

# Information content of the near field: three-dimensional samples

David G. Fischer,<sup>1,\*</sup> Richard A. Frazin,<sup>2</sup> Marius Asipauskas,<sup>3</sup> and P. Scott Carney<sup>4</sup>

<sup>1</sup>Research and Technology Directorate, NASA Glenn Research Center, Cleveland, Ohio 44135, USA

<sup>2</sup>Department of Ocean, Atmospheric and Space Science, University of Michigan, Ann Arbor, Michigan 48109, USA

<sup>3</sup>National Center for Space Exploration Research, NASA Glenn Research Center, Cleveland, Ohio 44135, USA

<sup>4</sup>Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA

\*Corresponding author: dgfischer@nasa.gov

Received October 8, 2010; accepted November 20, 2010;  
posted December 3, 2010 (Doc. ID 136061); published February 7, 2011

We present an analysis of the accuracy and information content of three-dimensional reconstructions of the dielectric susceptibility of a sample from noisy, near-field holographic measurements, such as those made in scanning probe microscopy. Holographic measurements are related to the dielectric susceptibility via a linear operator within the accuracy of the first Born approximation. The maximum-likelihood reconstruction of the dielectric susceptibility is expressed as a linear combination of basis functions determined by singular value decomposition of the weighted measurement operator. Maximum *a posteriori* estimates based on prior information are also discussed. Semianalytical expressions are given for the likely error due to measurement noise in the basis function coefficients, resulting in effective resolution limits in all three dimensions. These results are illustrated by numerical examples. © 2011 Optical Society of America

OCIS codes: 110.0180, 180.5810, 290.3200, 070.6020, 110.3000, 110.3010.

## 1. INTRODUCTION

Near-field microscopy [1,2] is an important optical imaging technology that provides resolution on scales much smaller than the Rayleigh–Abbe resolution limit [3]. Syngé first suggested the basic method, in which a thin sample is illuminated through a subwavelength aperture that is scanned very close to the sample while the intensity of transmitted light is recorded as a function of the position [4]. This method is now known as near-field scanning optical microscopy (NSOM) [5–8], and is often practiced by illuminating the sample and collecting the light through a subwavelength aperture at the end of a tapered optical fiber. Photon scanning tunneling microscopy (PSTM) is a related method, in which the sample is illuminated by an evanescent wave created by total internal reflection at a prism face and the subsequently scattered light is collected by a tapered fiber, as in NSOM.

Interpretation of near-field images is sometimes problematic. Under certain simplifying assumptions, such as homogeneity of the material dielectric susceptibility [6], the measurements may be related to the topography. Under more general conditions, a solution of the linearized scattering problem [9–14] may relate near-field optical measurements to the object structure. An experimental demonstration of the linearized inverse problem was presented in Ref. [15]. Nonlinear reconstructions, i.e., beyond the Born approximation, have been formulated, but present a serious challenge to implementation, namely, regarding the stability. The study of the linearized problem also provides a starting point for the analysis [16] of the nonlinear problem [17–20].

Important questions remain: what are the fundamental limits on information or resolution in near-field microscopy? How do noise, bounds on object size, and distance from probe to

sample affect these limits? This paper addresses these questions in a statistical context and builds on the work presented in a previous paper [21], henceforth called paper I. In paper I, the analysis was carried out for a sample that exhibits variation in only two dimensions (see also Refs. [22–27]). In the present paper, the solution of the linearized inverse scattering problem [10–13] is applied to the case of a three-dimensional, weakly scattering sample. The MLE of the sample susceptibility is expanded in terms of singular functions, and a semianalytic expression for the error covariance matrix of expansion coefficients is obtained for a Gaussian noise model. These formulas are then applied to examples in which object size, noise level, and the distance from the probe to the sample are varied. The reconstruction quality is assessed, and the information content is measured by counting the number of available singular functions.

## 2. SCATTERING MODEL

Consider an experiment in which a monochromatic plane wave is incident on a sample, located in the half-space  $z \geq 0$ , with dielectric susceptibility  $\eta(\mathbf{r}) = (n_s^2(\mathbf{r}) - 1)/4\pi$ , where  $n_s$  is the complex index of refraction of the sample and  $\mathbf{r} = (\rho, z)$  is the coordinate vector. The half-space  $z \geq 0$  has a (background) index of refraction of  $n = 1$ , while the other half-space,  $z < 0$ , has an index of refraction of  $n$ . Only nonmagnetic materials are considered, and, so, it is sufficient to only treat the electric field  $\mathbf{E}$ . The field satisfies the equation

$$\nabla \times \nabla \times \mathbf{E}(\mathbf{r}) - n^2(z)k_0^2 \mathbf{E}(\mathbf{r}) = 4\pi k_0^2 \eta(\mathbf{r}) \mathbf{E}(\mathbf{r}), \quad (1)$$

where  $k_0$  is the free-space wave number and  $n(z)$  is the  $z$ -dependent background index of refraction described above.

