Reconstruction of surface morphology from coherent x-ray reflectivity

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

J. A. Pitney, J. L. Libbert, and I. K. Robinson
Department of Physics, University of Illinois, Urbana, Illinois 61801
(Received 23 December 1996)

The observation of coherent diffraction effects in recent measurements of x-ray reflectivity from Si surfaces is explained with the development of a simple kinematical theory. Some properties of the derived formalism are explored. We apply an algorithm developed by Gerchberg and Saxton and demonstrate its application to the reconstruction of the surface morphology from its coherent diffraction pattern. Initial testing with experimental data shows the method to be effective. [S0163-1829(97)01320-9]

I. INTRODUCTION

Coherent x-ray diffraction (CXD) is a new experimental technique that has emerged from the development of brilliant synchrotron-radiation sources of x rays. It has great potential for studying structural fluctuations in all kinds of condensed-matter systems, and a few pioneering experiments have demonstrated this.[1–6] The relaxation times of critical fluctuations were observed near to the phase transition of Fe_3Al.[3] The Brownian motion of colloidal gold particles[5] and, more recently, colloidal palladium[6] have been studied to determine their diffusion time constants as functions of temperature and momentum transfer.

Static CXD patterns from $GaAs_xAlAs_{1-x}$ multilayer samples have been examined for the purpose of explaining the diffraction mechanism.[4] A *phase block* model was introduced in which the sample was modeled as uniform regions of differing relative scattering phase that gave rise to interference in the CXD pattern. The finite number of phases could be adjusted as least-squares fit parameters to produce a faithful representation of the observed intensity. The diffraction was observed to be one dimensional in nature; this was justified on the grounds that the beam footprint on the sample was highly elongated, so that different points of the sample across its narrow direction could all be assumed to have the same phase. Despite the limitations of the arbitrary discrete assumptions in the phase block model, reasonable fits to the data resulted.[4] In this paper, we develop a more general model in which the discreteness is reduced to an arbitrarily low level.

We have recently demonstrated that CXD can also be used to study surface morphology in the specular reflectivity geometry.[7] While incoherent x-ray reflectivity from a rough surface would give rise to a superposition of pure specular and diffuse components,[8] the equivalent CXD experiment yields a *speckle pattern* that intermixes both of these components: the specular component becomes broadened by the finite size of the illuminated sample area while the diffuse component adopts a finely speckled structure. The experimental geometry for surface CXD employed a coherent x-ray beam at grazing-incidence angle and examined the resulting distribution of intensity near to the specular direction.[7] This geometry was very similar to that used previously for studying the multilayers, again with the result that the CXD pattern was one dimensional. The restriction to Bragg diffraction no longer applies in the specular reflectivity case and so the perpendicular momentum transfer, q_z, can be varied in a continuous manner. The variation of the richness of the speckle with q_z was found to be a powerful way to characterize the surface roughness.[7] An empirical rule was established for the point of emergence of strongly modulated speckle features, $q_z \geqslant \pi/\sigma$, where σ is the root-mean-square roughness. This rule will be explained in the present work.

The requirements for coherence, necessary for these experiments to work, have been discussed previously.[4] The source coherence may be divided conveniently into two components, measured as lengths in the longitudinal and transverse directions. The longitudinal coherence length, $\xi_\parallel$, is the distance along the incident beam over which different rays largely retain their relative phases with respect to each other. It is limited by the distribution of wavelengths present, $\Delta\lambda/\lambda$, and so is a property of the monochromator used for the experiments, according to $\xi_\parallel = \lambda/(\Delta\lambda/\lambda) = \lambda^2/\Delta\lambda$. $\xi_\parallel$ is critical in a CXD experiment because it restricts the maximum path length difference (PLD) that can be tolerated between different rays passing from source to detector and is often dominated by beam penetration inside the sample. The lateral coherence length, $\xi_\perp$, is the distance across the beam over which all rays traveling from different points in the source largely retain their relative phases with respect to one another. It is limited by the source size, σ, according to $\xi_\perp = \lambda D/\sigma$, where D is the distance of the experiment from the source. $\xi_\perp$ is important because it determines the maximum size of the beam-defining aperture, $d \leqslant \xi_\perp$, that can be used. Because of the two coherence lengths, the requirement for coherence limits the maximum total flux that is available in a CXD experiment to an amount that is directly proportional to the brilliance of the source. In this paper we will discuss the origin of the speckle pattern seen in the surface reflectivity experiments, its asymptotic properties, and demonstrate an algorithm which can be used to *invert* it and reconstruct the actual morphology of the reflecting surface.

II. SPECKLE PATTERNS AND SURFACE MORPHOLOGY

All diffraction experiments detect the interference of waves scattered by a superposition of scatterers, either atoms or groups of atoms. In a disordered system, an ensemble average takes place over many examples of local configurations to give statistically meaningful information. The essential point of CXD is that the diffraction signal measured from a sample under coherent illumination conditions is the magnitude squared of the scattering amplitude *without the ensemble averaging*. We simply observe the direct summation over all atoms, at positions $\mathbf{r}_j$, within the illuminated volume,

$$A'(\mathbf{q}) \propto \sum_j f e^{i\mathbf{q}\cdot\mathbf{r}_j}. \tag{1}$$

The surface is conveniently defined by a single-valued boundary function $z=h(x,y)$, where the z direction is taken to be perpendicular to the surface, and (x,y) are the two coordinates in an ideal plane defining the average surface.[8] We impose a rigid periodicity perpendicular to this average plane so that the position of the jth column of atoms is at $r_{j,j_3}=(x_j,y_j,a_3 j_3)$, where a_3 is the vertical lattice spacing. For simplicity, we are assuming a sample made of a single kind of atoms of form factor, f.

When the surface is present, we can write the sum in Eq. (1) in the following way:

$$A'(\mathbf{q}) \propto \sum_j \sum_{j_3=-\infty}^{h(x_j,y_j)/a_3} f e^{i(q_x x_j + q_y y_j + q_z a_3 j_3)}.$$

Using the substitution $j_3 = j_3' + h(x_j,y_j)/a_3$ and summing over j_3', we obtain for the scattered amplitude,[9]

$$A'(\mathbf{q}) \propto F_{\mathrm{CTR}}(q_z) \sum_j e^{i(q_x x_j + q_y y_j)} e^{i q_z h(x_j,y_j)}, \tag{2}$$

where the summation has been reduced in dimension by summation over the *columns* of equally spaced atoms at lateral position x_j, y_j starting at height $h(x_j,y_j)$. The same result could be obtained equivalently in its integral form using Green's theorem.[8] The momentum transfer $\mathbf{q}=(q_x,q_y,q_z)$ has also been split into its components parallel to the surface, q_x and q_y, and perpendicular, q_z. The sum in the z direction is identical for all columns and takes the familiar form seen in the analysis of crystal-truncation rods (CTR's), which gives an amplitude that interpolates between the Bragg peaks in q_z,[10]

$$F_{\mathrm{CTR}}(q_z) = f[1-\exp(-iq_z a_3)]^{-1}. \tag{3}$$

To continue the derivation for ordinary x-ray diffraction from surfaces, the next step would be to introduce some correlation function for $h(x,y)$, and take the ensemble average $\langle |A'(\mathbf{q})|^2 \rangle$.[8] Instead for CXD, we can look directly at $|A'(\mathbf{q})|^2$, which should be compared with what is observed in the experiment.

