Method for Retrieval of the Three-Dimensional Object Potential by Inversion of Dynamical Electron Scattering

Wouter Van den Broek* and Christoph T. Koch

Institut für Experimentelle Physik, Universität Ulm, Albert-Einstein-Allee 11, 89081 Ulm, Germany

(Received 20 July 2012; published 10 December 2012)

Dynamical scattering of fast electrons can be inverted by recasting the multislice algorithm as an artificial neural network, enabling the iterative retrieval of the three-dimensional object potential. This allows a nonheuristic treatment of the modulation transfer function of the CCD, partial spatial and temporal coherence, and inelastic scattering through an absorptive potential. Furthermore, prior knowledge about the atomic potential shape and the sparseness and positivity of the object can be used. The method is demonstrated on simulated bright field images recorded at 40 kV.

DOI: 10.1103/PhysRevLett.109.245502 PACS numbers: 61.05.J−, 02.30.Zz, 07.05.Mh, 61.46.−w

Because of their relatively high elastic scattering cross section, electrons provide up to four orders of magnitude more information per amount of radiation damage than x-rays [1]. However, the field of electron microscopy seems to lag behind in developments like direct methods [2] or charge flipping [3] where the crystal structure is determined directly from x-ray measurements. The reason for this, too, is the high scattering cross section since it causes dynamical scattering, which is more difficult to treat accurately. Its effects can be partially reduced by, for example, electron beam precession diffraction [4] or high angle annular dark field scanning transmission electron microscopy [5]. Such methods, however, essentially rely on losing the dynamical part of the signal, while recently it has been shown that dynamical scattering provides extra information that can actually help solve for the object structure [6,7]. In this Letter, we take a further step in that direction by providing a general framework for the inversion of dynamical electron scattering.

The multislice (MS) formalism [8–12] describes the propagation of fast electrons through an object, neglecting backscattering. The object is divided in N slices normal to the optical axis, each supporting a transmission function t. The electron wave ψ reaching the slice is multiplied with t and its propagation to the next slice is described by a convolution with the Fresnel propagator p,

$$\psi^{j+1} = p \otimes (\psi^j t^j), \qquad t^j = e^{i\sigma V^j}. \tag{1}$$

The superscript j denotes the slice number, $\otimes$ a convolution, σ the interaction constant, and V^j the projected potential within the slice [13]. In the limit of very thin slices this deceptively simple formalism becomes equivalent to a Bloch wave calculation [9] and accounts for dynamical effects such as channeling along the atom columns [14] and forbidden reflections in the diffraction pattern. The objective lens either brings the exit wave ψ^{N+1} to the image plane or to the diffraction plane. The former is described as a real space convolution with the lens function [15] and the latter as a Fourier transform [16], in both cases the result is denoted as ψ^{N+2}. Finally, the phase of the wave is lost upon recording and the measurement I equals the squared amplitude of ψ^{N+2}. Since a numerical implementation requires discretization, the quantities are sampled on a three dimensional grid: ψ_k^j, p_k, V_k^j, and t_k^j, where k denotes the position within a slice.

Equation (1) is a nonlinear system with a large amount of unknowns, rendering direct inversion impractical and thus prompting us to opt for an iterative solution by minimizing an error metric E. This is not uncommon, in related fields—like electron tomography [17], coherent diffraction imaging [18], and ptychography [19]—solutions often are arrived at iteratively. These are true object estimates since they explain the measurements if the image formation process is applied to them. In an experimental situation, one does not know if an absolute minimum of E has been reached because the true object is unknown. In case of computer simulations however, the true object is known and can be used to verify the quality of the solution and suggest experimental conditions that allow for good convergence.

The error metric is defined as

$$E = \sum_k \frac{1}{2}(I_k - J_k)^2, \tag{2}$$

where I and J are the modelled and the measured intensities, k runs over the pixels in the measurements, and the factor of $1/2$ simplifies the subsequent mathematics. The solution to the inversion problem is the potential V that minimizes E. In this Letter, it is sought by a gradient descent algorithm. Numerically calculating the required derivatives of E with respect to the elements of V would take up a prohibitive amount of time since the whole MS algorithm would need to be run for each entry of V. However, this problem is evaded by recasting the algorithm as an artificial neural network (ANN) [20].

