Application of optimization technique to noncrystalline x-ray diffraction microscopy: Guided hybrid input-output method

Chien-Chun Chen,[1] Jianwei Miao,[2] C. W. Wang,[3] and T. K. Lee[1]
[1]*Institute of Physics, Academia Sinica, Nankang, Taipei 11529, Taiwan*
[2]*Department of Physics and Astronomy and California NanoSystems Institute, University of California, Los Angeles, California 90095, USA*
[3]*Department of Physics, University of Florida, Gainesville, Florida 32611, USA*
(Received 9 March 2007; revised manuscript received 22 June 2007; published 17 August 2007)

We have developed an algorithm that combines the concept of optimization with the conventional hybrid input-output (HIO) algorithm for phase retrieval of oversampled diffraction intensities. In particular, the optimization algorithm of guiding searching direction to locate the global minimum has been implemented. Compared with HIO, this guided HIO algorithm retrieves the lost phase information from diffraction intensities with much better accuracy.

DOI: 10.1103/PhysRevB.76.064113 PACS number(s): 68.65.Hb, 42.30.Rx, 68.37.Yz, 42.30.Wb

With the advance in nanoscience and nanotechnology, x-ray diffraction microscopy, a newly developed imaging technique, is becoming more and more important in structural determination of nonperiodic micro- or nano-objects.[1–12] The idea of possibly extending the methodology of x-ray crystallography to noncrystalline objects (i.e., x-ray diffraction microscopy) was first suggested by Sayre in 1980.[13] It was not until in 1999 that the first demonstration experiment was carried out by Miao *et al.*,[1] which was based on the oversampling phasing method. When the diffraction intensities of a finite object are sampled sufficiently finer than the Nyquist frequency so that the number of correlated intensities is more than the number of unknown variables in real space, phases are usually uniquely encoded in the diffraction intensities.[14,15] It is, however, a daunting task to find the unique phases since the number of correlated intensities (i.e., nonlinear equations) is enormously large. An efficient approach is to use iteration methods such as error reduction,[16] hybrid input output (HIO),[16] difference map,[17] or the Hamiltonian approach[18] to solve the nonlinear equations. Currently, the most widely used algorithm in x-ray diffraction microscopy is Fienup's HIO approach, which was developed partially based on the Gerchberg-Saxton algorithm.[19]

Although the HIO method has contributed to the significant progress in x-ray diffraction microscopy, the method has several difficulties. One of the central difficulties is that, quite often, its results are trapped in local minima. Different trials of reconstruction could produce different solutions and none of them is accurate enough to be selected as the "right" solution. It then becomes very difficult to decide which solution provides the "correct" reconstructed image (RI). This poses a serious problem for three-dimensional (3D) tomographic reconstruction. If the two-dimensional (2D) projected image at each angle is not accurate enough, then the 3D tomographic image becomes even less reliable.

The usual way of determining whether a solution obtained is accurate or not is by examining the difference between the calculated intensities and the experimental ones. Actually, this may be the only way if we do not have other information about the object we are trying to reconstruct. In the usual HIO method, if the accuracy obtained is not good enough, we randomly adjust the electron densities with the hope of geting a smaller difference or error. This procedure is actually quite similar to the optimization problem for systems with very large numbers of variables. The minimization of the cost function, which is the difference between the calculated intensities and the experimental ones, is what one aims at.

Realization of the similarity between the image reconstruction from measured x-ray diffraction intensities and solving an optimization problem opens up new possibilities for improving the algorithms. In particular, Lee and coworkers have recently developed a guiding algorithm[20–22] to search for the optimal solution of complex systems. It differs from most of the optimization algorithms[23,24] by using the cost function as the only criterion in directing the searching pathways, and variables are permitted to sample all its allowed values. In the guiding algorithm, instead of just concentrating on designing techniques to avoid trapping in local minima and increasing the sampling efficiency, it focuses more on the guiding of the searching directions. For most optimization problems, the optimal solution has physical properties quite different from most other solutions and these properties could help guide the searching path. This idea is well known in statistical physics and it is used to find the different ground states of many models. In this paper, we employ this idea of guiding with the traditional HIO method. The algorithm, the guided HIO method, is shown to be very effective in locating the optimal solution of two model systems by reconstructing the 2D image of a noncrystalline object from x-ray diffraction intensities only. The missing data due to a beam stop are naturally resolved in this algorithm.