The sample is assumed to be of a thickness not greater than  $\Delta z$  at any transverse spatial coordinate  $\rho$ .

The field can be decomposed into two parts as  $\mathbf{E} = \mathbf{E}^i + \mathbf{E}^s$ , where  $\mathbf{E}^i$  is the incident field, present when  $\eta = 0$ , and  $\mathbf{E}^s$  is the scattered field. It is convenient to separate the wave vectors into transverse and longitudinal ( $z$ ) components as follows:

$$\mathbf{k}(\mathbf{q}) = [\mathbf{q}, k_z(\mathbf{q})], \quad (2)$$

$$k_z(\mathbf{q}) = \sqrt{k_0^2 - |\mathbf{q}|^2}. \quad (3)$$

The modes for which  $|\mathbf{q}| < k_0$  are homogeneous, or propagating, plane waves. When  $|\mathbf{q}| > k_0$ , the modes are evanescent plane waves, decaying exponentially with increasing values of  $z$ . These waves are superoscillatory in the transverse plane and thus provide a means to probe spatial frequencies beyond the Rayleigh–Abbe limit in the far field [13].

The (possibly evanescent) incident field has the form

$$\mathbf{E}^i(\mathbf{r}) = A^i \hat{\mathbf{e}}^i \exp[i\mathbf{k}(\mathbf{q}^i) \cdot \mathbf{r}], \quad (4)$$

where  $\mathbf{q}^i$  is the transverse wave vector of the incident plane wave,  $\hat{\mathbf{e}}^i$  is the complex polarization unit vector, and  $A^i$  is the real field amplitude, so that  $(A^i)^2$  has units of energy/(time · area). In PSTM,  $k_0 < |\mathbf{q}^i| \leq nk_0$ , and in NSOM,  $|\mathbf{q}^i| \leq k_0$ .

Keeping only the first term in the Born series, the scattered field is given by the expression [12]

$$E_\alpha^s(\mathbf{r}) = k_0^2 A^i \int d^3 r' \eta(\mathbf{r}') G_{\alpha\beta}(\mathbf{r}, \mathbf{r}') \hat{e}_\beta^i \exp[i\mathbf{q}^i \cdot \boldsymbol{\rho}' + ik_z(\mathbf{q}^i)z'], \quad (5)$$

where the summation convention for repeated indices is implied,  $G$  is the half-space Green's tensor given in Appendix A, and  $\boldsymbol{\rho}'$  is equal to the transverse components of  $\mathbf{r}'$ .

The measurements are made in the plane  $z = z_d$ , which is illuminated by a reference plane wave

$$\mathbf{E}^r(\mathbf{r}) = A^r \hat{\mathbf{e}}^r \exp[i\mathbf{k}(\mathbf{q}^r) \cdot \mathbf{r}]. \quad (6)$$

It is important to note that for the NSOM and PSTM imaging modalities, as well as some forms of total internal reflection tomography (TIRT), the incident field also plays the role of the reference field. The total field in the detector plane is given by the superposition of the reference and scattered fields. The resulting total intensity is given by

$$I^{\text{total}}(\boldsymbol{\rho}, z_d) = E_\alpha^r E_\alpha^{r*} + E_\alpha^s E_\alpha^{s*} + E_\alpha^s E_\alpha^{r*} + E_\alpha^r E_\alpha^{s*}, \quad (7)$$

where the spatial arguments of the fields have been suppressed and the  $*$  notation indicates that the complex conjugate is to be taken. We emphasize that in this classical viewpoint  $I^{\text{total}}$  is a deterministic, noise-free quantity that is not measured and hence not exactly known. The associated measured quantity is a noisy version that is given by a random variable defined (in terms of photocounts) later in the paper. The first term represents a uniform background term at a fixed  $z_d$ . The second term is necessarily nonlinear in the susceptibility. Under weak scattering conditions, it is negligible compared to the third and fourth terms and is disregarded hereafter. However, if the reference field is evanescent, the

scattering term becomes more important and eventually the second term dominates as  $z_d$  increases. The third and fourth terms are conjugates that carry image information and together are called the holographic intensity:

$$I(\boldsymbol{\rho}, z_d) \equiv E_\alpha^s E_\alpha^{r*} + E_\alpha^r E_\alpha^{s*}. \quad (8)$$

It should be noted that  $I(\boldsymbol{\rho}, z_d)$  may be positive or negative.

