space of area 1 Å^2. Parameters (d_1, d_2) are the two topmost interlayer spacings. The real part of the inner potential is also optimized for each structure. The global minimum ($R_{P,\min} = 0.18$) is located at $(d_1, d_2) = (1.75, 1.81)$ Å. Secondary deep minima appear at $(d_1, d_2) = (2.14, 1.45)$ Å with $R_P = 0.42$, $(d_1, d_2) = (1.75, 1.48)$ Å with $R_P = 0.48$, and $(d_1, d_2) = (1.30, 2.20)$ Å with $R_P = 0.50$.

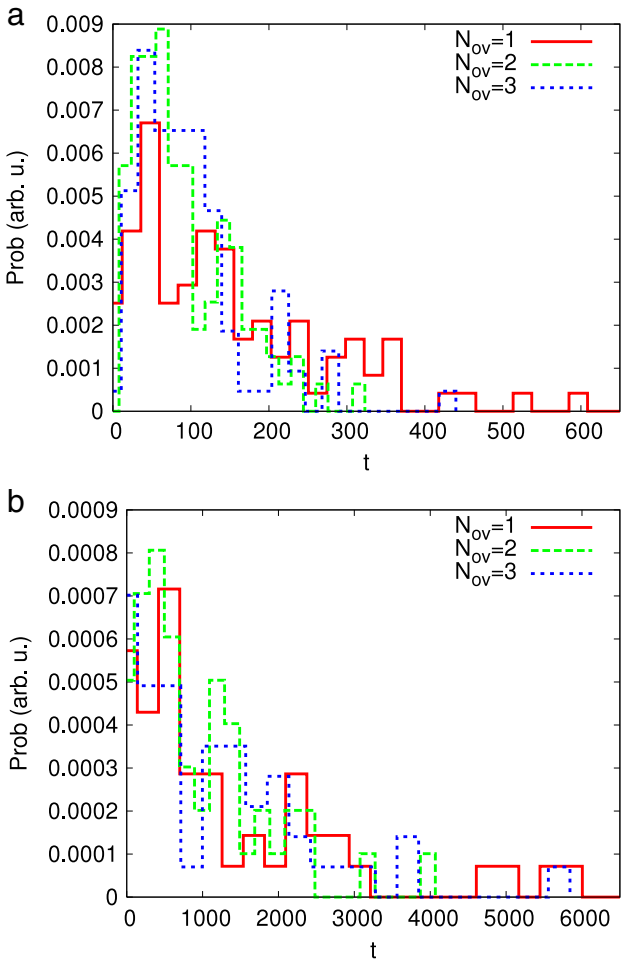

Figure 2. Panel (a) Cu(100) data in a search space of size 0.4 Å with $N_{stat} = 100$. Panel (b) Cu(100) data in a search space of size 1 Å with $N_{stat} = 50$. Search parameters are the three topmost interlayer spacings. The real part of the inner potential is also optimized for each structure. Searches are made under different inner shell oversampling conditions: $N_{ov} = 1$ (solid line), 2 (dashed) and 3 (dotted).

of the surface structure. Secondary deep minima are due to Bragg coincidences with the actual structure and shallow minima are due to random correlation between experimental and theoretical spectra. The (0.4 Å)^2 area boundaries are 1.57 and 1.97 Å, respectively. This search space contains several local minima but no secondary deep minima. Due to these differences in the R_P hypersurface topography, differences in the scaling law between both search spaces appear.

The effect of different degrees of overdetermination in the inner shell has been benchmarked for $N = 3$. Figure 2 shows the PDF of the computing effort t at values $N_{ov} = 1, 2, 3$ on both search volumes. $N_{stat} = 100$ and 50 on 1 Å^3 and (0.4 Å)^3, respectively. In both cases, $N_{ov} = 2$ appears to be a well converged value, and further overdetermination does not cause significant improvements in the efficiency.

For (0.4 Å)^3, the PDFs are non-symmetric, show a peak at low values of t and decay exponentially. The histograms can be fitted by least-squares to equation (9). If plain SO is used, i.e. $N_{ov} = 1$, decay is clearly slower. However, in the 1 Å^3 volume the $N_{ov} = 1, 3$ histograms do not show a peak, and are better fitted by an exponential function $\lambda e^{-\lambda t}$. Table 1 contains some average values obtained from those PDFs.

