Coherent x-ray diffraction imaging of silicon oxide growth

I. K. Robinson and J. L. Libbert
Department of Physics, University of Illinois, Urbana, Illinois 61801

I. A. Vartanyants
Department of Physics, University of Illinois, Urbana, Illinois 61801
and Institute of Crystallography RAS, Leninsky prospect 59, 117333 Moscow, Russia

J. A. Pitney
Department of Physics, University of Illinois, Urbana, Illinois 61801

D. M. Smilgies, D. L. Abernathy, and G. Grübel
European Synchrotron Radiation Facility, 38043 Grenoble, France
(Received 5 February 1999)

We have measured the morphology of Si samples as a function of time in air after stripping of the native oxide. For this purpose we examined the reflectivity of a coherent beam of x rays, which produces a structured diffraction pattern. We have made further progress in the development of an inversion algorithm for conversion of these patterns into one-dimensional height images. Nanometer-sized features are found to grow and evolve in waves across the surface on the time scale of minutes to hours. [S0163-1829(99)05038-9]

I. INTRODUCTION

Nanostructured layers of semiconductor material possess many interesting properties arising from the spatial confinement of their electronic structure. These properties have potential applications in the electronics industry, where lithographically derived structures are rapidly approaching the nanometer length scale. Nanometer-scale structures can arise spontaneously from instabilities in the growth,[1–3] which can be attributed to step-edge barriers that limit the transport of material from one atomic layer to the next.[4,5] There is therefore an urgent need to develop microscopic imaging techniques that are capable of probing these structures both in real time and under the ambient conditions associated with their formation.

Coherent x-ray diffraction (CXD) is one such technique. X rays interact weakly with matter, so they do not disturb the delicate balance of diffusion rates that can lead to the growth instabilities. X rays are sufficiently penetrating that changes which are internal to the material can be explored. A coherent beam is prepared from a partially coherent one by cutting with a small aperture of size d, smaller than the lateral coherence length of the x-ray source.[6,7] When a coherent beam is used, the diffraction pattern represents the *complete* Fourier transform of the entire illuminated object, without the ensemble averaging that is usually associated with diffraction. We previously showed how the morphology of the illuminated region of the surface is connected with the CXD pattern.[8–10] At each location across the coherent beam profile, (x,y), the relative *phase* of the incoming beam is modified by $q_z h(x,y)$, where q_z is the perpendicular momentum transfer and $h(x,y)$ describes the vertical height of the surface struck by the beam at position (x,y) within the profile. The amplitude of the signal seen as a function of position (angle) within the detector plane, equivalent to two parallel components of momentum transfer (q_x,q_y), is a coherent superposition of all the reflected waves within the finite illumination region, taken to extend from $-d/2$ to $+d/2$,

$$A(q_x,q_y)=\int_{-d/2}^{d/2} dx dy e^{iq_z h(x,y)} e^{iq_x x} e^{iq_y y}. \tag{1}$$

This resembles the Fourier transform of the *complex* density function, $\exp(iq_z h(x,y))$, which has unit amplitude and a height-dependent phase. Apart from this complex nature, the other main difference from the crystallographic situation is that the density function is nonperiodic. Since only the amplitude of the diffraction can be measured, while its phase information is lost, and since no suitable x-ray objective lens has yet been developed, one must resort to computational methods to produce an image.

In recent years, there has been considerable progress in computational methods for inverting these ''noncrystallographic'' diffraction patterns. First proposed by Sayre[11] in 1980 as a means of diffraction imaging, soft x-ray diffraction patterns from individual biological cells were successfully recorded in 1987 by Yun *et al.*[12] The feasibility of inversion of model data has since been demonstrated by several authors.[13–15] This and other methods of x-ray microscopy have been reviewed recently by Sayre and Chapman.[16] Sayre's original idea[11] depends on the fact that the diffraction pattern can be measured on a substantially finer scale than is attainable in diffraction from an analogous crystal, where a discrete set of Bragg peak intensities represents the entire measurement. As is well known, the direct inversion of crystallographic data is not possible without the use of difference techniques or exploitation of statistical properties of the entire data set. With sufficiently *oversampled* measurements of the continuous diffraction pattern of a nonperiodic object, it is now recognized that the Fourier transformation is overdetermined, and so can be inverted uniquely in principle. The mathematical problem is reduced to a compu-

tational one of intelligent searching for the correct solution, for which successful algorithms have been developed.[13–15] In the real situation of experimental diffraction data, containing statistical and systematic errors, the robustness of these algorithms has yet to be fully explored.

We have previously shown[8] that the grazing incidence of the x-ray beam used in the reflectivity geometry illuminates a thin stripe on the sample, and so leads to a coherent diffraction pattern which is approximately one dimensional (1D). The full expression of Eq. (1) can therefore be reduced to a 1D expression by omitting the dependence on y and q_y and referring with x and q_x to the direction along the beam's footprint alone. We also demonstrated[9,10] that these 1D patterns can be inverted using the Gerchberg-Saxton (GS) algorithm.[17] The computation time involved with inverting 1D data is relatively small, so we have since been able to explore the convergence properties of this particular algorithm. In the work we present here, we summarize these results and further explore certain assumptions underlying our model of the coherent diffraction itself.