The expression for $A'(\mathbf{q})$ in Eq. (2) has a very simple form. The first factor is the well-known CTR amplitude, given in Eq. (3). This tells us that the signal will be large near to Bragg peaks in q_z, and gets progressively smaller as we move away. Near to a bulk Bragg peak at G_z, the intensity falls off as $|(q_z - G_z)|^{-2}$. Most importantly, since we are concerned with reflectivity, this is also true for the origin of reciprocal space, $G_z=0$. The properties of F_{CTR} are discussed elsewhere[10] and are not relevant to the rest of the paper. The second factor in Eq. (2), which we call the speckle amplitude, can be rewritten in the continuum limit as a kind of Fourier transform,

$$A(q) = \sum_j e^{i(q_x x_j + q_y y_j)} e^{i q_z h(x_j,y_j)},$$

$$= \int\int_\Omega \rho(x,y) e^{i(q_x x + q_y y)} dx dy, \tag{4}$$

where

$$\rho(x,y) = e^{i\varphi(x,y)}, \quad \varphi(x,y) = q_z h(x,y). \tag{5}$$

This speckle amplitude $A(q)$ is the fundamental representation of the surface CXD, which we will be discussing in the remainder of the paper. There are two unusual aspects of this Fourier transform compared with the familiar crystallographic expressions relating electron density in a structure with its diffraction pattern. The first is that the domain of the function is *finite*, cutoff outside the illuminated region, Ω. We will assume a square beam for simplicity for which Ω is defined by the edges of the beam at positions $(x,y)=(\pm d/2, \pm d/2)$. The second is that the integrand, $\rho(x,y)$, is itself a complex function with unit amplitude and a position-dependent phase, which depends on both the height function and the perpendicular momentum transfer.

The expression in Eqs. (4)–(5) is a more general version of the model used in fitting the static speckle from the multilayer.[4] There, it had been assumed that the coherent beam could be divided into ''blocks'' that received different phases after diffracting from the discrete regions of the multilayer. The physical interpretation for $\varphi(x,y)$ is clear: the x-ray beam incident on the sample at position (x,y) becomes phase shifted by amount $\varphi(x,y)$, which depends on height in the case of surface CXD, or the position of the diffracting planes in the case of the multilayer. The *phase*-modulated beam undergoes mutual interference, which then results in the *amplitude* modulations seen in the speckle pattern.

The experimentally measured quantity is the intensity of CXD,

$$I(\mathbf{q}) \propto |A(\mathbf{q})|^2. \tag{6}$$

So, the measurement loses all the phase information of the amplitude $A(\mathbf{q})$ and reduces to the classical inversion problem of reconstructing the object from the intensity distribution of the scattered radiation. This is a more difficult problem than the inversion of a diffraction pattern to obtain the electron density because the quantity sought is itself a phase. For example, we can no longer use the inversion symmetry that arises from the real nature of the electron density. Inversion problems of this kind have been intensively studied in the past,[11] and there are some powerful theoretical methods available. The most relevant analogy is to the problem of holographic reconstruction: transmission holography is achieved by coherent illumination of an transparent object over a finite aperture; the interference of all transmitted

waves, recorded with sufficient resolution, is the hologram.[12,14,15] Except for the requirement in holography for a *reference* wave, this is the same thing as the CXD speckle pattern. If the object modulates the amplitude of the wave, this amplitude is the argument of the Fourier transform. But if the object has only phase contrast with no amplitude change, then the situation is identical to that of Eq. (4). For this reason, we propose to employ the image reconstruction algorithm described below.

The appearance of q_z in the expression for $\varphi(x,y)$ provides us with an opportunity that is specific to the surface CXD problem. If we can measure the amplitude $A(\mathbf{q})$ at a number of different q_z values, then we would have redundant information concerning the $h(x,y)$ functions sought, coupled to the phases $\varphi(x,y)$ through different coefficients q_z. The additional information could then be used in a reconstruction procedure.

On the other hand, if the surface morphology, i.e., the function $h(x,y)$ were known independently, then according to (4)–(5) the full function $A(\mathbf{q})$ could be calculated. This is illustrated at the top of Fig. 1, where the surface morphology of a silicon wafer was measured with a commercial profilometer.[7] The profilometer stylus rides over an approximately three micrometer-wide strip of surface. The 300-micrometer trace, $h(x,0)$, therefore covers about the same area as the beam in the surface speckle measurements.[7] The speckle intensity corresponding to this trace $I(q_x,q_z)=|A(q_x,0,q_z)|^2$ is plotted for this one-dimensional case by the direct evaluation of (4)–(6). The observed speckle patterns are horizontal cuts across this diagram, and the smooth evolution as q_z is varied can be seen immediately. For example, if the speckle pattern is measured at a particular value of q_z, and the function $h(x,y)$ is somehow derived from it, then the pattern can be verified at other values of q_z as a validity check.

Even though the q_z scale in Fig. 1 is a thousand times wider than that of q_x, the pattern is notably anisotropic, being strongly streaked in the q_z direction. This is a general characteristic of surface diffraction, and is not changed when the F_{CTR} term is restored, since that shares the same property. The V shape of the calculated pattern arises from the bimodal slope of the profilometer trace, as if there were (locally) two different specular directions, corresponding to the two characteristic slopes seen in the profilometer trace.

We can understand some general properties of surface speckle patterns, and their relation to surface morphology, by the examination of Eqs. (4)–(5). Any arbitrary rough surface can be considered as a superposition of roughness components on different length scales, often as a power spectrum. Such a description implies a statistical average over an ensemble of surfaces, and attaches a *root-mean-square roughness*, σ, which is just the spectral weight expressed as a height variation, to each length scale. This roughness spectrum will then determine the qualitative features of the speckle pattern as a function of q_z.