Figure 3 shows a log–log plot of the scaling law, represented in terms of the average computing effort $\langle t\rangle$ as a function of N. In both cases, the scaling law can be fitted to a polynomial law, $\langle t\rangle \propto N^\beta$, where $\beta = 3.45 \pm 0.02$ for 1 Å^N and $\beta = 2.3 \pm 0.2$ for (0.4 Å)^N. Averages are calculated upon $N_{stat} = 20$ and 40 for each N value, respectively. Figure 4 shows the average number of R_P evaluations, $\langle t_{R_P}\rangle$, at different N. Least-squares fits to $\langle t_{R_P}\rangle \propto N^{\beta_{R_P}}$ yield exponents $\beta_{R_P} = 1.93 \pm 0.07$ for 1 Å^N and $\beta_{R_P} = 1.0 \pm 0.2$ for 0.4 Å^N.

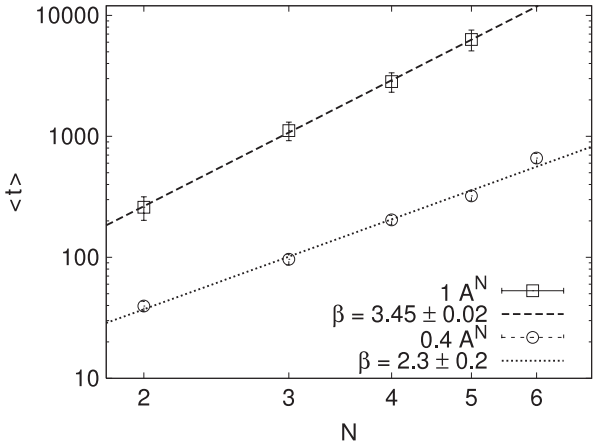

Figure 3. Scaling law for Cu(100) data with search spaces of sizes 1 Å^N (squares) and (0.4 Å)^N (circles) together with the corresponding least-squares fits to a polynomial law.

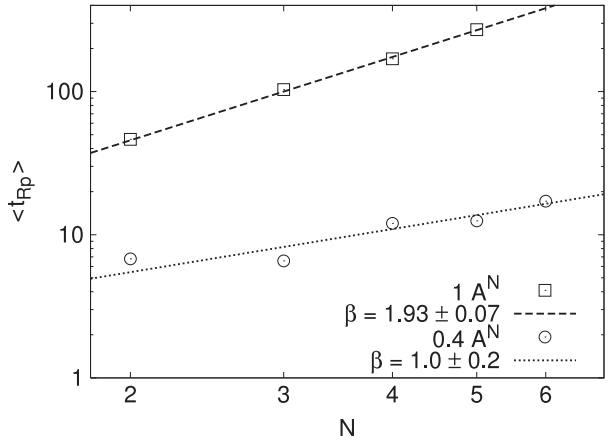

Figure 4. Scaling law for the number of R-factor evaluations in Cu(100) data with search spaces of sizes 1 Å^N (squares) and (0.4 Å)^N (circles) together with the corresponding least-squares fits to a polynomial law.

3.2. $Fe_{0.53}Al_{0.47}$(100)-1 × 1

Annealing of alloy surfaces may result in deviations of chemical composition in the surface region with respect to that of the bulk. Segregation in alloy surfaces is a well-known phenomenon that has been investigated in a number of binary alloys (see, for example, [30] and references therein). In the present section, we use experimental LEED $I(E)$ data from a $Fe_{0.53}Al_{0.47}$(100)-1 × 1 surface to benchmark the ability of SO to fit non-geometrical data, namely impurity concentration and/or thermal vibration amplitudes. FeAl crystal structure is of CsCl type, and the stoichiometric (100) surface consists of alternating Al and Fe layers. FeAl crystals exhibit a rich phase diagram, and the surface structure strongly depends on both the bulk stoichiometry and annealing temperature. Annealing of a $Fe_{0.53}Al_{0.47}$(100) surface at 650 K results in a c(2 × 2) reconstruction, whilst annealing at temperatures above 880 K produces a sharp 1 × 1 LEED pattern. We shall focus on the latter. Experimental details on surface preparation and spectra measurements can be found elsewhere [19, 31]. Early quantitative LEED calculations on this sample determined that the (100) face is Al-terminated, yielding a satisfactory $R_P = 0.12$ and the following interlayer distances: $d_1 = 1.24 \pm 0.02$ Å, $d_2 = 1.49 \pm 0.02$ Å, $d_3 = 1.47 \pm 0.02$ Å (the lattice parameter is $a_0 = 2.894$ Å [31]). The same analysis yields a ∼20% Al impurity concentration in the second layer. Since interstitial site formation is energetically unfavourable in FeAl, this result can be attributed to the strong contraction of the outermost Fe plane.