In the experimental work reported here, we examined the growth of oxide on silicon in air after removal of the native oxide with hydrofluoric acid. A clear time evolution was seen in the CXD patterns recorded subsequently. By inverting the patterns we could observe the time dependence of the oxidation process in our calculated images. To anticipate the results, the morphology can be described as a series of waves and steps with heights starting at around 5 Å, which evolve laterally across the surface with velocities up to 0.06 μm/s.

Previous studies of the formation of the native oxide on silicon in air have been made by several authors using x-ray photoelectron (XPS) and infrared (IR) spectroscopies.[18–22] These serve to quantify the rate of growth under a variety of ambient conditions (e.g., humid vs dry) and also to investigate the chemical reactions involved. None of the previous measurements have attempted to look for inhomogeneities in the spatial distribution of oxide. The general conclusions are that the hydrofluoric-acid-treated surface is terminated by Si-H species, with its characteristic IR vibration frequency at 2100 cm^{-1}. Over a period of several hours, this becomes converted into vibration lines characteristic of $SiH(O_3)$ or $SiH(O_2)$. The rate depends somewhat on humidity, wafer orientation, and doping type.[18] The XPS measurements showed an average oxide thickness of 6 Å after 10^4 min in humid air and essentially no growth in a dry O_2/N_2 mixture.[20] The critical rate-limiting reaction[18,21] is believed to be the formation of a surface silanol,

$$Si-H+H_2O \rightarrow Si-OH+H_2\uparrow,$$

followed by reaction with a second Si-H to form a Si-O-Si bridge bond which is the main structural component of the SiO_2 structure. The Si-H terminated surface is hydrophobic, making the initial reaction very unlikely to occur, and hence very slow. Once Si-O bonds have started to form, the surface becomes hydrophilic, and so more reactive towards further attack by H_2O and/or O_2. The data of Niwano *et al.*[18] clearly show a faster-than-linear decline of the Si-H vibration intensity with time, strongly suggesting that the oxidation reaction is *autocatalytic*. We therefore would expect inhomogeneities in the appearance of oxide to arise. Our time-resolved experiment was therefore intended to probe the sizes and typical length scales of those oxide inhomogeneities.

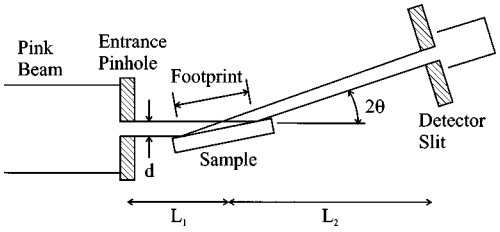

FIG. 1. Schematic drawing of the experiment installed on ID-10A at ESRF with the detector moving in the vertical plane. The setup employed a $d=5$ μm entrance pinhole at $L_1=70$ mm in front of the sample and a 20 μm detector slit at $L_2=1.25$ m behind the sample.

II. EXPERIMENTAL METHODS

Measurements were made at the Troïka beamline (ID-10A) of the European Synchrotron Radiation Facility in Grenoble. Unfiltered third harmonic undulator radiation of 8.8 keV ($\lambda=1.4$ Å) was reflected horizontally with a water-cooled polished silicon mirror onto a circular 5 μm aperture manufactured in a gold film lithographically. The electron-beam source size inside the undulator was small enough and the working distance large enough that the beam emerging from this aperture was spatially coherent in the transverse direction. The bandwidth of the U26 undulator was measured to be 3.7%, giving a longitudinal coherence length of 38 Å. With the storage ring operating at 160 mA, we obtained an incident flux of 2×10^9 photons per second passing through this pinhole onto the sample.

Different sample geometries were tried, but we found that the mechanical stability problems were reduced when the sample plane was oriented horizontally, and the detector arm was inclined in the vertical direction. The sample was mounted as shown in Fig. 1 at the center of a three-circle goniometer, at a distance of 70 mm from the entrance pinhole. For the results presented here, the incidence and exit angles were fixed at $\theta=1.28°$ to obtain a momentum transfer q_z of 0.20 $Å^{-1}$ perpendicular to the sample surface. After mounting each sample, its orientation angles on the goniometer were aligned first. Coherent diffraction data were then measured by scanning the vertical position of a 20 μm (vertical) $\times 100$ μm (horizontal) detector aperture through the reflected beam at a distance of 1.3 m from the sample. Neither the sample nor the incident beam was moved during the acquisition of the subsequent time series.

The samples were fragments of commercial Si (111) wafers, initially bearing a thin grown oxide. They were degreased and cleaned in a NH_3-H_2O_2 solution at 80 °C. After each sample was mounted and aligned, a short time series was collected to ensure that there was no drifting of its position or of the beam. The oxide in the center was then carefully removed by applying a drop of 50% hydrofluoric acid for about 30 s, draining it with a piece of filter paper, then wiping the whole face of the sample with another filter paper soaked with distilled water. Time-dependent data could be collected during the regrowth of the oxide following this preparation, typically starting within about 1 min. Each scan of typically 140 data points then took about 3 min to collect.