For example, if we consider a surface with just two length scales of importance, with long-range roughness σ_1 and a short-range roughness $\sigma_2 \ll \sigma_1$ with a finer length scale. The CXD pattern, given by Eqs. (4)–(5), will then split into three ranges of q_z. So long as $q_z \ll \pi/\sigma_1$ we will have $\rho(x,y)\simeq 1$ and a speckle pattern which is a simple slit function, as would be formed by the direct x-ray beam itself. When $\pi/\sigma_1 \ll q_z \ll \pi/\sigma_2$, the speckle pattern would be determined by the long-range roughness alone, and might still be relatively simple, not extending far in q_x. At the greatest values of q_z, when $q_z \gg \pi/\sigma_2$ the speckle pattern finally is sensitive to the short-range roughness as well, and will spread out much further in q_x. Thus, by tuning of the experimental variable q_z, patterns of increasing complexity could be generated and analyzed progressively. The evolution of complexity is clearly seen in Fig. 1 and corresponds closely with what was seen experimentally.[7] In practice this should be done by changing x-ray *energy*, rather than the angle of incidence, because the changing angle would also change the shape of the beam footprint on the sample. There are also severe limitations in the accuracy of diffractometer components that make it hard to vary the angle while illuminating exactly the same spot on the sample.

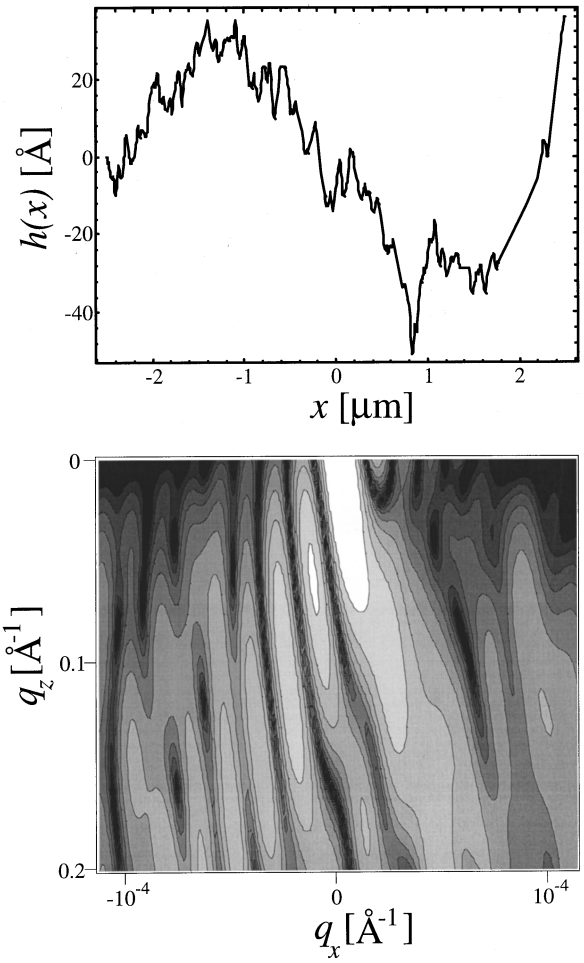

FIG. 1. Simulated speckle patterns from a one dimensionally rough surface specified by the $h(x)$ function in the top panel, which was measured from a Si(111) wafer using a profilometer. The x axis represents a position across the beam and has been shortened to account for the grazing incidence angle, for consistency with the rest of the paper. Below, logarithmic contours of $H(\mathbf{q})$ are plotted as a function of q_x and q_z. Note that both top and bottom panels use the same 1000:1 aspect ratio, so the ridges in the lower panel are seen to lie perpendicular to the sloped sides of the profile in the top panel.

One further consideration is very important for the design

of experiments to measure speckle from surfaces: the longitudinal coherence requirements. According to the formalism above, the PLD entering into the sum in Eq. (2) at $q_x=0$ is just $\lambda\varphi(x,y)$, which is very small indeed since the diffraction conditions (i.e., q_z) would be chosen to have $\varphi(x,y)$ not vary by more than a few π. Large PLD's *are* present between the terms in Eq. (1), but these all accumulate in the factor F_{CTR}, which is common to all terms of Eq. (2). The PLD constraint can therefore be written,

$$\lambda[q_z\delta h(x,y)+q_x d/2]<\xi_{\parallel}, \tag{7}$$

where $\delta h(x,y)$ is the excursion of the height, $h(x,y)$, over the range $-d/2<x<d/2$. The second term in Eq. (7) accounts for the lateral component of the PLD accumulated at the edges of the illuminated area. This latter term often dominates the roughness term, and can limit the effective range of scans in q_x that can be reached before coherence is lost. $\xi_{\parallel}$ is the longitudinal coherence length, defined in the introduction, which is determined in an experiment by the monochromator or the source characteristics. Even in the case where $\xi_{\parallel}$ is extremely short, say 70λ for a raw undulator beam, there will be broadening in q_z of $F_{\mathrm{CTR}}(q_z)$ in Eq. (3), but this is already a slow-varying function.

More sophisticated approaches to coherence exist in the optics literature, dealing with the case of *partial* coherence that occurs when the quantities appearing in inequality (7) become approximately equal. Recent progress has been made to measure the degree of coherence using x-ray diffraction methods.[2,16,17] The most important conclusion for the case of surface CXD is that expression (7) is considerably less demanding than the PLD constraints on bulk diffraction, where the large penetration depth can be insurmountable. Note that we are not necessarily constrained to the small-angle case of $q_z\approx 0$, if the surface is sufficiently flat.

The effects of longitudinal coherence are seen in the context of the speckle amplitude directly by a suitable construction in Fig. 1. The definition of $\xi_{\parallel}=\lambda/(\Delta\lambda/\lambda)$ means that the effect of longitudinal coherence can be considered to be a radial smearing of the **q**-space diffraction pattern due to the distribution of wavelengths, $\Delta\lambda/\lambda$. The features of Fig. 1 that are already streaked in the radial direction will be affected very little by this procedure, and so the limited coherence will not change the result. Conversely, features that are diagonal in the diagram, not emanating from the origin, will be strongly affected, and eventually washed away by overlap with their neighbors.

III. PROPERTIES OF THE SPECKLE AMPLITUDE

In coherent x-ray reflectivity experiments the sample is arranged at grazing incidence angles. Due to this arrangement, the footprint of the beam on the sample is highly elongated and its speckle pattern becomes one dimensional, to a good approximation. Throughout the remainder of the paper we will consider only this 1D case. The results obtained may be generalized to the 2D case in a straightforward manner. In 1D we have for the amplitude

$$A(q_x)=|A(q_x)|e^{i\alpha(q_x)}=\int_{-d/2}^{d/2}\rho(x)\exp(iq_x x)dx, \tag{8}$$

$$\rho(x)=e^{i\varphi(x)},\quad \varphi(x)=q_z h(x). \tag{9}$$

While deriving expressions (8)–(9) we implicitly assumed the following:

(i) The incident radiation is a coherent plane wave.