Another possible explanation is found in the coupling between chemical and vibrational parameters, which occur in LEED intensities via the Debye–Waller factor, if the average t-matrix approximation is used [32]. Blum *et al* [33] revisited this surface and found that the best structure fit can be achieved by either: (i) fitting the second Fe layer concentration of Al impurities, c_2 and the topmost Al atomic vibration amplitude v_1, or (ii) fitting two different vibration amplitudes v_1 and v_2 for the first Al and second Fe layers, respectively. Considering $V_{or}(E)$, too, these fits yielded improved R-factors, $R_P = 0.091$ and $R_P = 0.081$, respectively. Both fits are equally favourable, as they lie within the systematic error limits of the LEED analysis.

The experimental database consists of eight beams measured at $T = 120$ K in the energy range 40–500 eV with a step of 3 eV, producing a total data base width $\Delta E = 1880$ eV. The beam list is {(10), (11), (20), (21), (22), (30), (31), (32)}. In the $I(E)$ evaluations, $l_{max} = 10$ provides convergence, and the imaginary part of the inner potential is kept fixed at $V_{oi} = 6$ eV. We use $R_P \leqslant 0.12$ as the exiting condition for SO. Independent searches have been made for $N = 6$ parameters: V_{or}, three structural and two non-structural parameters. V_{or} lies in the range 6–14 eV. Structural parameters are the three topmost interlayer distances, d_1, d_2, d_3, in the ranges $d_1 = 1.16$–1.46 Å, $d_2 = 1.30$–1.60 Å and $d_3 = 1.36$–1.56 Å. The two non-structural parameters are, according to Blum *et al* [33], (i) c_2 and v_1, or (ii) v_1 and v_2. c_2 lies in the range 0–50% and the vibration amplitudes lie in the range 0.087–0.15 Å. Other vibrations are kept fixed at their bulk values, namely: v_b(Fe) = 0.09 Å and v_b(Al) = 0.12 Å. We have performed $N_{stat} = 40$ statistical searches for each type of fit using $N_{ov} = 2$. The corresponding computing effort PDFs are shown in figure 5, and average values are $\langle t \rangle = 570 \pm 70$ for a type (i) search and $\langle t \rangle = 470 \pm 70$ for type (ii).

4. Discussion

The simultaneous optimization (SO) algorithm has been successfully applied to parameter extraction from highly nonlinear scattered intensities in ideal scenarios, namely phase shift retrieval from single atom electron scattering and surface structure recovery from simulated LEED $I(E)$. In these examples, SO has performed efficiently, and works well with minimal data sets in the inner shell, i.e. it uses N datapoints $\mathcal{S}' = (E_1, \ldots, E_N)$ to fit N parameters $\vec{x} = (x_1, \ldots, x_N)$. The inner shell solves a nonlinear system of equations, a problem which is equivalent to finding the global minimum of a projected R-factor on the $\mathcal{S}'$ data subset. The