As what is usually done by the oversampling technique,[14] the sample is embedded in a loose support region surrounded by a no-density region. The whole region is discretized to be an array. The size of the array is determined by sampling frequency of the diffraction intensities, which is finer than the Nyquist frequency (i.e., the inverse of the size of the specimen).

The HIO method we used is essentially similar to what has been described elsewhere.[25–27] In particular, we follow

Ref. 28 for most of the iteration procedures except that we also include the missing data caused by the beam stop as the extra unknown variables to be solved. In our simulations, each iteration of the algorithm consists of the following steps. The Fourier modulus, $|F(k)|$, is combined with phases generated from the previous iteration to produce the new guessed Fourier transform $G'_n(k)$ for the next iteration:

$$G'_n(k) = \begin{cases} G_n(k) & \text{if } k \notin D \\ |F(k)/G_n(k)|G_n(k) & \text{if } k \in D, \end{cases} \quad (1)$$

where n is the iteration number, k is the scattering vector, and $k \notin D$ represents the central region blocked by the beam stop. $G_n(k)$ is the Fourier transform of the electron density $\rho_n(r)$ obtained from the previous iteration by the following equations:

$$\rho_n(r) = \begin{cases} \rho'_{n-1}(r) & \text{if } r \in S, \text{ real } [\rho'_{n-1}(r)] \geqslant 0 \\ \rho_{n-1}(r) - \beta\rho'_{n-1}(r) & \text{otherwise}, \end{cases} \quad (2)$$

where S denotes the loose support region and the parameter β is chosen to be 0.9. $\rho'_{n-1}(r)$ is the inverse Fourier transform of $G'_{n-1}(k)$. Equation (2) is designed so that, at the end of the iteration, the real part of the electron density inside the loose support should be non-negative and that outside should be zero. In this paper, we neglect absorption and consider only real electron density. Thus, the imaginary part is adjusted toward zero at the end of the iteration. The algorithm starts by assigning arbitrary phases to the measured $|F(k)|$ to obtain $G'_0(k)$. We then apply inverse Fourier transform to $G'_0(k)$ to obtain the electron density $\rho'_0(r)$. After $\rho_1(r)$ is obtained from Eq. (2), by taking $\rho_0(r)=\rho'_0(r)$, we calculate its Fourier transforms $G_1(k)$ and $G'_1(k)$ using Eq. (1). We repeat the cycle many times until there is no more improvement of the solution.

Here, we introduce a function, erf F, to determine the quality of the electron density obtained after a certain number of iterations. For example, after n iterations, we calculate $\sigma_n(r)$ as defined by the following equations:

$$\sigma_n(r) = \begin{cases} |\rho_n(r)| & \text{if } r \in S \\ 0 & \text{otherwise}. \end{cases} \quad (3)$$

The Fourier transform of $\sigma_n(r)$ is $H_n(k)$ and the erf F is defined as

$$\text{erf } F = \frac{\sum_{k \in D} \| |F(k)| - |H_n(k)| \|}{\sum_{k \in D} |F(k)|}. \quad (4)$$

Minimization of this function is what we hope to achieve by using the HIO method. However, most of the times, $\rho(r)$ of RIs are not convergent using this HIO method. The lost information in the central region due to the beam stop provides some difficulty. However, the most serious difficulty is its inability to find the global minimum when there is noise or the support is not tight enough. This is not surprising as the HIO method does not include optimization explicitly. Thus, to improve the HIO method, we need to combine it with the optimization principle. The guided algorithm[20–22] used quite successfully in some of the optimization problems is employed here.

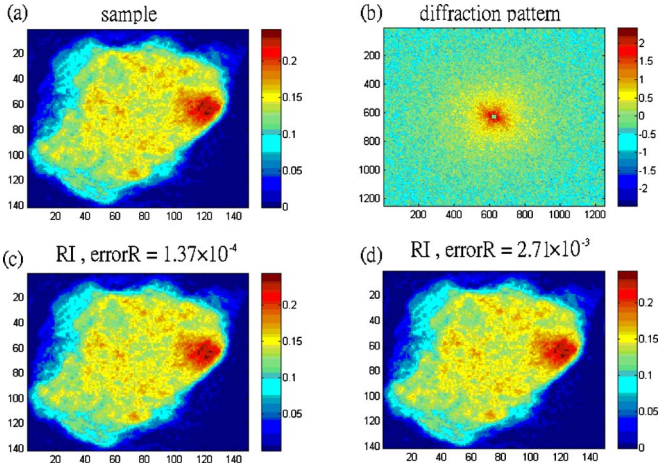

FIG. 1. (Color online) (a) The electron density of a sample model and (b) the Fourier modulus of (a) in a logarithmic scale. The central region of the diffraction pattern (23×27 pixels) is assumed to be blocked by a beam stop. The best RI by using (c) the HIO method and (d) the DM method with $\beta=1$, $\gamma_F=1.4$, and $\gamma_S=-0.8$.