(ii) The rough surface is an ideal *phase object* which changes only the phase $\varphi(x)$ of the reflected beam across the surface. This implies that the amplitude of the reflected beam is constant.

(iii) The aperture of the entrance slit alone determines the size of the illuminated region on the sample. The sample is in the near field, a short distance, L, behind the entrance slit of size d, satisfying $d^2\gg\lambda L$.

(iv) The registration (detection) plane is far away. This is the far field approximation which gives rise to the Fraunhofer limit of diffraction. This assumption is valid if $d^2\ll\lambda L_1$, where L_1 is the distance from the sample to the detector.

These assumptions are the reason that expressions (8)–(9) are so simple. However, certain real experimental conditions, for example, the finite distance from the entrance aperture to the sample, can be allowed for (see below) without changing the main results of the following analysis. We will now analyze what information can be obtained in the framework of our model in different limiting cases.

A. Asymptotic behavior as $q_x\to\infty$

Integrating (8) by parts we obtain

$$\begin{aligned}A(q_x)&=\int_{-d/2}^{d/2}e^{i\varphi(x)}e^{iq_x x}dx\\&=e^{i\varphi(x)}\frac{e^{iq_x x}}{iq_x}\Bigg|_{d/2}^{-d/2}\\&\quad-\frac{1}{iq_x}\int_{-d/2}^{d/2}e^{i\varphi(x)}e^{iq_x x}(i\varphi'(x))dx,\end{aligned} \tag{10}$$

where $\varphi'(x)$ is the spatial derivative of the phase. For large q_x values we can neglect the second term in Eq. (10) relative to the first one and we would have for the intensity far from the center of the speckle pattern,

$$I(q_x)\propto|A(q_x)|^2=\frac{4}{q_x^2}\sin^2\left[\frac{q_x d+\Delta\varphi}{2}\right], \tag{11}$$

where $\Delta\varphi=\varphi(d/2)-\varphi(-d/2)=q_z[h(d/2)-h(-d/2)]$.

We can see from this result that the intensity distribution in this limit is just a slit function, $I_0(q_x)=(4/q_x^2)\sin^2(q_x d/2)$, shifted by the phase value $\Delta\varphi$. It is possible in principle to determine the phase shift $\Delta\varphi$ directly from the shift of the fringes of the intensity distribution relative to those of the direct beam, although in the experimental situation this would be equivalent to a change in the sample alignment. From the shift of the fringes, the height difference across the sample $[h(d/2)-h(-d/2)]$ could be determined. Indeed, a change in this height difference is nothing other than a small rotation of the sample, which leads to a shift of the pattern across the detector.

If we continue our expansion over the terms $(1/q_x)$ in Eq. (10), integrating the second term of Eq. (10) by parts, we would obtain the next term in the expansion of the intensity over small $(1/q_x)$ values

$$I(q_x) \propto \frac{4}{q_x^2}\sin^2\left[\frac{q_x d+\Delta\varphi}{2}\right]\{1-1/q_x[\varphi'(d/2)+\varphi'(-d/2)]\}. \tag{12}$$

B. Asymptotic behavior as $q_x \to 0$

The other limiting case to consider is small q_x values, giving the intensity distribution in the center of the speckle pattern. In this case we can expand the exponent in the integral of Eq. (8), $e^{iq_x x}=1+iq_x x-\cdots$ and obtain for the amplitude,

$$A(q_x)=\int_{-d/2}^{d/2} e^{i\varphi(x)}[1+iq_x x]dx = d[M_0+i(q_x d)M_1], \tag{13}$$

where the complex numbers M_0 and M_1 are the average value $M_0=\langle e^{i\varphi(x)}\rangle_d$ and the first moment $M_1=\langle (x/d)e^{i\varphi(x)}\rangle_d$ of the function $e^{i\varphi(x)}$. Note that the symbols $\langle\rangle_d$ denote *specific* averages over the finite range of integration of Eq. (13) and *not* ensemble averages. Hence we obtain for the intensity,

$$I(q_x)\propto d^2\{|M_0|^2-2q_x d\,\mathrm{Im}[M_0^* M_1]\}, \tag{14}$$

where we have neglected the small quadratic $(q_x d)^2$ term.

From this result we can see that the intensity in the center of the speckle pattern is determined by the average value,

$$I(q_x)\propto d^2|M_0|^2 = d^2|\langle e^{i\varphi(x)}\rangle_d|^2. \tag{15}$$

For example for a rough surface which is flat in the range of interest so that $\langle\varphi(x)\rangle_d=0$, we obtain

$$M_0=\langle e^{i\delta\varphi(x)}\rangle_d \simeq e^{-(1/2)\langle\delta\varphi(x)^2\rangle_d}=e^{-(1/2)q_z^2\sigma_d^2}, \tag{16}$$

where $\delta\varphi(x)=\varphi(x)-\langle\varphi(x)\rangle_d$ and σ_d is the *specific* roughness of the surface averaged over the range $-d/2<x<d/2$, since $\langle\delta\varphi(x)^2\rangle_d=q_z^2\sigma_d^2$. So, in this case we have for the intensity in the center,

$$I(q_x=0)\propto d^2 e^{-q_z^2\sigma_d^2}, \tag{17}$$

which resembles the classical (incoherent) result. For a rough surface with a sufficiently large σ_d value we find that $I(q_x=0)\to 0$.

C. Perturbation limit, $\varphi \ll \pi$

In the case of the small phases φ we can expand our density function,

$$\rho(x)=e^{i\varphi(x)}=1+i\varphi(x)-\cdots, \tag{18}$$

and hence obtain the speckle amplitude,

$$A(q_x)=\int_{-d/2}^{d/2}[1+i\varphi(x)]e^{iq_x x}dx = A_0(q_x)+i(q_z d)H(q_x), \tag{19}$$

where

$$H(q_x)=\frac{1}{d}\int_{-d/2}^{d/2} h(x)e^{iq_x x}dx \tag{20}$$

is the Fourier transform of the height function $h(x)$ itself and $A_0(q_x)=(2/q_x)\sin(q_x d/2)$ is the slit function.

This gives for the intensity,

$$I(q_x)\propto |A_0(q_x)|^2-2q_z d A_0(q_x)\mathrm{Im}[H(q_x)], \tag{21}$$

where we have neglected the small quadratic term $|H(q_x)|^2$. Thus the height information in $H(q_x)$ appears as a perturbation from the ideal Fraunhofer diffraction of the direct beam.