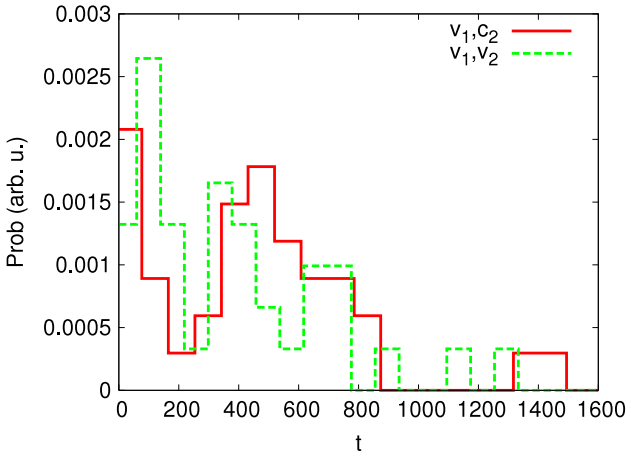

Figure 5. PDFs of the computing effort in FeAl searches using two types of non-structural parameters. The solid line corresponds to fitting c_2 and v_1, and the dotted line to fitting v_1 and v_2. Both histograms are built upon $N_{\text{stat}} = 40$ independent searches.

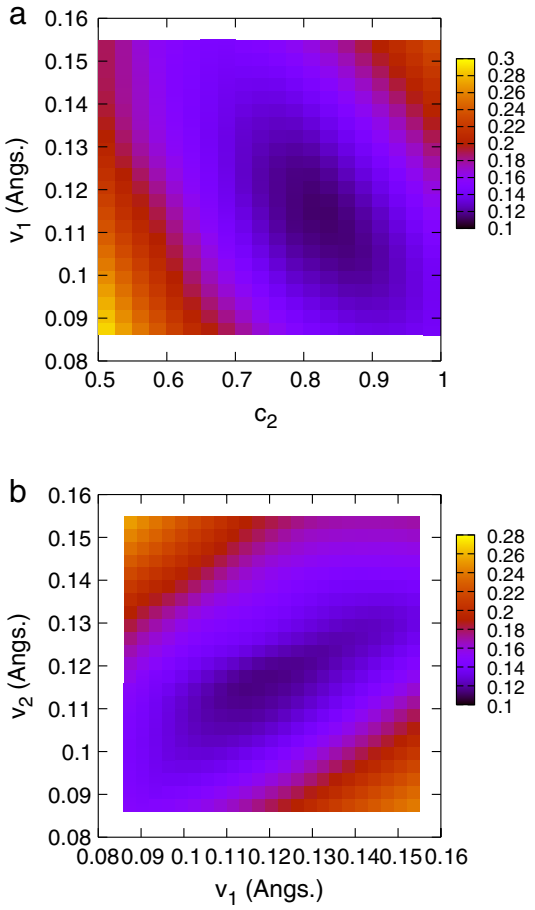

Figure 6. R_P map for $Fe_{0.53}Al_{0.47}(100)$ data as a function of (a) (v_1, c_2) or (b) (v_1, v_2) (see text), with structural parameters fixed at their best fit values.

solutions provided by the system solver must be validated against the whole data base in an outer shell by evaluating the R-factor. Therefore, SO can explore a large number of meaningful configurations in the investigated hypervolume at a low computational cost. The configurations provided by the inner shell have been shown to exhibit local minima of the R-factor at the worst case (i.e. when the system solver fails to return a solution to the system because it stagnates at a local minimum of the projected R-factor). However, when dealing with real experimental databases this is not always the case. We have found that it is not guaranteed that solutions to the partial problems correspond to local minima of the R-factor, as spurious peaks in the experimental spectra may be eventually fitted. Therefore, we generalize the implementation of SO so it can be applied to realistic situations with the same efficiency or robustness as in the theoretical case. Spurious peak fitting results in high frequency corrugation of the R-factor hypersurface. By using overdetermined systems of equations, i.e. taking $\mathcal{S}' = (E_1, \ldots, E_{N_{\text{eq}}})$ with $N_{\text{eq}} > N$, SO can filter out those corrugations, since the projected R-factor topography captures more features of the global R-factor. This method has been benchmarked using experimental LEED $I(E)$ curves in the Cu(100) surface and a search space of hypervolume 1 Å^N. After determining the optimum oversampling degree, it is found that the computational effort scales polynomially as N^β, with exponent $\beta = 3.45 \pm 0.02$, a value which is of the same order as the previously found exponent for SO working under ideal noiseless conditions. It is noteworthy that only $N_{\text{eq}} = 2(N + 1)$ are needed in the inner shell to achieve convergence in the search success rate, as shown in figure 2 for $N = 3$. Since $N_{\text{eq}} \ll N_D$, it is ensured that the number of different system choices is high enough to ensure ergodicity.