In many trials through the HIO method with different initial phases during our study of a theoretical model, we found that although RIs are not quite identical to the exact solution, they all seem to have some common features. If we can keep these correct features for the future iterations, then we shall be able to approach the exact solution much more easily. A simple way to implement this idea is to select the solution with the smallest erf F to be a template to guide future iterations. Here, we use an example to show how this is executed.

Let us consider a sample model shown in Fig. 1(a). This irregularly shaped 2D sample has electron density varying between 0 and 0.25 at each pixel. We have chosen the loose support region to be 141×149 pixels and the total region including the no-density region is 1249×1249 pixels, which corresponds to a linear oversampling ratio of 8.9 and 8.4 in the x and y axes, respectively. The color bar on the right shows the magnitude of the electron density. The Fourier modulus, $|F(k)|$, is shown in Fig. 1(b) with an array of 1249×1249 pixels. The color bar on the right shows the distribution of the modulus in a logarithmic scale. The central region in the diffraction pattern with 23×27 pixels is blocked by a beam stop. These $|F(k)|$ are considered as the experimental data for our model. Using a traditional HIO method and the difference map (DM) method[17] with 10 000 iterations for 16 independent runs, we obtained the best results shown in Figs. 1(c) and 1(d). The parameters in the DM method were chosen such that erf F is smallest in the scanning region of $\beta=1$, $\gamma_F=-2$ to 2, and $\gamma_S=-2$ to 2. In addition to erf F, the quality of the RI can also be examined by the electron density difference, defined as erf $R = \Sigma_{r \in S} \| \rho^{reconstructed}(r)| - |\rho^{sample}(r) \| / \Sigma_{r \in S} |\rho^{sample}(r)|$. Based on our result, the best erf R for the RI shown in Figs. 1(c) and 1(d) is around 10^{-4}. Although the erf R of the RI through DM[17] is higher than that of HIO, the convergence factor in DM, $\epsilon \sim 10^{-5}$ is smaller.

TABLE I. erf R, erf F, and the moment of density corresponding to the HIO, DM, and GHIO methods. GHIO1 means the best results after the first generation of GHIO.

Method	erf R	erf F	Moment of density
Template	0	0	(−3.0771, −3.8422)
HIO	1.37×10^{-4}	1.41×10^{-3}	(−3.0763, −3.8430)
DM	2.71×10^{-3}	2.61×10^{-2}	(−3.0712, −3.7801)
GHIO0	3.08×10^{-2}	1.55×10^{-2}	(−3.0681, −3.7635)
GHIO1	1.36×10^{-3}	3.95×10^{-4}	(−3.0759, −3.8401)
GHIO2	4.18×10^{-6}	1.82×10^{-6}	(−3.0770, −3.8422)
GHIO3	1.99×10^{-9}	9.19×10^{-10}	(−3.0771, −3.8422)
GHIO4	1.87×10^{-12}	7.38×10^{-13}	(−3.0771, −3.8422)
GHIO5	6.08×10^{-15}	2.16×10^{-14}	(−3.0771, −3.8422)

Now, we apply the guided algorithm with the HIO method (or GHIO method) in the following steps. (i) By assuming that the electron density is real, we modify the choice of initial phases. Instead of having a complete random selection of phases, we assign k and $-k$ with the conjugate phases. Then, we start the HIO algorithm described above with 1000 iterations for each of the 16 runs with different initial phases. (ii) After the first 500 iterations, we save the RI with the smallest erf F. (iii) For the next 500 iterations, we adopt a slightly different procedure. We use the following equations:

$$\rho_n(r) = \begin{cases} \rho'_{n-1}(r) & \text{if } r \in S, \text{ real } [\rho'_{n-1}(r)] \geqslant 0 \\ \alpha\rho'_{n-1}(r) & \text{otherwise,} \end{cases} \tag{5}$$