According to this expression we can determine the imaginary part of $H(q_x)$ from the difference between an experimental curve $I^{\mathrm{exp}}(q_x)$ and $|A_0(q_x)|^2$, as measured for the direct beam,

$$\mathrm{Im}[H(q_x)]=\frac{I^{\mathrm{exp}}(q_x)-|A_0(q_x)|^2}{2|A_0(q_x)|}. \tag{22}$$

In the case of incoherent diffraction from a rough surface[8] the resulting diffraction pattern clearly separates into the sum of a δ function and diffuse components. With coherent diffraction, on the other hand, we generally cannot make this separation. As shown above, a partial separation is possible in the limit of small phases, when the surface structure perturbs the tails of the slit function $|A_0(q_x)|^2$, leaving the central part of the intensity nearly unchanged.

IV. RECONSTRUCTION OF $h(x)$ FROM STATIC CXD

The ideal CXD experiment we have described is formally identical to the measurement of the Gabor hologram of the object under investigation.[13] There an object, usually of spatially varying absorption, is illuminated by coherent radiation, and all scattered waves are collected in the far field, with a sufficient resolution to resolve all the interference fringes. In this case, the *amplitude* of the function $\rho(x)$ would represent the absorption, and its phase would be constant, ignoring refraction effects. If the object field is suitably transparent that a large fraction of the beam passes unmodified, or alternatively a *reference wave* is deliberately added, we would have the equivalent of the small phase approximation described above, in which the hologram is a perturbation of the ideal diffraction pattern of the beam itself.

From the considerable literature on the subject of image reconstruction, we have selected an algorithm[18] that has the potential of inverting Eq. (8) to recover the height function $h(x)$. In one dimension, the algorithm is summarized by the pair of equations,

$$A(q_x)e^{i\alpha(q_x)} := \int_{-\infty}^{\infty} B(x)e^{i\varphi(x)}\exp(iq_x x)dx,$$

$$\rho(x)e^{i\varphi(x)} := \frac{1}{2\pi}\int_{-\infty}^{\infty}\sqrt{I(q_x)}e^{i\alpha(q_x)}\exp(-iq_x x)dq_x, \tag{23}$$

where we have used the symbol ''$:=$'' to mean ''is replaced by.'' The method iterates back and forth between real space and reciprocal space by the repeated application of Eq. (23). In the numerical application, the Fourier transforms would be achieved by fast Fourier transforms (FFT's), as described below. In both real and reciprocal space, constraints are ap-

plied to guide the algorithm to find a solution consistent both with the observed data and with the expected form of $\rho(x)$. The speckle amplitude in reciprocal space, $A(q_x)$, generated during one cycle is overwritten by the observed amplitude, $\sqrt{I(q_x)}$. Similarly, the amplitude of the real-space density, $\rho(x)$, is reset to a predefined real function, $B(x)$, while retaining the phase information, $\varphi(x)$. For the present discussion, we consider $B(x)$ to be a ''box'' function, with unit value within the illuminating aperture, and zero outside.

In the context of holography, the procedure would be identical if the object being imaged were a *phase object* that modified the phase of the illuminating radiation. More commonly in holographic reconstruction the object under investigation would have an amplitude contrast, often blocking the beam completely over certain regions. In the latter case, the function $\rho(x)$ would be constrained to be real in the return half of the procedure instead by resetting its phase to zero upon every cycle of the algorithm.

We experimented with this algorithm and found that there were some important details that needed attention. These are discussed in the following subsections. They lead to some practical constraints on the procedure.

A. Illumination filter function

Over the course of the cycle, the real-space density, $\rho(x)$, leaks outside the range $\pm d/2$ and requires cutting off; however, if this is done abruptly, strong fringes will appear in the diffraction. These fringes are the desired Fraunhofer fringes of the speckle pattern, but in the experiment they are smeared by the detector resolution and imperfect beam coherence. These effects can be partly accounted for by choosing in Eq. (23) an appropriate filter function $B(x)$, such as a rounded box function.

This filter function can also be used to account for the change of the illuminated area on the sample due to its finite separation from the defining aperture. The form of $B(x)$ would then depend on the distance L of the sample from the entrance slits, in a way that can be predicted. The derivation requires retaining the quadratic terms in the Fresnel limit of diffraction.[20] In this near-field approximation, $L \leqslant d^2/\lambda$, we obtain the following expression for the illumination across the sample:

$$B(x)=\frac{1}{\sqrt{2}}\{[C(z_2)-C(z_1)]+i[S(z_2)-S(z_1)]\}, \quad (24)$$

where $C(z)$ and $S(z)$ are Fresnel integrals,

$$C(z)=\int_0^z \cos\left(\frac{\pi t^2}{2}\right)dt, \quad S(z)=\int_0^z \sin\left(\frac{\pi t^2}{2}\right)dt,$$

and $z_1=\sqrt{2/\lambda L}(x-d/2)$, $z_2=\sqrt{2/\lambda L}(x+d/2)$.

As we can see from this expression, the function $B(x)$ is complex, meaning that the illumination changes phase across the sample and is no longer an ideal plane wave. In Fig. 2 we plot the absolute value, $|B(x)|$, as solid curves for different distances, L. For small distances $L \ll d^2/\lambda$, as in the top panel of Fig. 2, the phase is almost constant across the central part of the aperture, $-d/2<x<d/2$. Important phase deviations start when $L \approx d^2/\lambda$, shown in the middle panel, and the magnitude $|B(x)|$ has significant tails. In our fitting procedure, these additional phase shifts would be absorbed into the experimentally determined phase $\varphi(x)$. Finally, in the limit of big distances, $L \gg d^2/\lambda$ in the bottom panel, the function becomes real again and would itself look like a real-space slit function $B(x)=\sin(\beta x)/(\beta x)$, where $\beta=\pi d/\lambda L$. In this limit, the illumination is diffraction limited, with width $\lambda L/d$, and no longer resembles the shape of the defining slit at all. In practice, this large-L limit would not be interesting in experiments requiring a small illuminated area on the sample, because a *larger* aperture, d, could have been used for a gain in flux and would have resulted in a *narrower* illumination function.