As expected, the search space size affects the scaling behaviour. In particular, if we restrict the hypervolume sufficiently to contain only one minimum inside (local or not) and if we use gradient-directed methods, we expect an ideal scaling, $\beta = 2$. It is interesting to notice that the best possible scaling law, $\beta = 1$, implies that R_P depends linearly on the parameters, and the searching hypervolume is such that the minimum stays in one of the corners of the simplex. These are too restrictive conditions, and are not likely to happen for complex problems. On the other hand, the 1 Å^N hypervolume size is big enough to contain in most of the cases several local minima in the R_P hypersurface [20], as we have confirmed in our calculations (figure 1). Therefore, it is not surprising that by reducing the searching hypervolume the exponent decreases towards the ideal value of 2. Statistical search averages yield an exponent $\beta = 2.3 \pm 0.2$ for the same system in a $(0.4\ \text{Å})^N$ hypervolume (see figure 3), which is similar to the value $\beta = 2.5$ found for RSA working in a space of the same size [19]. As shown in figure 4, the number of R_P evaluations in the outer shell is a linear function of N in the smaller space, while it is quadratic in 1 Å^N. Therefore, we conclude that the total scaling law of SO in the former case appears to be dominated by the inner shell, explaining the high efficiency of the method.

We have also tested the applicability of SO to non-geometrical parameter retrieval. Introducing V_{or} in the search as an additional parameter does not alter significantly the scaling law in the Cu(100) example. Other non-structural parameters, such as vibrational and chemical ones, can be optimized by SO. In the studied application, $Fe_{0.53}Al_{0.47}(100)$,

Al intermixing in the second layer and surface vibration amplitude are satisfactorily characterized by SO using $\langle t \rangle = 570$, which is more than the reported value for the RSA algorithm (345 trial structures) working under equivalent search conditions [19]. Interestingly, corrugation in the R_P topography is mainly due to structural parameters. Figure 6 shows R_P as a function of chemical and/or vibrational parameters with structural parameters fixed at their best fit values. Both maps have a deep flat minimum. Therefore, we expect that the statistics in this example are biased by the search for interlayer spacings.

5. Conclusions

The simultaneous optimization (SO) algorithm has been applied to experimental LEED data and characterized in detail so it can be compared with other alternatives in the literature. The use of noisy experimental LEED data does not imply a reduction of reliability or efficiency with respect to its performance on ideal noiseless databases [22]. SO brings together two apparently contradictory widespread concepts in structural work by LEED: while large parts of the database are redundant and promising structures can be obtained from a reduced number of datapoints [34], one must keep in mind that only by using large experimental databases can high accuracy and reliability be obtained. Finally, the use of the ergodic principle confers global character to the search [20].

Acknowledgments

This work has been financed by the CYCIT (MAT-2005-3866). MBR acknowledges Spanish CSIC for a I3P fellowship. The Barcelona Supercomputer Center (http://www.bsc.es/) is acknowledged for computing time.