where α is linearly changed from 1 to 0 in 500 iterations. This way, we impose a stronger condition to comply with the requirement that the density should be non-negative in the loose support region and zero outside. (iv) After we have completed all 16 runs, we select the image with the smallest erf F to be the template, $\rho^{0,template}(r)$. (v) We align the other 15 RIs to the template by comparing the density correlation function.[29] These 16 RIs are defined to be the zeroth generation. (vi) To start the next generation, we use the density of the template to modify the other 15 images by taking the square root of the product of $\rho^{0,template}(r)$ and the mth image $\rho^{0,m}(r)$,

$$\rho^{g+1,m}(r) = \sqrt{\rho^{g,template}(r) \times \rho^{g,m}(r)}, \tag{6}$$

where m goes from 1 to 16 and g stands for the gth generation ($g=0$ in this case). (vii) After we have the initial 16 images or $\rho^{1,m}(r)$ for the first generation, we start the HIO algorithm again. (viii) We repeat the above procedures to generate RIs for the higher generations.

Table I shows erf R and erf F of the best RIs obtained from the HIO, DM, and GHIO methods, indicating that the RIs of GHIO have small errors. To further quantify the error, we calculate the moment of density [i.e., $R = \Sigma_i r_i \rho(r_i) / \Sigma_j \rho(r_j)$]. As shown in Table I, both HIO and DM get somewhat different values even after 10 000 iterations. For the GHIO method, after the third generation (only 4500 iterations), the moment of density is exactly the same as that of the model. The most serious consequence of the error in the HIO and DM methods is the inability to form a reliable three-dimensional tomographic reconstruction by using the projected RIs measured at different angles, since these RIs have their moment of density (which is similar to the center of mass) at the incorrect positions. Table I also demonstrates the speed of convergence in the GHIO method. Both erf F and erf R approach zero exponentially. It should be noted that the correct result is obtained in spite of missing a substantial amount of data due to the beam stop.

To further understand the difference between the RIs obtained by GHIO, HIO, and DM, we examine the density difference, $\| |\rho^{reconstructed}(r)| - |\rho^{sample}(r)| \|$, more carefully. Most noticeable errors in the results obtained from HIO and DM lie between the boundary of the object and the loose support. Although these errors contribute little in erf R, they are not negligible in determining the moment of density. It is not surprising that HIO and DM have these errors as both methods do not have a particular way of enforcing the zero-density region inside the loose support to be zero. Furthermore, in Fourier space, the information of the edge corresponds to the higher spatial frequency. However, both HIO and DM aim to match the higher intensities first, which corresponds to the lower spatial frequency. Hence, these methods usually show fuzzy edges.

The idea of guiding algorithm is to merge the best image with each of the 16 images, so that the "favorable gene" (i.e., smaller erf F) will be passed on to the succeeding generations. There are many possible ways to do this. Our choice of using geometric average to generate the next generation actually has a special advantage. It provides a fairly efficient way of making the zero-density region inside the support approach zero. Since the density generated in the zero-density region in each iteration is always quite small, taking the square root of the product with the template will usually make it even smaller. The arithmetic average will not be as efficient.

The previous model studied has a graduate spatial variation and without any symmetry. Here, we consider another model with opposite characteristics to test the accuracy of the GHIO algorithm. Figures 2(a) and 2(b) show the electron density distribution and the Fourier modulus of a three-dimensional model with x ray incident at a specific angle where a ball has a diameter of 41 pixels and an inner cube has side length of 11 pixels. In this ball-and-cube model (BC model), the density of the cube is twice larger than that of the ball. The loose support we have chosen in this case is 41 $\times$ 41 pixels and the total region including the no-density region is 799 $\times$ 799 pixels, corresponding to the linear oversampling ratio of 19.5. The color bar on the right is the same as that of the sample model shown in Fig. 1 and the central region in the Fourier amplitude (29 $\times$ 29 pixels) is blocked by a beam stop. The reconstruction is much more difficult in the BC model because of the larger missing data area and higher symmetry. The RIs of HIO after 10 000 iterations and of GHIO after the ninth generation are shown in Figs. 2(c) and 2(d), respectively. There are several apparent differences in the RIs. The HIO result in the cube region has larger density fluctuation. The boundary of the RI obtained by GHIO is clearly sharper. In addition, at the opposite side of