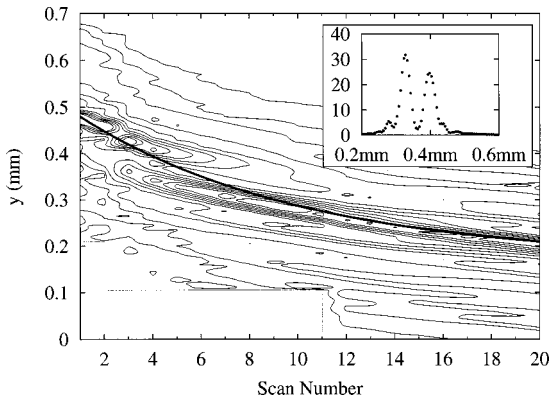

FIG. 2. Measurements of coherent x-ray diffraction from a Si(111) sample as a function of time after removal of its oxide. Scans of duration 3 min were measured in immediate succession. The time axis is therefore the serial number of the scan. The observed CXD intensity is plotted as contours with logarithmic spacings. The shaded region indicates where no measurements were made because the scan was terminated. The solid line is an exponential fit to the center-of-mass position of each scan; this movement is attributed to mechanical drift as discussed in the text. The inset shows the raw intensity (in units of thousands of counts per 0.3 s) as a function of detector-slit position during the fifth scan.

III. RESULTS

From a number of experimental runs with qualitatively the same behavior, the most interesting time series is shown in Fig. 2. Contours of intensity are plotted both as functions of time running horizontally and of the scanned position of the detector pinhole as the vertical axis, with a typical scan inset. Scans were taken in repetition, spaced 3 min apart. Because of the 3-min measurement time, some time evolution took place during the measurement itself; however, because of the relatively slow evolution of the data, regarding each scan as a ''snapshot'' does not lead to a serious distortion. For example, there is never an ambiguity concerning how the peaks in one diffraction pattern are connected to those of the next scan. It is immediately obvious that the diffraction pattern evolves more rapidly at the start of the series than at the end. This can be seen in the center of mass of the intensity distribution, superimposed on Fig. 2 as a thick line, which is a fit to an exponential function in time; it can also be seen that the relative intensities of different diffraction features generally change more dramatically at earlier times.

Using the model of the surface coherent diffraction process arising from our earlier work, which is summarized above, we attempted to invert the data, one scan at a time. Because the data (Fig. 2) are in the form of a time series of related diffraction patterns, the reliability of the data-inversion procedure can be assessed from the similarities between the inverted images.

IV. INVERSION OF CXD DATA

The most important assumption in our model[9] of coherent diffraction is the definition of the incident radiation that provides the illumination of the sample. We assume that the sample is in the near field of the entrance pinhole (Fig. 1) ($L_1 \leqslant d^2/\lambda \simeq 180$ mm), where λ is the wavelength of radiation and L_1 is a distance between the pinhole and the sample. The detector is then assumed to be in the far field of the sample ($L_2 \gg d^2/\lambda$), where L_2 is the distance from the sample to detector. Under these assumptions, the 1D version of Eq. (1) becomes

$$A(q_x) = \int dx \rho_{eff}(x) e^{iq_x x}, \tag{2}$$

where

$$\rho_{eff}(x) = B(x) e^{iq_z h(x)} \tag{3}$$

is a complex density function. In Eqs. (2) and (3) we have used the definitions $q_z = 2k \sin\theta$ and $q_x = k\eta/L_2$, where $k = 2\pi/\lambda$ is the wave vector, 2θ is the scattering angle, the coordinate x is taken across the beam at the position of the sample, and η is the position of the detector slit. Due to the grazing incident angle conditions the illuminated area is substantially elongated along the beam direction. We therefore define a coordinate $x' = x/\sin\theta$ along the sample surface itself (see Fig. 1). This fiftyfold elongation between x and x' also allows us to make the approximation that isotropic features in the height of the surface will predominately affect the diffraction in the direction of elongation. In other words we can assume a one-dimensional rather than two-dimensional scattering process.

A reciprocal-space view of this elongation effect was discussed in an earlier paper we published on coherent x-ray diffraction from a GaAs/AlAs multilayer sample.[23] In that case also, one-dimensional speckle was obtained using a symmetric beam. The reciprocal-space direction scanned by the detector is also inclined at an angle θ with respect to the surface plane, so the component of momentum transfer parallel to the surface becomes *contracted* by $q_x' = q_x \sin\theta$. The two factors of $\sin\theta$ then cancel each other when the Fourier transform is evaluated in the primed coordinates.[23]

The function $B(x)$ in Eq. (2) is called the illumination function because it accounts for the propagation of the wave front of the incoming coherent beam from the entrance aperture to the sample. In the simplest approximation $B(x)$ would be just a box function or spatial cutoff as in Eq. (1). More generally, it will account for Fresnel diffraction by the aperture defining the incoming beam, when it becomes the complex function[24]

$$B(x) = |B(x)| e^{i\phi_F(x)}. \tag{4}$$

We have shown in our previous paper[9] examples of the illumination function $|B(x)|$ for the case of ideal slits with sharp edges. Then it depends on the slit size d, the distance from the slits to the sample L_1, and the wavelength of radiation in the form of a Fresnel integral.[24] We have estimated the illumination function $B(x)$ for the conditions used in this experiment (Fig. 1) which gives a strong maximum in the center and has pronounced side fringes, while $\phi_F(x)$ is relatively constant over the central part of the illumination. However, the function $B(x)$ also accounts for any internal phase structure in the beam due to the conditioning optics. In addition, it can partially compensate for such effects as nonsharpness of the slit edges, partial coherence of the beam and convolution with the detector slits. While it can be estimated

from the geometry of the entrance aperture, it must strictly be regarded as an experimental unknown. It is desirable to know the ''exact'' shape of the illumination function because it is centrally incorporated in the reconstruction procedure. In principle we should be able to optimize the shape of $B(x)$ by investigating its effect on the convergence of the reconstruction algorithm or on the ultimate agreement with data. However, we would not expect the resulting phase structures to depend sensitively on the exact details of $B(x)$.