Determination of  $\eta$  is facilitated by taking the 2D Fourier transform of the holographic intensity  $I(\boldsymbol{\rho}, z_d)$ , i.e.,

$$\tilde{I}(\mathbf{q}, z_d) \equiv \frac{1}{2\pi} \int d^2 \rho I(\boldsymbol{\rho}, z_d) \exp[i\mathbf{q} \cdot \boldsymbol{\rho}]. \quad (9)$$

Suppressing the dependence on  $z_d$  and substituting Eqs. (5) and (6) into Eq. (8) yields

$$\begin{aligned} \tilde{I}(\mathbf{q}) = & \int_0^{\Delta z} [H(\mathbf{q}, \mathbf{q}^r, \mathbf{q}^i, z) \tilde{\eta}(\mathbf{q} + \mathbf{q}^i - \mathbf{q}^r, z) \\ & + H^*(-\mathbf{q}, \mathbf{q}^r, \mathbf{q}^i, z) \tilde{\eta}^*(-\mathbf{q} + \mathbf{q}^i - \mathbf{q}^r, z)] dz, \end{aligned} \quad (10)$$

where  $H(\mathbf{q}, z)$  is given by

$$\begin{aligned} H(\mathbf{q}, \mathbf{q}^r, \mathbf{q}^i, z) = & \frac{i}{2\pi} k_0^2 A^r A^i \hat{e}_\beta^{r*} \hat{e}_\alpha^i \frac{h_{\alpha\beta}(\mathbf{q}^r - \mathbf{q}, z)^r}{k_z(\mathbf{q}^r - \mathbf{q})} \\ & \times \exp\{i[k_z(\mathbf{q}^r - \mathbf{q}) - k_z^*(\mathbf{q}^r)]z_d \\ & + i[k_z(\mathbf{q}^i) - k_z(\mathbf{q}^r - \mathbf{q})]z\}, \end{aligned} \quad (11)$$

where the tensor  $h$  is given in Appendix A.

Equation (10) is our forward model for the data and provides a basis for the solution of the inverse problem (i.e., the determination of  $\eta$ ). The inverse problem is overdetermined in the general case ( $\mathbf{q}^i \neq \mathbf{q}^r$ ), as might be expected from the fact that data are collected in six dimensions and the unknown ( $\eta$ ) is a function of only three variables.

It is worthwhile to discuss the way in which one might expect the holographic intensity to change with the transverse spatial frequencies  $\mathbf{q}$ ,  $\mathbf{q}^i$ , and  $\mathbf{q}^r$ . Inspecting Eq. (11), one can see several regimes in which the kernel function  $H$  decays exponentially with  $z$  or  $z'$  and one regime in which it does not. These regimes may be classified as follows:

1.  $|\mathbf{q}^i| \leq k_0$ ,  $|\mathbf{q}^r| \leq k_0$ , and  $|\mathbf{q}^r - \mathbf{q}| \leq k_0$ : there is no exponential decay. A necessary condition for this is  $|\mathbf{q}| < 2k_0$ . This is the regime of standard (far-field) diffraction tomography, where the reconstructions are bandlimited to the Ewald limiting sphere of radius  $2k_0$  [3].
2.  $|\mathbf{q}^i| \leq k_0$ ,  $\mathbf{q}^r = \mathbf{q}^i$ , and  $|\mathbf{q}^i - \mathbf{q}| > k_0$ : there is one factor of exponential decay. This regime is accessible in scattering NSOM.
3.  $k_0 < |\mathbf{q}^i| \leq nk_0$ ,  $|\mathbf{q}^r| \leq k_0$ , and  $|\mathbf{q}^r - \mathbf{q}| \leq k_0$ : the incident field decays exponentially, but the reference and scattered waves propagate to the far field. This regime is accessible in TIRT, a variant of total internal reflection microscopy [11,13].
4.  $k_0 < |\mathbf{q}^i| \leq nk_0$ ,  $\mathbf{q}^r = \mathbf{q}^i$ , and  $|\mathbf{q}^i - \mathbf{q}| \leq k_0$ : there is one factor of exponential decay. A necessary condition for this is  $|\mathbf{q}| < (n+1)k_0$ . This condition sets the limit for the  $(n+1)k_0$  cutoff observed in paper I. This regime is again accessible in TIRT.