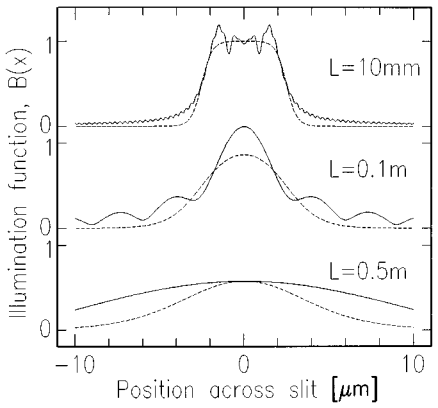

FIG. 2. The solid curves are the magnitude of the illumination function, $|B(x)|$, for different distances, L, of the sample from the entrance slits: sample near the slits (top); intermediate distances L (middle); limit of large distances $L \gg d^2/\lambda$ (bottom). Curves are calculated for $\lambda=0.15$ nm and $d=5$ μm.

Instead of calculating the exact illumination function, $B(x)$, for a particular experimental setup, and because $B(x)$ also needs to include the aforementioned filters, we choose to approximate it with fairly good accuracy as a product of a pair of Fermi functions,

$$B(x)=f(x-d/2)f[-(x+d/2)], \quad (25)$$

where $f(x)=1/[1+\exp(x/w)]$ and the slope parameter w determines the sharpness of the edge. The width of this filter function $B(x)$ is equal to the width of the slits and the slope parameter $w=\frac{1}{4}\sqrt{\lambda L}$ takes into account the smearing effects of the distance of the sample from the beam-defining aperture. The expression for w was obtained by differentiation of $|B(x)|$ at the slit edges $x=\pm d/2$. These approximate illumination functions are superimposed on the exact ones as dashed curves in Fig. 2. Clearly this approximation is good only for values of $L \leqslant d^2/\lambda$, which ensures the sample is in the near field of the slit (see Fig. 2).

Having defined $B(x)$ we can extend the limits of our Fourier transform (8) to infinity,

$$A(q_x)=\int_{-\infty}^{\infty}\rho(x)\exp(iq_x x)dx, \quad (26)$$

where

$$\rho(x)=B(x)e^{i\varphi(x)}. \quad (27)$$

We have constructed a new density function which adopts the amplitude of the illumination function $|\rho(x)|=|B(x)| \neq \text{const}$ and its argument into the sought-after phase $\varphi(x)$.

B. Uniqueness problem

The question of uniqueness must always be taken into consideration when we are dealing with inverse problems in which the phase information is lost in an experiment. When we have one solution to (8)–(9), we can always generate a second one, the so-called *dual solution*, using the properties of the Fourier transformation. If a complex function $\rho(x)=|\rho(x)|e^{i\varphi(x)}$ is the first solution, then

$$\tilde{\rho}(x)=\rho^*(-x)=|\rho(-x)|e^{-i\varphi(-x)}, \quad -d/2<x<d/2 \tag{28}$$

is its dual. $\tilde{\rho}(x)$ gives the identical intensity in the speckle pattern.

The general mathematical problem of the construction of a complex function given its modulus and the modulus of its Fourier transform as well as the problem of the uniqueness of the solution have been investigated in a series of papers (see references in Ref. 19). Even if there exist only two solutions, without other prior knowledge, it is impossible to choose between them. There are some situations where the problem can be avoided, however. For example, if we have data in the form of a *time series*, in which one scan is only slightly different from the previous one, the internal self-consistency could help to establish the uniqueness of the solution. Similarly, as mentioned above, if data are measured on the same area of the sample but with different q_z values, the fact that all measurements must correspond to the same $h(x)$ function can under some circumstances constrain the uniqueness, as we will discuss next.

The phase obtained from any calculation of this kind will always be bounded in some range such as $-\pi<\varphi(x)\leqslant\pi$. At any position, the phase value $\varphi(x)$ is indistinguishable from the phase value $\varphi(x)+2n\pi$. For this reason the derived phase function will be highly degenerate. The real height function, $h(x)$, is of course not necessarily bounded, but it is expected to be a continuous function. In principle this could be used to resolve any ambiguity in the mapping $\varphi(x)=q_z h(x)$, except for a single arbitrarily chosen phase point. In practice, the problem is not so straightforward because of the finite spacing of the q points (see below) and the effective coupling between them due to application of the Fourier transforms in applying the algorithm.

Let us consider an example of how this last kind of nonuniqueness could cause problems in trying to invert surface CXD data. If the surface height function $h(x)$ happened to contain a fairly abrupt step δh which resulted in a localized phase jump of $\delta\varphi=q_z\delta h\simeq 2\pi$, this would be almost invisible in measurements at that particular q_z. The fitting algorithm (23) could find a solution in which the step was omitted, and the agreement with the data would be already quite good, but not perfect. There would be little chance of such a feature being introduced *gradually* by the cycling of the algorithm because a broad or smaller version of the step would fit much worse than the full δh. To avoid this problem with safety, it would be necessary either to have data at more than one q_z, or to include testing for phase jumps, by inserting them between cycles to see if they are supported, as part of the general application procedure.

C. Data sampling

In the real experimental situation, data are measured within a finite region of reciprocal space, $q_{\min}\leqslant q_x\leqslant q_{\max}$. However, the domain of the inverse Fourier transformation must extend to infinity. The data must, therefore, be extended smoothly beyond the range of measurement. It is clear that, if they are cutoff abruptly, this will give rise to additional oscillations in real space which have no physical meaning. According to the properties of the Fourier transform, the far tails in reciprocal space will influence only the higher spatial-frequency features in real space. One of the simplest ways to extend the data is just to smoothly *cutoff* the long tails.

A more accurate way of extending the experimental data would be to make use of the asymptotic behavior of speckle intensity as $q_x\to\infty$, which we derived in Eq. (11). We could continue our experimental data smoothly outside the range $q_{\min}\leqslant q_x\leqslant q_{\max}$ using expression (11). Then the inverse Fourier transformation in the range $-\infty<q_x<q_{\min}$ and $q_{\max}<q_x<\infty$ can be performed analytically with this asymptotic function.

In numerical calculations, Fourier integrals such as Eq. (8) are calculated as discrete Fourier transforms (DFT), which are Fourier series,

$$A_n=\sum_{m=0}^{N-1}\rho_m\exp\left[\pm i\frac{2\pi mn}{N}\right], \quad m,n=0,\ldots,N-1, \tag{29}$$

where N is the total number of points of calculation, usually with small prime factors, as in the FFT. According to Eq. (29), the real and reciprocal spaces are connected to each other. Really, the nth point in real space is equal to a fraction of some unit cell length, X, by $x_n=Xn/N$, so the mth point in the reciprocal space according to Eq. (29) is equal to $q_m=2\pi m/X$. So, for the steps in the real and reciprocal space we have $\Delta x=X/N$ and $\Delta q=2\pi/X$ or for their product,

$$\Delta x\Delta q=2\pi/N. \tag{30}$$

Due to relationship (30) the steps in the real and reciprocal space are correlated and strictly cannot be taken independently. In practice, the interval Δq is fixed by the step size in scanning the diffraction pattern, or the pixel size of the detector, but we are usually free to choose the sampling interval Δx with which we represent the real-space functions.