The amplitude of $B(x)$ defines the magnitude of the illumination of the sample and the phase $\phi_F(x)$ gives an additional phase shift. This phase combines with that of the sample to give a final expression for the speckle amplitude (2) seen at the detector

$$A(q_x)=\int dx|B(x)|e^{i\Phi(x)}e^{iq_x x}, \tag{5}$$

where the total phase $\Phi(x)$ is equal to the sum of the phases

$$\Phi(x)=\phi_F(x)+q_z h(x). \tag{6}$$

Equation (5) is precisely the Fourier transform of the complex quantity $\rho_{eff}(x)=|B(x)|e^{i\Phi(x)}$ with the amplitude $|B(x)|$ and the phase $\Phi(x)$. What is measured corresponds to $|A(q_x)|^2$. It is therefore the combined density function $\rho_{eff}(x)$ that we seek to derive from the data analysis procedure. We have some prior knowledge of $|B(x)|$ and $\phi_F(x)$ from the sample geometry, while $h(x)$ is the image of the sample morphology we wish to know.

There are trivial ambiguities in any situation where only the amplitude of a Fourier transform is measured. In our case, the following three quantities are indistinguishable: $\rho_{eff}(x)$, $\rho'_{eff}(x)=|B(x)|e^{i\Phi(x)+\Phi_0}$, and $\rho''_{eff}(x)=\rho^*_{eff}(-x)$, where Φ_0 is an arbitrary real constant and $*$ denotes complex conjugation. The quantity $\rho''_{eff}(x)$ is called the ''dual'' solution. In the following, we study only nontrivial characteristics of the phase problem for $\rho_{eff}(x)$.

Up to now we have been considering the static speckle pattern. We now consider what happens when the height function becomes time dependent. It follows from the superposition properties of the Fourier transform that a height function with a component that is *linear* in position just leads to the same diffraction amplitude as the height function without the linear component, but with an origin shift in reciprocal space. Such a time drift of the centering of the data was indeed observed in the experiment and is shown in Fig. 2 by a solid line. This behavior can easily arise from mechanical drift in the experiment. We are interested instead in higher-order variations of the height function, so we systematically remove the linear term by recentering the data. A similar ''flattening'' procedure is used in scanning probe techniques (scanning tunneling microscopy or atomic force microscopy) for measuring surfaces. For this reason, we have centered all experimental scans to the position of the numerical center of mass of the data.

For the reconstruction of the phase $\Phi(x)$ (6) from the measured speckle pattern intensity we use the algorithm proposed by Gerchberg and Saxton,[17] which we have employed before.[9] The idea of this algorithm is to iterate back and forth between real and reciprocal space using the direct [Eq. (5)] and inverse Fourier transforms. After each step of the GS algorithm the amplitudes are updated. In real space, the amplitude is changed to the calculated value of $|B(x)|$; in reciprocal space, $|A(q)|$ is changed to the observed value, $\sqrt{I_{obs}(q)}$. In both cases, the obtained value of the phase is maintained. After a number of iterations, the algorithm converges to a value of the phase that satisfies our constraint in both real and reciprocal space. The detailed discussion of the algorithm and results of its testing on simulated as well as real experimental data were given in our previous publications.[9,10] Here we will concentrate on the results of the reconstruction of the phase in the time-dependent series of Fig. 2 by application of the GS algorithm.

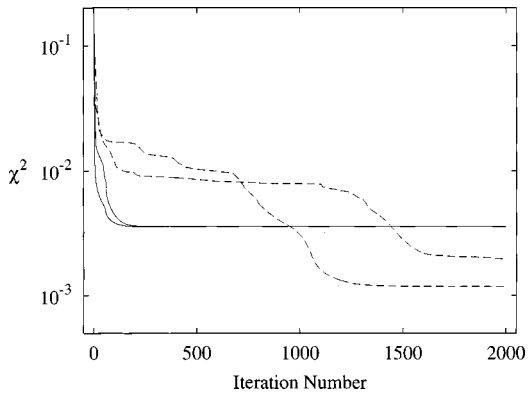

FIG. 3. Convergence trajectories of the χ^2 defined in Eq. (7) as a function of number of iterations of the GS algorithm. The two dashed curves are representiative of the conditions used in Fig. 4(a), while the two solid curves correspond to Fig. 4(b).

Direct (5) and inverse Fourier integrals in the iteration algorithm were calculated using the discrete fast Fourier transform (FFT). This requires that the continuous functions of x and q be replaced by arrays x_i and q_j, sampled at uniformly spaced discrete points in the real and reciprocal spaces. In our calculations, typically 140 data points were zero padded to $N=1024$. This gave sufficient definition of real space features that they appeared to be continuous functions.