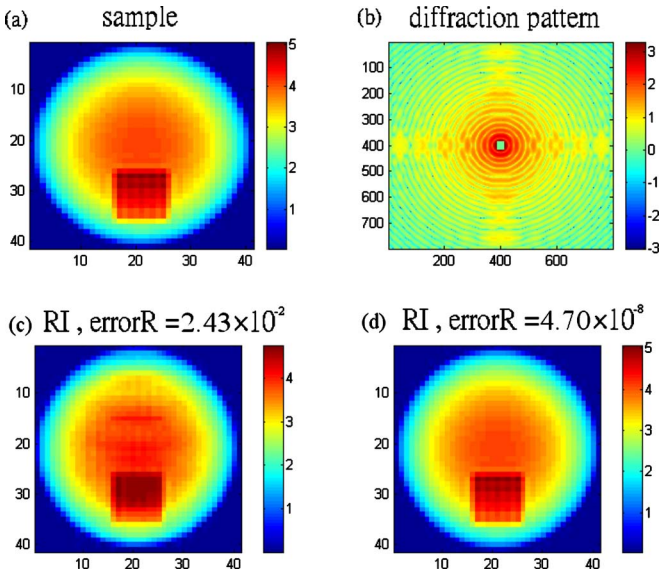

FIG. 2. (Color online) [(a) and (b)] The electron density distribution of the ball-and-cube model and the Fourier modulus in a logarithmic scale. The central region of the Fourier modulus (i.e., 29×29 pixels) is assumed to be blocked by a beam stop. The best RI using (c) the HIO method after 10 000 iterations and (d) the GHIO method after the ninth generation. The value of erf R is shown above the figure.

the cube in Fig. 2(c), there is a faint shadow of the cube. This is usually encountered in HIO as the Fourier modulus cannot distinguish the object and its inverted image.

The erf R distribution from zeroth to ninth generation is shown in Fig. 3. Each point represents a RI and there are 16 points in each generation (i.e., one column). After nine generations, we get the erf R and erf F around 4.7×10^{-8} and 3.3×10^{-9}, which are much smaller than the 2.4×10^{-2} and 8.5×10^{-3} from the best RI of the HIO method. It is worthy to point out that some RIs have converged to the best solution but some still have larger erf R. This demonstrates that the landscape of the solution space is not smooth and some RIs are trapped in local minima. This is why, in most cases, HIO cannot find the optimal solution. The advantage of the guided algorithm helps us avoid trapping.

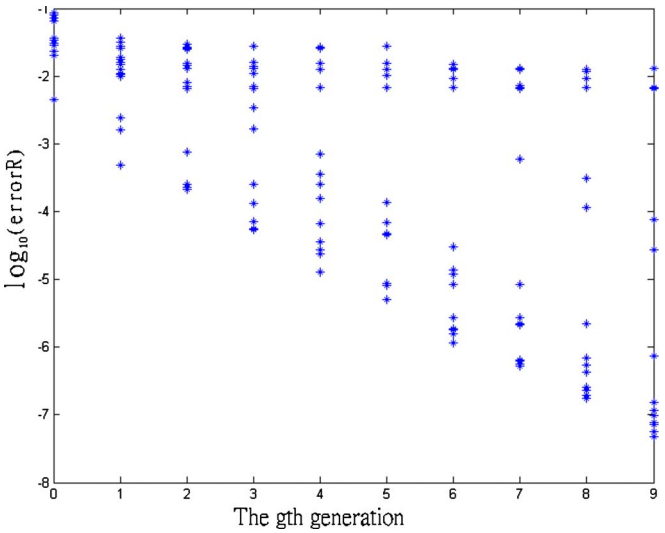

FIG. 3. (Color online) The distribution of erf R as a function of the generation for the GHIO method. The y axis is erf R in a logarithmic scale and the x axis is the gth generation. Each point represents a RI.

In summary, we have developed an algorithm, i.e., the GHIO method, to improve the image reconstruction from x-ray diffraction intensities. This method combines the traditional HIO method with the idea of optimization. We demonstrated the effectiveness of the method by reconstructing almost exactly the image of two model systems. The effect of a beam stop is naturally resolved by this method. With such an accurate algorithm, we will have more confidence in analyzing real experimental data. The possible error in the RI will be most likely due to the experimental errors instead of the algorithm. With a much more accurate RI, it is then possible to construct a more reliable three-dimensional tomogram by aligning many of these RIs measured at successive rotation angles.[30]

J.M. acknowledges the support by the U.S. Department of Energy, Office of Basic Energy Sciences (DE-FG02-06ER46276) and the U.S. National Science Foundation, Division of Materials Research (DMR-0520894).