A related practical consideration of using the DFT is the problem of *aliasing*, which arises from the periodic boundary conditions inherent in Eq. (29). This is usually avoided by choosing a suitably large value of N, and retaining a healthy amount of empty space within both the real- and reciprocal-space unit cells. This also allows room for the smooth continuation of the diffraction pattern outside the measured range, mentioned above. Relationship (30) will then guarantee that the spacing Δx is sufficiently small that the discretely sampled numerical function, $\varphi(x)$, will be oversampled and will therefore appear to be smooth.

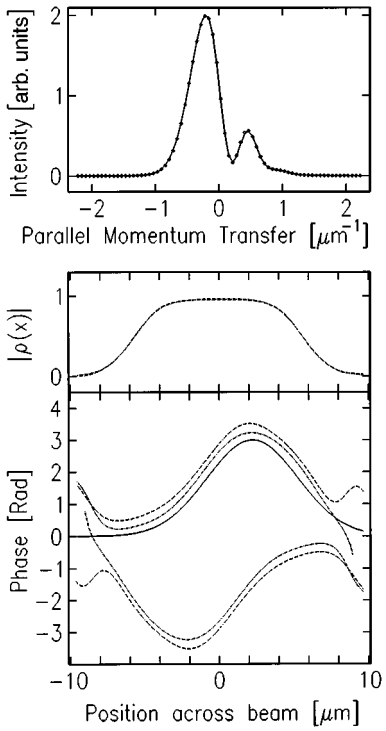

FIG. 3. Test of the Gerchberg-Saxton algorithm applied to a simple height function in the form of a Gaussian phase bump. Top: derived speckle pattern (points) and eventual fit. Middle: derived illumination as a function of position across the incident beam for two reconstruction attempts. Bottom: starting height function (solid) and two versions of the reconstruction (dashed and dash-dot) together with their duals. The two versions, which differ substantially only outside the illuminated aperture, were obtained using different random starting phases.

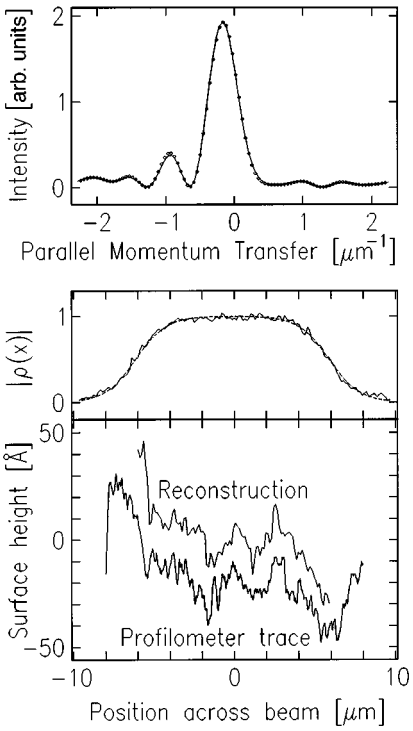

FIG. 4. Test of the algorithm on profilometer data. Top: derived speckle pattern (points) and eventual fit. Middle: assumed illumination as a function of position across the incident beam and eventual fit. Bottom: starting height function and reconstructed version.

D. Tests of the algorithm

Our first test was simply to show that a simple structure could be recovered from simulated data. This is shown in Fig. 3 for a Gaussian-shaped phase bump as the test structure. The simulated speckle pattern is shown in the upper panel as data points and the line passing through them was obtained using the Gerchberg-Saxton fitting algorithm we have described. The real-space amplitude and phase are shown in the lower panel. The data were generated using an illumination function $B(x)$ given by Eq. (25) with a width $d=11$ μm and slope parameter $w=1.25$ μm. This same illumination function was asserted on every cycle of the algorithm, except a slope parameter $w=1\mu$m was used instead, as a test. For starting phases, a random number was chosen for each value of x with $-\pi<\varphi(x)\leqslant\pi$. The procedure converged within 40 iterations to the correct solution or to its dual (about half the time for each), depending on the starting phase set.

For a simple function as in this example, the method was fairly reproducible, even when the illumination function was not chosen exactly the same as the true one. The two examples of reconstructed profiles shown in Fig. 3 by dashed and dash-dot curves were obtained with different starting phases. They are not quite the same: they both resemble the original in the center, but begin to diverge dramatically outside the range where the illumination function is large, $-d/2<x<d/2$. Since the phase has little weight outside this window region, it does not contribute much to the data and so is poorly constrained there.

The convergence of the algorithm was found to be data dependent. It has been reported that if the solution and its dual start to resemble each other, this can interfere with the convergence.[21] This is a well-documented *stagnation* failure of the Gerchberg-Saxton method when it becomes stuck with a function that adopts features of both the solution and its dual.[21] Stagnation was apparently not a problem in this simple test case, perhaps because of the particular choice of test function used.

Our second test, shown in Fig. 4, was more demanding. We considered a different example of a real surface morphology as measured with a profilometer.[7] A typical 300 μm trace was compressed into 12 μm to represent the grazing incidence angle of the beam on the sample. A rounded Fermi-function illumination function, $B(x)$, was assumed as in Eq. (25). The speckle amplitude was generated and then used, together with the same $B(x)$ and random starting phases to recover the surface height. The data are only shown in the center of the pattern, but have important small features extending far beyond the edges of the figure; retaining these is important for the fitting to work. The recovered surface structure as well as the initial one are shown in the bottom panel of Fig. 3. Clearly, the coarser features of the profile have been largely recovered by the algorithm but there remain some differences on a finer scale, which might be attributed to the finite range of the reciprocal-space Fourier transform.