While using iterative algorithms, it is desirable that the obtained result should not depend on the choice of the starting phases and should always converge to the same solution reproducibly. The number of iterations necessary for the procedure to converge can also depend on the choice of the starting phases. Examples of the convergence of the GS algorithm are shown in Fig. 3. Our experience is that the convergence is usually very fast for the first 50 iterations, but then, depending on the choice of the starting phases, it can either continue to converge or can develop long stagnation periods, as shown in Fig. 3. This stagnation behavior of the GS algorithm is well known and there have been proposed different ways to overcome it.[25]

Taking into account that the computation time for such a fitting procedure based on FFT's in one dimension is relatively small, we developed the following strategy. A set of random starting phases was generated and the GS algorithm was applied to it with several thousands of iterations. A second set of random phases was generated and the process repeated. To ascertain the reproducibility we made 100 fits each with a different choice of random phases. This represents a rather crude way to avoid the problem of stagnation

at the expense of computation time. At the end of every iteration the quantity

$$\chi^2 = \sum_j \{\sqrt{I_{obs}(q_j)} - |A(q_j)|\}^2 / \sum_j I_{obs}(q_j) \qquad (7)$$

was calculated and the results of all fits were sorted in order of ascending χ^2 to keep the ten ''best'' fits out of 100. The χ^2 in Eq. (7) represents the *fidelity* of the fitting procedure. Because of the uncertainties of choosing random starting phases, a separate assessment of *reproducibility* was made,

$$\Delta\Phi = \max_{\{1 \leq k,l \leq 10\}} \{|\Phi_k(x_i) - \Phi_l(x_i)|\}|_{x_i = \pm d/2}. \qquad (8)$$

This represents the maximum deviation of the phases over the width of the illuminated area for the ten best fits with the lowest χ^2 value. It is an empirical quantity that may be related to the mathematical ''uniqueness'' of the reconstructed phase. Of course, the most desirable situation would be simultaneously to obtain low values of both χ^2 and $\Delta\Phi$. In optimizing the performance of the algorithm, we considered different forms of the illumination function, $B(x)$, while tracking χ^2 and $\Delta\Phi$. We note that these two parameters are essentially complimentary to each other, so we might anticipate opposing trends as we systematically vary the illumination function.

The problem of uniqueness of the GS algorithm has been widely discussed and has been proven for the class of analytical functions.[26] In the numerical procedure we are looking for a solution of a set of nonlinear equations

$$|A(q_j)| = \left| \sum_{i=0}^{N-1} \rho_{eff}(x_i) \exp(i q_j x_i) \right|,$$

$$|B(x_i)| = \left| \frac{1}{2\pi N} \sum_{j=0}^{N-1} A(q_j) \exp(-i q_j x_i) \right|, \qquad (9)$$

with known amplitudes $|A(q_j)|$ and $|B(x_i)|$ and unknown phase $\Phi(x_i)$. The phase of the Fourier amplitude $A(q_j)$ is also an unknown parameter, although we are not interested in that explicitly. The mathematical problem is fully determined by these equations for which we hope to find a unique solution. In the simplest 1D case for FFT's of size N, Eq. (9) provides $2N$ equations (N in real space and N in reciprocal space) for $2N$ unknown phases. However, the real situation is that there are fewer than N data points and fewer than N nonzero points for the real space amplitude, $B(x)$. Errors in the data would technically count as inequalities, rather than exact equations, so the situation of assessing whether or not the problem is overdetermined becomes more complicated.

We have selected one of the typical scans from the time series (scan 5 of Fig. 2, also shown in the inset) for a detailed investigation of the problem of convergence, agreement, and reproducibility. First, we made simulations with the illumination filter function expected for the experiment, provided by the Fresnel integrals with $d = 5$ μm and $L_1 = 100$ mm. The results were quite encouraging and gave good agreement for the amplitude functions in both the real and reciprocal domains as shown in Fig. 4(a). Good agreement is seen even for the low intensity tails. However, the poor reproducibility of the phase among the ten best fits is clearly seen in Fig.

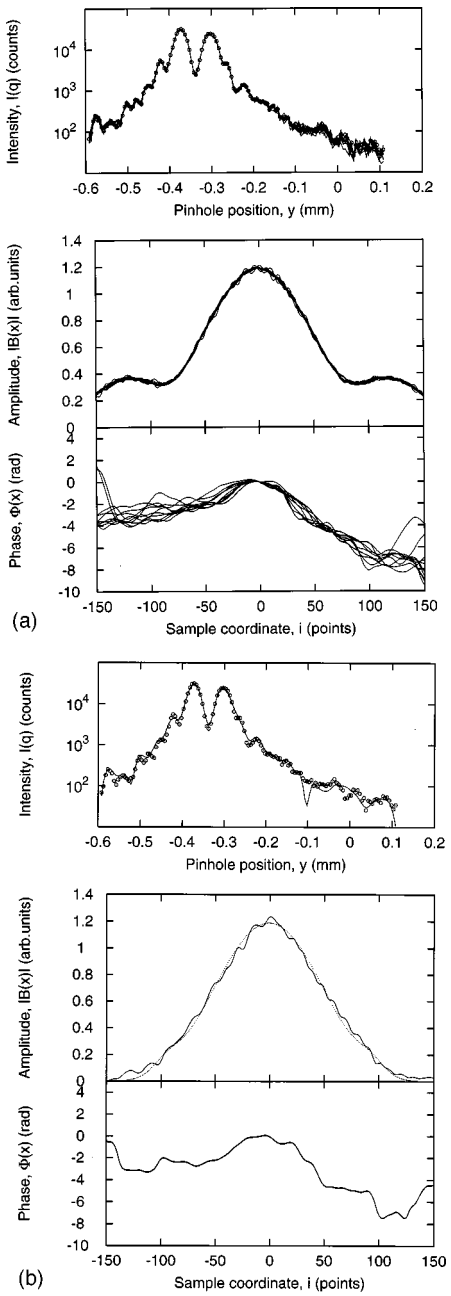

FIG. 4. Results of fitting the data of scan 5 of Fig. 2 using different illumination functions. The GS algorithm has been used with 100 attempts of 2000 iterations each, as described in the text, with FFT's of size $N = 1024$. The top panels of each part show the reciprocal-space data with the ten best fits out of 100 superimposed. The middle panels denote the real-space amplitude fits to the illumination function. The bottom panels show the ten best real-space phase functions derived. (a) top panels, using the ''open'' illumination function given by the theoretical Fresnel form and the approximate parameter values used in the experiment. (b) bottom panels, using the cutoff illumination function, as in (a) except smoothly truncated on both sides.