An ANN consists of nodes representing functions, interconnected with edges carrying a certain weight. Each node

accepts as its input the output of other nodes multiplied with the weight of the interconnecting edges. This flexible scheme allows implementation of a broad variety of phenomena, including convolutions, Fourier transforms and the error metric from Eq. (2). In Fig. 1 the ANN corresponding to the current problem is given; note that the entries of the transmission function are acting as weights.

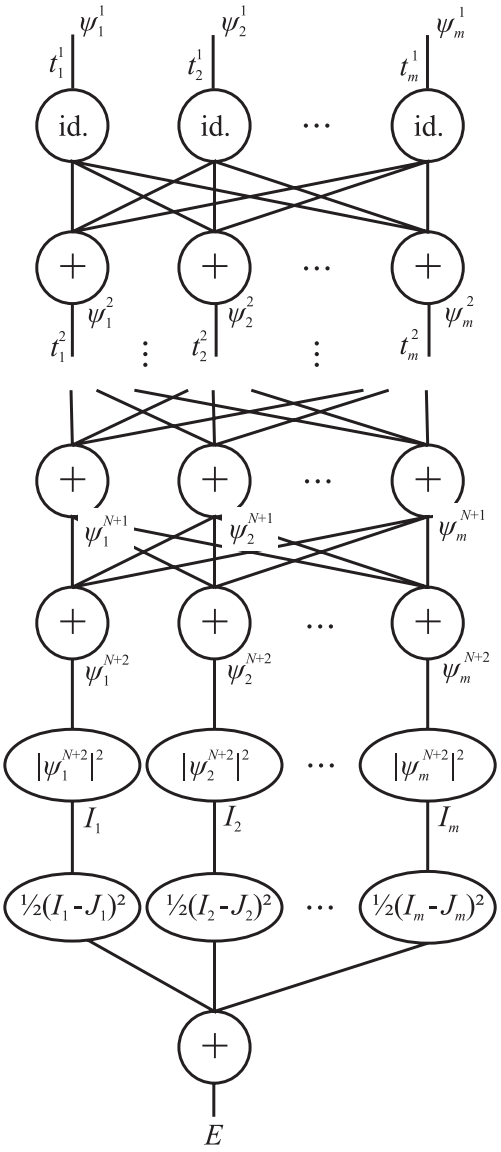

FIG. 1. The MS algorithm written as an ANN. The incoming wave, ψ^1, is multiplied with the transmission function t^1 of the first slice, then $\psi^1 t^1$ is fed into the identity function (id.). The next layer of nodes and edges realizes a real space convolution with the propagator p, resulting in ψ^2. This is repeated until the exit wave ψ^{N+1} is reached. The next layer encodes a real space convolution with the lens function if in imaging mode, or a Fourier transform if in diffraction mode, resulting in ψ^{N+2}. The last layers convert ψ^{N+2} to an intensity I and calculate the error metric E from the measurements J.

All derivatives of E with respect to the node inputs, i.e., $\partial E/\partial(\psi_k^j t_k^j)$, can be computed with one extra pass through the ANN by the well-established backpropagation algorithm [20]. The derivatives of interest are then written as

$$\frac{\partial E}{\partial V_k^j} = \frac{\partial E}{\partial \psi_k^j t_k^j}\frac{\partial \psi_k^j t_k^j}{\partial V_k^j} + \text{c.c.}, \tag{3}$$

$$= -2\Im\left(\psi_k^j t_k^j \frac{\partial E}{\partial \psi_k^j t_k^j}\right), \tag{4}$$

where c.c. denotes the complex conjugate of the previous term and $\Im$ takes the imaginary part. These derivatives are used in a steepest descent algorithm where the step length is computed with one iteration of the secant method [21,22].