References

[1] Pendry J B 1974 *Low-Energy Electron Diffraction* (London: Academic)
[2] Van Hove M A, Weinberg V H and Chan C-M 1986 *Low-Energy Electron Diffraction* (Berlin: Springer)
[3] Van Hove M A and Tong S Y 1979 *Surface Crystallography by LEED* (Berlin: Springer)
[4] Woodruff D P and Bradshaw A M 1994 Adsorbate structure determination on surfaces using photoelectron diffraction *Rep. Prog. Phys.* **57** 1029
[5] Stout G H and Jensen L H 1989 *X-ray Structure Determination: a Practical Guide* (New York: Wiley)
[6] Robinson I K and Tweet D J 1992 Surface x-ray diffraction *Rep. Prog. Phys.* **55** 599
[7] Norman D 1986 X-ray absorption spectroscopy (EXAFS and XANES) at surfaces *J. Phys. C: Solid State Phys.* **19** 3273
[8] Woodruff D P and Delchar T A 1994 *Modern Techniques of Surface Science* (Cambridge: Cambridge University Press)
[9] Zangwill A 1998 *Physics at Surfaces* (Cambridge: Cambridge University Press)
[10] Kinniburgh C G and Pendry J B 1978 The phase problem in LEED *J. Phys. C: Solid State Phys.* **11** 2415
[11] Pendry J B 1980 Reliability factors for LEED calculations *J. Phys. C: Solid State Phys.* **13** 937
[12] Rous P J, Pendry J B, Saldin D K, Heinz K, Müller K and Bickel N 1986 Tensor LEED: a technique for high-speed surface-structure determination *Phys. Rev. Lett.* **57** 2951
[13] Pendry J B, Heinz K and Oed W 1988 Direct methods in surface crystallography *Phys. Rev. Lett.* **61** 2953
[14] Saldin D K, Chen X, Vamvakas J A, Ott M, Wedler H, Reuter K, Heinz K and de Andres P L 1997 Holographic LEED: a review of recent progress *Surf. Rev. Lett.* **4** 991
[15] Cowell P G and de Carvalho V E 1987 Unconstrained optimization in surface crystallography by LEED—preliminary results of its application to CdTe(110) *Surf. Sci.* **187** 175–93
[16] Rous P J, Van Hove M A and Somorjai G A 1990 Directed search methods for surface structure determination by LEED *Surf. Sci.* **226** 15
[17] Adams D L 2002 A simple and effective procedure for the refinement of surface structure in LEED *Surf. Sci.* **519** 157–72
[18] Döll R and Van Hove M A 1996 Global optimization in LEED structure determination using genetic algorithms *Surf. Sci.* **355** L393
[19] Kottcke M and Heinz K 1997 A new approach to automated structure optimization in LEED intensity analysis *Surf. Sci.* **376** 352
[20] Rous P J 1993 A global approach to the search problem in surface crystallography by low-energy electron diffraction *Surf. Sci.* **296** 358
[21] Nascimento V B *et al* 2001 The fast simulated annealing algorithm applied to the search problem in LEED *Surf. Sci.* **487** 15
[22] Blanco-Rey M and de Andres P L 2006 Surface diffraction structure determination from combinatorial simultaneous optimization *Surf. Sci.* **600** L91
[23] Broyden C G 1965 A class of methods for solving non-linear simultaneous equations *Math. Comput.* **19** 577
[24] Press W H, Teukolsky S A and Vetterling W T 2002 *Numerical Recipes* (Cambridge: Cambridge University Press)
[25] Kleinle G, Moritz W, Adams D L and Ertl G 1989 A novel procedure for fast surface structural-analysis based on LEED intensity data *Surf. Sci.* **219** L637–45
[26] Otten R H J M and van Ginneken L P P P 1989 *The Annealing Algorithm* (Boston, MA: Kluwer)
[27] Müller S, Kinne A, Kottcke M, Metzler R, Bayer P, Hammer L and Heinz K 1995 In-plane lattice reconstruction of Cu(100) *Phys. Rev. Lett.* **75** 2859
[28] Walter S, Blum V, Hammer L, Müller S, Heinz K and Giesen M 2000 The role of an energy-dependent inner potential in quantitative low-energy electron diffraction *Surf. Sci.* **458** 155
[29] Rundgren J 1999 Electron inelastic mean free path, electron attenuation lenght, and low-energy electron-diffraction theory *Phys. Rev.* B **59** 5106
[30] Hammer L, Meier W, Blum V and Heinz K 2002 Equilibration processes in surfaces of the binary alloy Fe–Al *J. Phys.: Condens. Matter* **14** 4145–64
[31] Kottcke M, Graupner H, Zehner D M, Hammer L and Heinz K 1996 Segregation-induced subsurface restructuring of FeAl(100) *Phys. Rev.* B **54** R5275
[32] Gauthier Y, Joly Y, Baudoing R and Rundgren J 1985 Surface-sandwich segregation on nondilute bimetallic alloys: $Pt_{50}Ni_{50}$ and $Pt_{78}Ni_{22}$ probed by low-energy electron diffraction *Phys. Rev.* B **31** 6216
[33] Blum V, Hammer L, Meier W and Heinz K 2001 Quantification of substitutional disorder and atomic vibrations by LEED—the role of parameter correlations *Surf. Sci.* **488** 219–32
[34] Kleinle G, Moritz W and Ertl G 1990 An efficient method for LEED crystallography *Surf. Sci.* **238** 119