4(a). Note that, because of the arbitrary choice of the phase offset, Φ_0, all phases have been set to zero at the central point of the plot.

In our investigation of the reproducibility of repeated fitting of the same data, we noticed a very interesting general characteristic. When the $B(x)$ function was confined to be

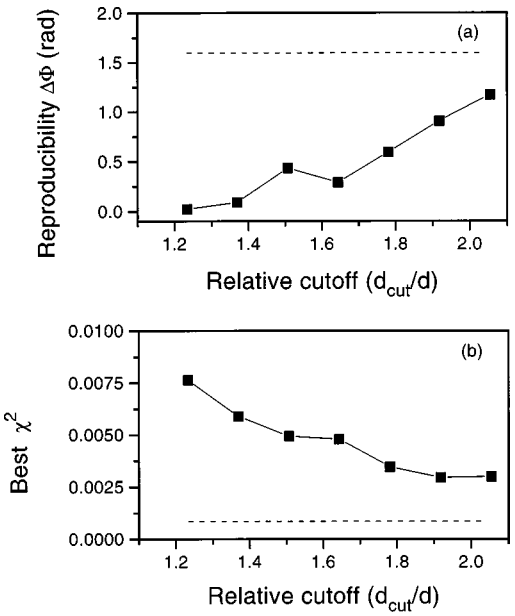

FIG. 5. Variation of the reproducibility, $\Delta\Phi$ (top) and ''best'' χ^2 (bottom) as a function of the cutoff parameter d_{cut}. Each point represents 100 GS fits to the same data with different random starting values for the phases. The assumed value of the width of the incident beam, corresponding to the pinhole size *d*, was fixed at 73 points and the FFT's were all of size $N=1024$.

different from zero only over narrow domain in x_i, the spectrum of the resulting χ^2 values became discrete, with *identical* values of χ^2 occurring multiple times each. For each discrete χ^2, the solutions for the phase, allowing for duals, were also identical over the central illuminated region. The lowest of these χ^2 values was naturally taken to correspond to the ''correct'' solution. We assume this result arose because the inversion problem of Eq. (9) becomes overdetermined when the domain of x_i is sufficiently small: there are more data points, each representing a simultaneous nonlinear equation in Eq. (9), than variables, each being the phase of one point in the domain of x_i.

We illustrate this by comparing fits to the same data with the confined illumination function in Fig. 4(b): the domain of illumination has been smoothly suppressed beyond a separation, d_{cut}, wider than the intrinsic width, *d*, by a factor of 1.3. This time the best ten phase solutions, corrected for duals, are identical, but the final χ^2 value is somewhat higher. The convergence behavior for the two cases is compared in Fig. 3. The solid curves are two of the ten best χ^2 trajectories for the cutoff version of the illumination function, for different starting phases. Though not identical at the start, they both fall quickly to the same final χ^2 value. Conversely, the two dashed curves in Fig. 3 are examples of two of the ten best χ^2 convergences for the full, wide-open $B(x)$ function used in Fig. 4(a); notable stagnation periods occur, but both examples do ultimately reach better χ^2 values.

This property was analyzed systematically for these experimental data as a function of d_{cut}/d in Fig. 5. The χ^2 and $\Delta\Phi$ values can clearly be seen to trade off. A large illumination region leads to good fidelity, represented by small values of χ^2, but bad reproducibility, meaning large $\Delta\Phi$. A small illumination region does the opposite. This is consistent with the idea mentioned above that the same solution will occur most frequently, i.e., $\Delta\Phi=0$, when the problem is sufficiently overdetermined. However, the smallest χ^2 will occur when there are the greatest number of degrees of freedom, regardless of whether the solutions are uniquely defined.

In the phase retrieval problem of nonperiodic objects lies the central theoretical concept of *support* for the real space domain,[27–29] which for a 3D real object is simply the shape of the boundary of the object. For our 1D case, the support can be defined as the illuminated length of the sample. By decreasing the size of the support, we are effectively decreasing the number of variables in the set of nonlinear equations of Eq. (9) for unknown phases in the real-space domain. We know from the literature that the uniqueness of the solution to this general problem essentially depends upon the spatial dimension, on the type of the object (real, phase, or complex) and upon the available reciprocal-space information. The work of Miao, Sayre, and Chapman[29] defines a ratio σ_{MSC} between the total array size, *N*, and the number of points in the support. In their tests with simulated data, they saw convergence to the correct solution whenever $\sigma_{MSC}>2.5$ approximately, in both two and three dimensions.