The current ANN is readily extended to include more aspects of the image formation. The CCD's modulation transfer function (MTF) [23,24] and the partial spatial coherence [25] (characterized by the illumination convergence semi-angle α) are accounted for by an extra convolution after taking the intensity. The partial temporal coherence (characterized by the focal spread Δf) is modeled as an extra convolution following the lens function [26]. Small object tilts are incorporated by shifting the wave function between slices [12]. In the case of zero-loss filtered data acquisition, inelastic scattering may be treated by an absorptive potential. Therefore an imaginary part iW_k^j is added to the regular potential [27], yielding

$$\frac{\partial E}{\partial W_k^j} = -2\Re\left(\psi_k^j t_k^j \frac{\partial E}{\partial \psi_k^j t_k^j}\right), \tag{5}$$

where $\Re$ takes the real part.

In Refs. [28,29], it is shown to be highly advantageous to incorporate prior knowledge about the shape of the potential. Note that

$$V_k^j + iW_k^j = [V_0 \otimes v^j + iW_0 \otimes w^j]_k, \tag{6}$$

where V_0 and W_0 are the generalized regular and absorptive potentials of individual atoms and v and w are arrays of Dirac delta functions centred on the atom positions, with an amplitude roughly proportional to the atomic number.

TABLE I. The simulation parameters. U is the acceleration voltage, C_1 the focus value, C_3 the spherical aberration constant, d_{xy} the CCD pixel size, d_z the slice thickness, and a, c, and d define the MTF as $a\exp(-ct) + (1-a)\exp(-d^2t^2)$, with t the spatial frequency in pm^{-1}.

U	C_1	C_3	α	Δf
40 kV	−10 nm	14 μm	0.10 mrad	1.0 nm
d_{xy}	d_z	a	c	d
15 pm	0.21 nm	0.58	68 pm	98 pm

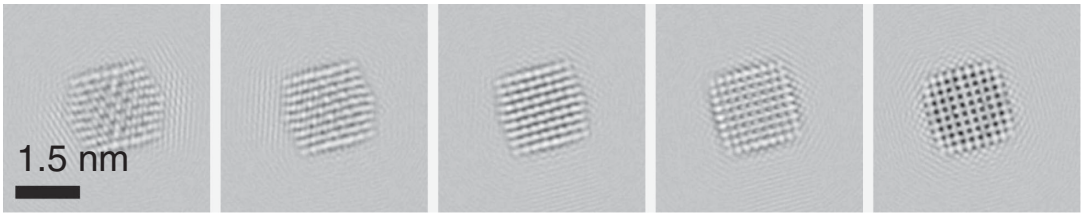

FIG. 2. Five typical measurements, the tilt around the horizontal axis of these images is 0° and around the vertical axis, from left to right, −10°, −5°, 0°, 5°, and 10°.

The error metric can then be optimized with respect to v and w instead of V and W, by using

$$\frac{\partial E}{\partial v_k^j} = -2\left[\Im\left(\frac{\partial E}{\partial V^j} \otimes V_{0,180^\circ}\right)\right]_k, \tag{7}$$

$$\frac{\partial E}{\partial w_k^j} = -2\left[\Re\left(\frac{\partial E}{\partial W^j} \otimes W_{0,180^\circ}\right)\right]_k, \tag{8}$$

where $V_{0,180^\circ}$ and $W_{0,180^\circ}$ equal V_0 and W_0 rotated over 180°. Since v and w are sparse by construction, this constraint can be imposed by adding a term $\mu\ell_1$ to the error metric E [30]. μ is a user defined constant, and the so-called ℓ_1-norm is the sum of the absolute values of v_k^j and w_k^j. ℓ_1 regularization is a well-established method in the compressed sensing community [31] to retrieve the sparsest solution to an under-determined problem. Furthermore, as in charge flipping (CF), positive and peaked solutions are favored by flipping the sign of intermediate solutions wherever they fall below a small, positive threshold, the so-called CF threshold. In practice, charge flipping has proven to successfully solve global, nonconvex optimization problems in x-ray analysis.