[1] J. Miao, P. Charalambous, J. Kirz, and D. Sayre, Nature (London) **400**, 342 (1999).
[2] I. K. Robinson, I. A. Vartanyants, G. J. Williams, M. A. Pfeifer, and J. A. Pitney, Phys. Rev. Lett. **87**, 195505 (2001).
[3] J. Miao, T. Ishikawa, B. Johnson, E. H. Anderson, B. Lai, and K. O. Hodgson, Phys. Rev. Lett. **89**, 088303 (2002).
[4] S. Marchesini, H. He, H. N. Chapman, S. P. Hau-Riege, A. Noy, M. R. Howells, U. Weierstall, and J. C. H. Spence, Phys. Rev. B **68**, 140101(R) (2003).
[5] J. Miao, K. O. Hodgson, T. Ishikawa, C. A. Larabell, M. A. LeGros, and Y. Nishino, Proc. Natl. Acad. Sci. U.S.A. **100**, 110 (2003).
[6] X. Xiao and Q. Shen, Phys. Rev. B **72**, 033103 (2005).
[7] D. Shapiro *et al.*, Proc. Natl. Acad. Sci. U.S.A. **102**, 15343 (2005).
[8] J. Miao, Y. Nishino, Y. Kohmura, B. Johnson, C. Song, S. H. Risbud, and T. Ishikawa, Phys. Rev. Lett. **95**, 085503 (2005).
[9] H. N. Chapman *et al.*, J. Opt. Soc. Am. A **23**, 1179 (2006).
[10] M. A. Pfeifer, G. J. Williams, I. A. Vartanyants, R. Harder, and I. K. Robinson, Nature (London) **442**, 63 (2006).
[11] G. J. Williams, H. M. Quiney, B. B. Dhal, C. Q. Tran, K. A. Nugent, A. G. Peele, D. Paterson, and M. D. de Jonge, Phys. Rev. Lett. **97**, 025506 (2006).
[12] J. Miao, C. C. Chen, C. Song, Y. Nishino, Y. Kohmura, T. Ishikawa, D. Ramunno-Johnson, T. K. Lee, and S. H. Risbud, Phys. Rev. Lett. **97**, 215503 (2006).

[13] D. Sayre, in *Imaging Processes and Coherence in Physics*, Springer Lecture Notes in Physics Vol. 112, edited by M. Schlenker *et al.* (Springer, Berlin, 1980), pp. 229–235.
[14] J. Miao, D. Sayre, and H. N. Chapman, J. Opt. Soc. Am. A **15**, 1662 (1998).
[15] In some rare cases, the phases may not be unique, i.e., different structures correspond to the same oversampled diffraction pattern. This nonuniqueness problem can be overcome by using x-ray fields with phase curvature. [K. A. Nugent, A. G. Peele, H. N. Chapman, and A. P. Mancuso, Phys. Rev. Lett. **91**, 203902 (2003).].
[16] J. R. Fienup, Appl. Opt. **21**, 2758 (1982).
[17] V. Elser, Acta Crystallogr., Sect. A: Found. Crystallogr. **59**, 201 (2003).
[18] R. Blankenbecler, Phys. Rev. B **69**, 064108 (2004).
[19] R. W. Gerchberg and W. O. Saxton, Optik (Jena) **35**, 237 (1972).
[20] C. I. Chou and T. K. Lee, Acta Crystallogr., Sect. A: Found. Crystallogr. **58**, 42 (2002).
[21] C. I. Chou, R. S. Han, T. K. Lee, and S. P. Li, Lect. Notes Comput. Sci. **2690**, 447 (2003).
[22] C. I. Chou, R. S. Han, S. P. Li, and T. K. Lee, Phys. Rev. E **67**, 066704 (2003).
[23] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, Science **220**, 671 (1983).
[24] M. Melanie, *An Introduction to Genetic Algorithms* (MIT Press, Cambridge, MA, 1996).
[25] J. R. Fienup, Appl. Opt. **21**, 2758 (1982).
[26] Y. Nishino, J. Miao, and T. Ishikawa, Phys. Rev. B **68**, 220101(R) (2003).
[27] V. Elser, J. Opt. Soc. Am. A **20**, 40 (2003).
[28] J. Miao and D. Sayre, Acta Crystallogr., Sect. A: Found. Crystallogr. **56**, 596 (2000).
[29] G. R. Strobl and M. Schneider, J. Polym. Sci., Polym. Phys. Ed. **18**, 1343 (1980).
[30] The method of using the moment of density to align RIs obtained at different tilt angles will be presented in a follow-up paper.