5.  $k_0 < |\mathbf{q}^i| \leq nk_0$ ,  $\mathbf{q}^r = \mathbf{q}^i$ , and  $|\mathbf{q}^i - \mathbf{q}| > k_0$ : there are two factors of exponential decay. This regime is accessible in PSTM.

The general case where  $\mathbf{q}^i \neq \mathbf{q}^r$  might be thought of as a near-field form of Leith–Upatnieks holography where the reference field is not identical with the illumination and so avoids the so-called conjugate image problem. The restricted case that  $\mathbf{q}^i = \mathbf{q}^r$  is analogous to Gabor holography, in which the incident and reference fields are identical. For scattering NSOM and PSTM, the Gabor-type hologram is obtained by a natural extension of the standard instrument. We consider the restricted case for the rest of the paper.

The susceptibility  $\boldsymbol{\eta}(\mathbf{r})$  is to be estimated from a set of  $P$  noisy measurements. Paper I dealt with the  $P = 2$  case under the assumption that  $\boldsymbol{\eta}$  has no structure in the  $z$  direction apart from finite support. For the purpose of this paper, the  $j$ th measurement is defined to be a set of  $M^2$  experimental determinations of total intensity  $\{I_{j,mn}^{\text{total}}, 0 \leq m, n \leq M-1\}$  at transverse coordinates  $\rho_{mn}$ . The entire dataset consists of  $P$  such measurements with distinct values of the experimental parameters that determine the properties of the incident and reference beams. Each value of the index  $j$  corresponds to a different pair of incident and reference wave vectors  $\mathbf{q}_j^r = \mathbf{q}_j^i$  and their polarization vectors  $\hat{\mathbf{e}}_j^r = \hat{\mathbf{e}}_j^i$ . If two (or more) measurements are made with identical experimental parameters, they may be averaged and treated as a single measurement (with the measurement uncertainty adjusted accordingly).

It is useful to define the measurement operator  $\mathbf{H}$ , which maps  $\mathcal{L}_2(0, \Delta z) \times \mathcal{L}_2(0, \Delta z)$  to  $\mathcal{C}^P$  [where  $\mathcal{L}_2(0, \Delta z)$  is the space of square integrable functions on the interval  $(0, \Delta z)$  and  $\mathcal{C}^P$  is the space of complex vectors with  $P$  components] via its kernel:

$$\mathbf{H}(\mathbf{q}, z) \equiv \begin{bmatrix} H(\mathbf{q}, \mathbf{q}_0^i, \mathbf{q}_0^i, z) & H^*(-\mathbf{q}, \mathbf{q}_0^i, \mathbf{q}_0^i, z) \\ \vdots & \vdots \\ H(\mathbf{q}, \mathbf{q}_{p-1}^i, \mathbf{q}_{p-1}^i, z) & H^*(-\mathbf{q}, \mathbf{q}_{p-1}^i, \mathbf{q}_{p-1}^i, z) \end{bmatrix}, \quad (12)$$

as well as the vector of holographic intensities:

$$\tilde{\mathbf{I}}(\mathbf{q}) \equiv \begin{pmatrix} \tilde{I}_0(\mathbf{q}) \\ \vdots \\ \tilde{I}_{p-1}(\mathbf{q}) \end{pmatrix}, \quad (13)$$

and the two-component vector containing the pair:

$$\tilde{\boldsymbol{\eta}}(\mathbf{q}, z) \equiv \begin{pmatrix} \tilde{\eta}(\mathbf{q}, z) \\ \tilde{\eta}^*(-\mathbf{q}, z) \end{pmatrix}. \quad (14)$$

With this notation the inner product is denoted

$$\tilde{\boldsymbol{\eta}}^H(\mathbf{q}, z) \tilde{\boldsymbol{\eta}}(\mathbf{q}, z) = \int_0^{\Delta z} dz \{ |\tilde{\eta}(\mathbf{q}, z)|^2 + |\tilde{\eta}(-\mathbf{q}, z)|^2 \}, \quad (15)$$

and the  $H$  superscript indicates the Hermitian adjoint. Similarly,  $\tilde{\boldsymbol{\eta}}(\mathbf{q}, z_1) \tilde{\boldsymbol{\eta}}^H(\mathbf{q}, z_2)$  denotes an outer product, which is a  $2 \times 2$  matrix of functions on  $\mathcal{L}_2(0, \Delta z) \times \mathcal{L}_2(0, \Delta z)$ .