The method is tested on a simulated data set of 25 bright field images of a cuboctahedron composed of 309 Au atoms arranged in an fcc lattice. The simulation parameters are listed in Table I and yield a point resolution of 0.15 nm. Since, in practice, an exact focus is difficult to realize, each image is taken with a random but known defocus between −1.5 and 1.5 nm. The specimen has undergone a double tilt of −10°, −5°, 0°, 5°, or 10° around the two axes. Poisson noise equivalent to a dose of 100 electrons per pixel is applied right before the MTF. The absorptive potential W is approximated as $V/5$ [32]. In Fig. 2, five typical measurements are displayed.

Although the crystal is relatively thin, the low acceleration voltage causes significant dynamical effects. In Fig. 3, a dynamical, full MS, calculation is compared to a kinematical simulation (obtained through the weak phase object approximation).

The function $E + 0.6\ell_1$ is minimized with respect to v and w. The CF threshold is set to 1.1 times the standard deviation of the intermediate solution, as recommended in Ref. [3]. V_0 and W_0 are taken as the projected potential of Au, the factor of $1/5$ for W_0 is not used as prior knowledge.

To avoid wrap-around errors induced by the Fourier transforms, a region twice as wide as the measurements is reconstructed, but only the central part of the solution is of interest and will be displayed. Since the measurements have $320 \times 320 \times 25$ pixels and the reconstructed object $640 \times 640 \times 10$ voxels, the problem is under determined and possibly has infinitely many solutions. However, due to the object's atomicity, the physically correct solution must be sparse, and the ℓ_1-regularization does just that: restricting the solution space to sparse objects.

The object and its reconstruction are shown in Fig. 4. The reconstruction of the regular potential is faithful, all atoms are reconstructed on the correct positions. Note how a finite vertical resolution causes the intensity of the atoms to protrude in adjacent slices. In the reconstruction of the absorptive potential, however, several atoms are missing, most likely due to the weak overall contribution of the absorptive signal to the noisy measurements. For the moment, we are not interested in the absorptive signal itself [33], but since it can account for effects not included in a purely elastic MS simulation, its simultaneous reconstruction can help improving the reconstruction of the regular potential.

In this Letter, we have detailed a method to invert the dynamical scattering by recasting the multislice algorithm as an artificial neural network. This versatile approach can describe both imaging and diffraction mode and allows for a nonheuristic treatment of the modulation transfer function of the CCD, partial spatial and temporal coherence, and inelastic scattering through an absorptive potential. Furthermore, prior knowledge about the atomic potential

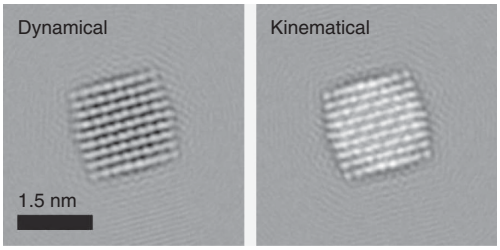

FIG. 3. Comparison between a dynamical and a kinematical simulation. The particle is imaged at zero tilt.