The final test was on the reconstruction of an *unknown* surface structure from an example of data from silicon wafer

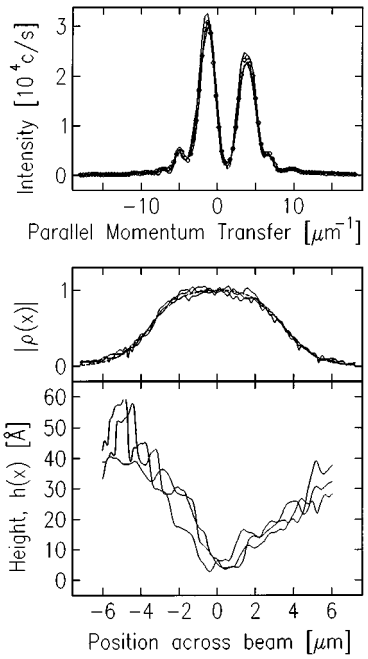

FIG. 5. Fit of a height function $h(x)$ to a 1D speckle pattern obtained by reflection from a silicon wafer at the ESRF using the methods described in Ref. 7. A perpendicular momentum transfer of $q_z=0.2$ Å^{-1}, a nominal pinhole of $d=7\,\mu$m and wavelength $\lambda=0.136$ nm were used. Three examples of reconstructions are shown superimposed.

samples, shown in the top panel of Fig. 5. These data were measured at ID10 of the European Synchrotron Radiation Facility (ESRF) using the methods described in Ref. 7. The illumination function was assumed to be of the form (25) with parameters estimated from the dimensions of the experimental setup: $d=7$ μm and $w=0.8$ μm. In our fitting procedure we made several attempts (starting with random phases) to fit the data and the solution was regarded as found when χ^2 values for both the real-space amplitude constraint $\Sigma[|\rho(x)|-|B(x)|]^2$ and the reciprocal-space data constraint $\Sigma\{\sqrt{I^{\rm obs}(q_x)}-|A(q_x)|\}^2$ became small. Of course we have no independent way of checking the validity of the derived surface profile, so we show three independent solutions, obtained with different random starting phases. The similarity of these gives confidence that the solutions may be correct.

V. CONCLUSIONS

We have shown how the coherent diffraction profiles seen in reflectivity experiments can be derived with a kinematic formalism. The expression for the reflected amplitude factorizes into a crystal-truncation part and an integral of a phase factor that depends on the height function, $h(x)$, describing the surface. One consequence of this factorization is that the reflection geometry is very favorable from the point of view of longitudinal coherence, so wide-band optics can be utilized for a considerable advantage in flux.

Interesting mathematical properties of the derived formalism were explored. An expression was derived for the far tails of the diffraction pattern, which are oscillatory but shifted in phase by an amount that depends on the overall height difference across the sample. In the perturbation limit, it was discovered that there is a threshold value of the perpendicular momentum transfer, $q_z\simeq\pi/\sigma$, beyond which a surface of a given roughness, σ, will start to produce strong speckle features in its coherent diffraction pattern. This was already seen experimentally.[7]

Finally, we tested the possibility of using a reconstruction algorithm, proposed by Gerchberg and Saxton, and widely used in the optics literature since. The algorithm switches back and forth between real space and reciprocal space, asserting a sample illumination function, $B(x)$, in the former and the observed data in the latter. The illumination function also acts as a filter to allow for partial coherence and the finite detector resolution. Our tests showed that images of surface morphology could be recovered from simulated data, and that satisfactory fitting of real data could be achieved.

Further investigations of the capabilities of this data-fitting algorithm are in progress. We intend to pursue the questions of the uniqueness of the solution from the empirical direction, and to explore the sensitivity to errors in the data and in the illumination function. We will investigate other kinds of real-space intervention in the Gerchberg-Saxton cycle, which might improve the convergence. We can extend the application to other kinds of data, for example, 2D speckle or speckle due to antiphase domains rather than a continuous height function. We would also like to look at time-series and q_z-dependent data. We believe the methods outlined here will help to establish CXD as a powerful analytical tool for probing the structure of matter on the nanometer length scale.

ACKNOWLEDGMENTS

We have enjoyed helpful discussions of these ideas with S. K. Sinha, G. B. Stephenson, S. B. Dierker, R. Pindak, V. G. Kohn, and C. Jacobsen. The data in Fig. 5 were obtained with the help of G. Grübel, D. Abernathy, and D. Smilgies. The work was supported by the NSF under Grant No. DMR93-15691.

[1] 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).

[2] Z. H. Cai, B. Lai, W. B. Yun, I. McNulty, K. G. Huang, and T. P. Russell, Phys. Rev. Lett. **73**, 82 (1994).

[3] S. Brauer, G. B. Stephenson, M. Sutton, R. Brüning, E. Dufresne, S. G. J. Mochrie, G. Grübel, J. Als-Nielsen, and D. L. Abernathy, Phys. Rev. Lett. **74**, 2010 (1995).

[4] 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).

[5] S. B. Dierker, R. Pindak, R. M. Fleming, I. K. Robinson, and L. Berman, Phys. Rev. Lett. **75**, 449 (1995).

[6] G. Grübel, D. Abernathy, T. Thurn-Albrecht, W. Steffen, A. Patkowski, G. Meier, and E. W. Fischer, ESRF Newsletter **26**, 10 (1996).

[7] J. L. Libbert, R. Pindak, S. B. Dierker, and I. K. Robinson (unpublished).
[8] S. K. Sinha, E. B. Sirota, S. Garoff, and H. B. Stanley, Phys. Rev. B **38**, 2297 (1988).
[9] I. K. Robinson, E. H. Conrad, and D. S. Reed, J. Phys. (Paris) **51**, 103 (1990).
[10] I. K. Robinson, Phys. Rev. B **33**, 3830 (1986).
[11] *Inverse Source Problems in Optics*, edited by H. P. Baltes, Topics in Current Physics Vol. 9 (Springer-Verlag, Berlin, 1978).
[12] G. W. Stroke, *An Introduction to Coherent Optics and Holography* (Academic Press, New York, 1966).
[13] D. Gabor, Nature (London) **161**, 777 (1948).
[14] M. Howells, C. Jacobsen, J. Kirz, R. Feder, K. McQuaid, and S. Rothman, Science **238**, 514 (1987).
[15] C. Jacobsen, M. Howells, J. Kirz, and S. Rothman, J. Opt. Soc. Am. **7**, 1847 (1990).
[16] T. Salditt, H. Rhan, T. H. Metzger, J. Peisl, R. Schuster, and J. P. Kotthaus, Z. Phys. B **96**, 227 (1994).
[17] A. Gibaud, J. Wang, M. Tolan, G. Vignaud, and S. K. Sinha, J. Phys. (France) I **6**, 1085 (1996).
[18] R. W. Gerchberg and W. O. Saxton, Optik **35**, 237 (1972).
[19] H. A. Ferwerda, in *Inverse Source Problems in Optics*, edited by H. P. Baltes, Topics in Current Physics Vol. 9 (Springer-Verlag, Berlin, 1978).
[20] M. Born and E. Wolf, *Principles of Optics*, 4th ed. (Pergamon Press, London, 1970).
[21] J. R. Fienup and C. C. Wackerman, J. Opt. Soc. Am. **3**, 1897 (1986).