Equation (10) refers to only one measurement. Making use of the definitions in Eqs. (12)–(14), Eq. (10) generalizes to the following system of equations:

$$\tilde{\mathbf{I}}(\mathbf{q}) = [\mathbf{H}\tilde{\boldsymbol{\eta}}](\mathbf{q}) = \int_0^{\Delta z} dz \mathbf{H}(\mathbf{q}, z) \tilde{\boldsymbol{\eta}}(\mathbf{q}, z), \quad (16)$$

where the integration over  $z$  is implied in the operator notation. Note that  $\mathbf{H}(\mathbf{q})$  in Eq. (16) is an operator and  $\mathbf{H}(\mathbf{q}, z)$  in Eq. (12) is the kernel of that operator. As with  $I^{\text{total}}$  in Eq. (7),  $\tilde{\mathbf{I}}(\mathbf{q})$  is noise free, deterministic, and not known.

Without further constraints on the solutions, the measurement operator  $\mathbf{H}$  is not invertible because the problem is underdetermined for the case of finite data. Under conditions that allow a nonsingular representation of  $\mathbf{H}$ , it resembles the Laplace transform, and its poor conditioning makes the solution unstable to noise. The inversion algorithm must be regularized to find an approximate, smooth, and stable solution [28].

Because  $\mathbf{H}$  is a singular linear operator, there exists a family of solutions to Eq. (16). The solution of minimum  $\mathcal{L}_2$  norm is denoted by  $\tilde{\boldsymbol{\eta}}^+$ . Because it is the minimum-norm solution,  $\tilde{\boldsymbol{\eta}}^+$  lies in the orthogonal complement to the null space of  $\mathbf{H}$  [29], and  $\tilde{\boldsymbol{\eta}}^+$  satisfies Eq. (16):

$$\tilde{\mathbf{I}}(\mathbf{q}) = \mathbf{H}(\mathbf{q}) \tilde{\boldsymbol{\eta}}^+(\mathbf{q}). \quad (17)$$

### 3. STATISTICAL MODEL

An analytical forward model for the holographic intensity is given above. Real data are, of course, obtained on a discrete grid and are generally noisy. Below, the problem is recast in a discrete form and data are treated as random variables. A particular noise model is introduced, and the inverse problem is solved by employing regularization.

#### A. Discretization

The intensity in a given plane outside the sample is effectively bandlimited and therefore allows a representation on a discrete grid in coordinate space. It is assumed that the detector is of area  $L \times L$ , lies in the plane  $z = z_d$ , and that it is sampled at  $M \times M$  points, loosely referred to as “pixels,” that give measurements of the total intensity at points  $\{\rho_{mn}\}$ .  $M$  is assumed to be an odd integer, so that a symmetric spatial-frequency grid is obtained. It is also assumed that the detector has sufficient surface area so that the effects of truncation of the Fourier transform are negligible. This model leads to the following discretization of the spatial-frequency variable  $\mathbf{q}$ . The discrete spatial frequency is defined as  $\mathbf{q}_{mn} \equiv (\beta_m, \beta_n)$ , where  $\beta_n$  is

$$\beta_n = \frac{2\pi}{L} n, \quad 0 \leq n \leq \frac{M-1}{2}, \\ = \frac{2\pi}{L} [n - M], \quad \frac{M-1}{2} + 1 \leq n \leq M-1, \quad (18)$$

and  $\mathbf{q}_{M-m, M-n} = -\mathbf{q}_{mn}$ .

#### B. Photocount Statistics

In this discussion, the semiclassical model for photoelectric detection is used, i.e., the electromagnetic field incident on the detector is treated classically, but the detection is quantized in units of photoevents (photocounts) [30]. The photocount statistics are taken to be photon limited, for which the noise at each detector pixel is statistically independent of the others and follows a Poisson distribution. Under the

weak hologram approximation, however, the photon noise will be dominated by that due to the reference beam and, hence, will be independent of the holographic intensity (i.e., signal independent). In addition, for sufficiently high count rates, the noise can be modeled as approximately Gaussian, of the same mean and variance. Consequently, the noise at each pixel can be treated as additive Gaussian noise.