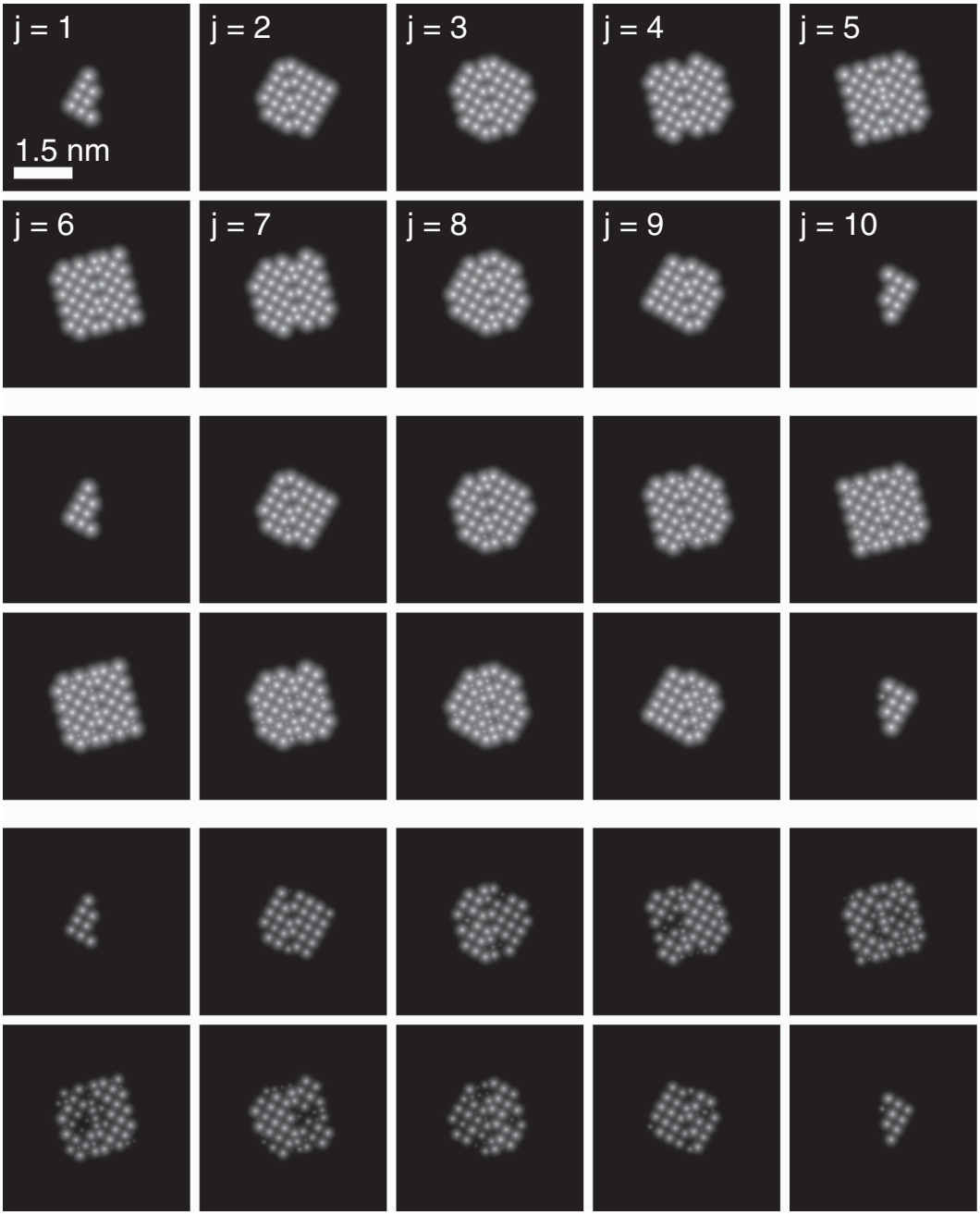

FIG. 4. *Upper two rows*: Slices with the projected potential of the Au particle; the gray scale is logarithmic. *Middle two rows*: Reconstructed projected potential, for comparison, the solution has been convolved with V_0 after intensities below the CF threshold have been set to zero. *Lower two rows*: Reconstructed absorptive potential.

shape and the sparseness and positivity of the object can be used. The lattice type or occupancy are not needed as prior knowledge. The algorithm is validated with a simulation of bright field images of a Au cuboctahedron. The potential is reconstructed faithfully and all atoms are retrieved. The scheme represented here is easily extended, for example by including a phase plate [34,35] in the back focal plane, by using a confocal setup [36] with a CCD instead of a pinhole detector or by replacing the current error metric by the negative log-likelihood function [37].

The authors acknowledge the Carl Zeiss Foundation as well as the German Research Foundation (Grant No. KO 2911/7-1).