We have carried out further tests to check the dependence on the support, noting that this corresponds precisely to the relative width of our illumination function, given by d_{cut}. For our case, $\sigma_{MSC}=N/N_x^{cut}$, where N_x^{cut} is the number of nonzero $B(x_i)$ sampling points within the d_{cut} window. The value of $d=5$ μm used for the tests corresponds to 150 sampling points, while $N=1024$. Thus our critical cutoff at which irreproducibility starts to occur in Fig. 5 corresponds to $\sigma_{MSC}\simeq 4$. When we changed to $N=512$ with the same data, we found this led to the same σ_{MSC}, but when we interpolated the experimental data to artificially generate 280 data points instead of 140, we found that N_x^{cut} had to be reduced by the same factor so that $\sigma_{MSC}\simeq 8$ gave the threshold behavior. We therefore conclude that the number of (nonzero) data points is also relevant to the argument. In further tests, we found that cutting down the total number of data points directly led to better reproducibility for the same N_x^{cut}.

Once we had found the maximum illuminated length, $d_{cut}=6.6$ μm, that permitted reproducible fitting, we varied the internal shape of $B(x)$ using the size of the pinhole, *d*, and the sample-pinhole distance, L_1, as parameters. We found $d=4.8$ μm and $L_1=100$ mm gave the smallest χ^2, so describing the best-fitting functional form for $B(x)$. These parameters were then used for the fitting of the whole time series to generate the real-space *images* in Fig. 6. The performance of the GS algorithm did vary somewhat across the series with five scans in the middle having slightly worse reproducibility. An additional problem was caused by the missing experimental data in some of the scans, shaded in Fig. 2. Here we simply appended the missing tail from the next (full) scan and reconstructed the modified data.

The curves plotted in Fig. 6 have been screened for ''duals'' and reversed where necessary to align with their neighbors. Because of similarities from one scan to the next, there was never any ambiguity in the choice of duals. The distortion of the results shown in Fig. 6 due to the additional Fresnel phase, $\phi_F(x)$, was found to be negligible.

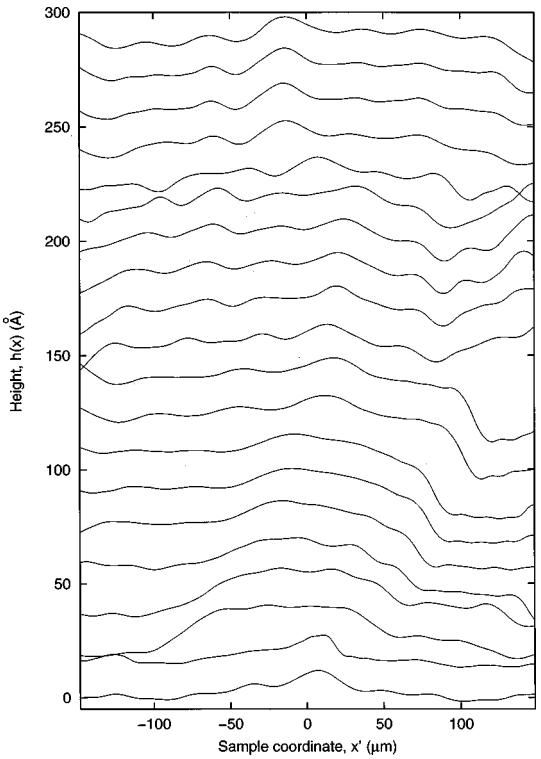

FIG. 6. Final height profiles resulting in the reconstruction of the whole time series. Each profile was reconstructed independently using random starting phases, following the procedure described in the text. Time runs from the bottom of the figure to the top and each profile has an arbitrary vertical offset. The fifth curve up is the same fit as shown in Fig. 4; however, both axes have been converted to physical units.

V. DISCUSSION

The phase structures plotted in Fig. 6 have been converted into units of height using the fixed value of perpendicular momentum transfer, 0.2 Å^{-1}, applicable to all the measurements. The horizontal axis has also been converted into units of distance across the sample, x', as defined above. As such, they represent a series of images of the sample as a function of time after the initiation of oxidation, with time running from the bottom of the figure to the top.

The first image (bottom of Fig. 6) is fairly featureless, with a single bump of 7 Å in height located near the center. Its lateral extent is about 25 μm and its edges are rounded. The total data range that was measured was ± 0.35 mm motion of the pinhole, corresponding to $q_x = \pm 1.21 \times 10^{-3}$ Å^{-1} range of lateral momentum transfer. This should provide an effective lateral resolution of 0.26 μm in x, or 12 μm in x', beyond which features should not be distinguishable. This initial bump is therefore slightly broader than our resolution limit. In the following scan, the bump has become taller with a slightly sharper edge which is closer to the resolution limit.

In subsequent scans (3)–(10), the initial bump spreads laterally across the sample at a significant rate, with its right-hand edge moving at 0.06 μm/s and becoming taller as it moves, reaching 20 Å by the time it starts to disappear from view on the extreme right-hand side. Again the width of the edge remains constant at around 15 μm, limited by the lateral resolution. We presume that the material to the right of this migrating edge remains bare silicon (hydrogen terminated), so we can use the trend to make a crude estimate of the initial growth rate: 20 Å of oxide has covered half of the surface within view in 30 min, implying 0.3 Å/min. This in rough agreement with the oxidation rates observed spectroscopically.[20,21]

The steplike feature discussed above is the most apparent, but there are smaller waves that also slide across the surface in both directions at speeds nearer to 0.01 μm/s. The fastest rate of 0.06 μm/s corresponds to a growth front consuming 100 unit cells of Si per second. While the growth rates are reasonable, it is remarkable that nucleation of oxide is apparently such a rare event that it occurred just once within the 5 μm$\times$230 μm field of view during the course of our experiment. It is possible that this is not just a coincidence but due to some beam-induced effects as ozone creation, which would of course be greatest in the center of the beam.