*wouter.vandenbroek@uni-ulm.de

[1] R. Henderson, Q. Rev. Biophys. **28**, 171 (1995).
[2] H. Schenk, *An Introduction to Direct Methods: The Most Important Phase Relationships and their Application in*

Solving the Phase Problem (University College Cardiff Press, Cardiff, Wales, 1984).
[3] G. Oszlányi and A. Sütő, Acta Crystallogr. Sect. A **64**, 123 (2008).
[4] R. Vincent and P. A. Midgley, Ultramicroscopy **53**, 271 (1994).
[5] P. Hartel, H. Rose, and C. Dinges, Ultramicroscopy **63**, 93 (1996).
[6] C. T. Koch, Acta Crystallogr. Sect. A **65**, 364 (2009).
[7] C. T. Koch, Ultramicroscopy **111**, 828 (2011).
[8] J. M. Cowley and A. F. Moodie, Acta Crystallogr. **10**, 609 (1957).
[9] P. Goodman and A. F. Moodie, Acta Crystallogr. Sect. A **30**, 280 (1974).
[10] J. G. Allpress, E. A. Hewat, A. F. Moodie, and J. V. Sanders, Acta Crystallogr. Sect. A **28**, 528 (1972).
[11] D. F. Lynch and M. A. O'Keefe, Acta Crystallogr. Sect. A **28**, 536 (1972).
[12] E. J. Kirkland, *Advanced Computing in Electron Microscopy* (Springer, New York, 2010), 2nd ed.
[13] To simplify notations, σ is absorbed in V for the remainder of this work.
[14] S. Amelinckx, Acta Crystallogr. Sect. B **51**, 486 (1995).
[15] The inverse Fourier transform of the wave transfer function.
[16] M. De Graef, *Introduction to Conventional Transmission Electron Microscopy* (Cambridge University Press, Cambridge, England, 2003).
[17] P. A. Midgley and M. Weyland, Ultramicroscopy **96**, 413 (2003).
[18] J. R. Fienup, Appl. Opt. **21**, 2758 (1982).
[19] A. M. Maiden and J. M. Rodenburg, Ultramicroscopy **109**, 1256 (2009).
[20] R. Rojas, *Neural Networks—A Systematic Introduction* (Springer-Verlag, Berlin, 1996).
[21] J. Nocedal and S. J. Wright, *Numerical Optimization* (Springer, New York, 1999).
[22] R. J. Shewchuk, *An Introduction to the Conjugate Gradient Method Without the Agonizing Pain*, Carnegie Mellon University, Report No. CMU-CS-94-125, 1994, www.cs.cmu.edu/~quake-papers/painless-conjugate-gradient.pdf.
[23] I. Daberkow, K. H. Herrmann, L. Liu, and W. D. Rau, Ultramicroscopy **38**, 215 (1991).
[24] W. Van den Broek, S. Van Aert, and D. Van Dyck, Microsc. Microanal. **18**, 336 (2012).
[25] C. T. Koch, Ultramicroscopy **108**, 141 (2008).
[26] K. Ishizuka, Ultramicroscopy **5**, 55 (1980).
[27] K. Ishizuka, Ultramicroscopy **90**, 71 (2002).
[28] W. Van den Broek, S. Van Aert, and D. Van Dyck, Ultramicroscopy **109**, 1485 (2009).
[29] W. Van den Broek, S. Van Aert, and D. Van Dyck, Ultramicroscopy **110**, 548 (2010).
[30] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, IEEE Journal on Selected Topics in Signal Processing **1**, 606 (2007).
[31] J. Romberg, IEEE Signal Process. Mag. **25**, 14 (2008).
[32] D. Paganin, S. C. Mayo, T. E. Gureyev, P. R. Miller, and S. W. Wilkins, J. Microsc. **206**, 33 (2002).
[33] If we were, there would be far better ways of measuring it, e.g., by filtering out the zero-loss signal.
[34] F. Zernike, Physica (Amsterdam) **9**, 686 (1942).
[35] K. Kanaya, H. Kawakatsu, K. Ito, and H. Yotsumoto, J. Appl. Phys. **29**, 1046 (1958).
[36] S. P. Frigo, Z. H. Levine, and N. J. Zaluzec, Appl. Phys. Lett. **81**, 2112 (2002).
[37] W. M. J. Coene, A. Thust, M. Op de Beeck, and D. Van Dyck, Ultramicroscopy **64**, 109 (1996).