In conclusion, we have developed the inversion techniques for coherent x-ray diffraction to such an extent that we have imaged credible changes on silicon due to the initial stages of oxidation. The oxidation we observed proceeds in distinct waves, a few monolayers in height that migrate across the sample at rates of up to 0.06 μm/s. We believe the GS inversion method to be reliable because of the smooth connection between the members of the time series of related images that emerges, even though random starting phases were assumed. In order to ensure reproducible results it was found to be necessary to restrict the assumed illumination of the sample within a window about 1.3 times the size of the entrance aperture.

ACKNOWLEDGMENTS

We thank the staff of ESRF, particularly P. Feder, H. Gleyzolle, and H. Müller for their enormous help in carrying out the experiments described. The work was supported by the U.S. Department of Energy under Contract No. DEFG02-96ER45439 to the Materials Research Laboratory of the University of Illinois.

[1] M.D. Johnson, C. Orme, A.W. Hunt, D. Graff, J. Sudijono, L.M. Sander, and B.G. Orr, Phys. Rev. Lett. **72**, 116 (1994).

[2] J.E.Van Nostrand, S.Jay Chey, M.A. Hasan, D.G. Cahill, and J.E. Greene, Phys. Rev. Lett. **74**, 1127 (1995).

[3] W. Theis, N.C. Bartelt, and R.M. Tromp, Phys. Rev. Lett. **75**, 3328 (1996).

[4] G. Ehrlich and F.G. Hudda, J. Chem. Phys. **44**, 1039 (1966).

[5] S.C. Wang and G. Ehrlich, Phys. Rev. Lett. **67**, 2509 (1991).

[6] M. Sutton, S.G.J. Mochrie, T. Greytak, S.E. Nagler, L.E. Berman, G.A. Held, and G.B. Stephenson, Nature (London) **352**, 608 (1991).

[7] G. Grübel and D. L. Abernathy, Proc. SPIE **3154** 103 (1997).

[8] J.L. Libbert, R. Pindak, S.B. Dierker, and I.K. Robinson, Phys. Rev. B **56**, 6454 (1997).

[9] I.A. Vartanyants, J.A. Pitney, J.L. Libbert, and I.K. Robinson, Phys. Rev. B **55**, 13 193 (1997).

[10] I.K. Robinson, J.A. Pitney, J.L. Libbert, and I.A. Vartanyants, Physica B **248**, 387 (1998).
[11] D. Sayre, in *Imaging Processes and Coherence in Physics*, edited by M. Schlenker, M. Fink, J. P. Goedgebuer, C. Malgrange, J. C. Viénot, and R. H. Wade, Lecture Notes in Physics Vol. 112 (Springer-Verlag, Berlin, 1980), p. 229.
[12] W.B. Yun, J. Kirz, and D. Sayre, Acta Crystallogr., Sect. A: Found. Crystallogr. **43**, 131 (1987).
[13] R.P. Millane and W.J. Stroud, J. Opt. Soc. Am. A **14**, 568 (1997).
[14] J.R. Fienup and C.C. Wackerman, J. Opt. Soc. Am. A **3**, 1897 (1986).
[15] C. Jacobsen, M. Howells, J. Kirz, and S. Rothman, J. Opt. Soc. Am. A **7**, 1847 (1990).
[16] D. Sayre and H.N. Chapman, Acta Crystallogr., Sect. A: Found. Crystallogr. **51**, 237 (1995).
[17] R.W. Gerchberg and W.O. Saxton, Optik (Stuttgart) **35**, 237 (1972).
[18] M. Niwano, J. Kageyama, K. Kiruta, I. Takahashi, and N. Miyamoto, J. Appl. Phys. **76**, 2157 (1994).
[19] H. Ogawa, K. Ishikawa, C. Inomata, and S. Fujimura, J. Appl. Phys. **79**, 472 (1996).
[20] M. Morita, T. Ohmi, E. Hasegawa, M. Kawakami, and K. Suma, Appl. Phys. Lett. **29**, 562 (1989).
[21] M. Morita, T. Ohmi, E. Hasegawa, M. Kawakami, and M. Ohwada, J. Appl. Phys. **68**, 1272 (1990).
[22] P. Jakob and Y.J. Chabal, Chem. Phys. Lett. **95**, 2897 (1991).
[23] I.K. Robinson, R. Pindak, R.M. Fleming, S.B. Dierker, K. Ploog, G. Grübel, D.L. Abernathy, and J. Als-Nielsen, Phys. Rev. B **52**, 9917 (1995).
[24] M. Born and E. Wolf, *Principles of Optics*, 4th ed. (Pergamon Press, London, 1970).
[25] J.R. Fienup, Appl. Opt. **21**, 2758 (1982).
[26] H.A. Ferwerda, in *Inverse Source Problems in Optics*, edited by H.P. Baltes, Topics in Current Physics Vol. 9 (Springer-Verlag, Berlin, 1978).
[27] R.P. Millane, J. Opt. Soc. Am. A **7**, 394 (1990).
[28] R.P. Millane, J. Opt. Soc. Am. A **10**, 1037 (1993).
[29] J. Miao, D. Sayre, and H.N. Chapman, J. Opt. Soc. Am. A **15**, 1662 (1998).