The number of photons counted in the  $mn$ th pixel for the  $j$ th measurement is represented by the random variable  $N_{j,mn}$ . The expected number of photons is  $\overline{N_{j,mn}} = W I_{j,mn}^{\text{total}}$ , where  $W = \Delta a \Delta t / \hbar c k_0$ , and the overbar denotes the expectation operator;  $\Delta a$ ,  $\Delta t$ ,  $c$ , and  $\hbar$  are the effective pixel area (including the quantum efficiency), the measurement time, the speed of light, and Planck's constant divided by  $2\pi$ , respectively. The expected number of reference beam photocounts (or incident beam photocounts, because they are same for PSTM) is represented by  $N_{0,j} = W (A_j^r)^2 \exp\{i[k_z(\mathbf{q}_j^r) - k_z^*(\mathbf{q}_j^r)]z_d\}$ . We see that the reference (incident) beam amplitude decreases exponentially with  $z_d$  by an amount that depends on  $k_z(\mathbf{q}_j^i)$ ; hence, in photon-limited imaging, the noise depends on  $\mathbf{q}_j^i$ , and so the noise variance is dependent on the measurement number. It follows that the random photocount variable  $c_{j,mn} \equiv N_{j,mn} - N_{0,j}$  represents a measure of the holographic intensity. As expected, for zero-mean noise,  $\overline{c_{j,mn}} = W I_{j,mn}$ .

Let  $\mathbf{c}_j = \{c_{j,mn}, 0 \leq m, n \leq M-1\}$  represent the holographic photocount vector corresponding to the  $j$ th measurement and  $\tilde{\mathbf{c}}_j = \{\tilde{c}_{j,mn}, 0 \leq m, n \leq M-1\}$  its 2D discrete Fourier transform (DFT), where

$$\tilde{c}_{j,mn} = \frac{1}{M^2} \sum_{k=0}^{M-1} \sum_{l=0}^{M-1} c_{j,kl} \exp\left\{2\pi i \left[\frac{km}{M} + \frac{ln}{M}\right]\right\}. \quad (19)$$

The joint probability density function (PDF) of the 2D DFT of the holographic photocount vector is given by

$$\begin{aligned} \ln \mathcal{P}_{\tilde{\mathbf{c}}_j}(\tilde{\mathbf{c}}_j) &= \frac{-M^2}{2\sigma_j^2} \left[ (\tilde{c}_{j,00} - \overline{\tilde{c}_{j,00}})(\tilde{c}_{j,00}^* - \overline{\tilde{c}_{j,00}^*}) \right. \\ &+ 2 \sum_{m=1}^{\frac{M-1}{2}} [(\tilde{c}_{j,0m} - \overline{\tilde{c}_{j,0m}})(\tilde{c}_{j,0m}^* - \overline{\tilde{c}_{j,0m}^*}) \\ &+ (\tilde{c}_{j,m0} - \overline{\tilde{c}_{j,m0}})(\tilde{c}_{j,m0}^* - \overline{\tilde{c}_{j,m0}^*})] \\ &\left. + 2 \sum_{n=1}^{M-1} \sum_{m=1}^{\frac{M-1}{2}} (\tilde{c}_{j,mn} - \overline{\tilde{c}_{j,mn}})(\tilde{c}_{j,mn}^* - \overline{\tilde{c}_{j,mn}^*}) \right] + \text{const.}, \end{aligned} \quad (20)$$

where  $\sigma_j^2$  is the variance of  $c_{j,mn}$ . It can be seen from Eq. (20) that  $\tilde{\mathbf{c}}_j$  is an example of a so-called circular (or proper) Gaussian random vector. For a circular random variable  $\mathbf{x}$ , the real and imaginary parts of the vector ( $\mathbf{x} - \bar{\mathbf{x}}$ ) are independent and of the same variance, and the joint PDF has the form

$$\mathcal{P}_{\mathbf{x}}(\mathbf{x}) = \frac{1}{(2\pi)^m |\mathbf{C}|} \exp\left[-\frac{1}{2}(\mathbf{x} - \bar{\mathbf{x}})^H \mathbf{C}^{-1}(\mathbf{x} - \bar{\mathbf{x}})\right], \quad (21)$$

where  $\mathbf{C}^{-1}$  is Hermitian symmetric [31]. In this case, the covariance is given by

$$\overline{(\mathbf{x} - \bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})^H} = 2\mathbf{C}. \quad (22)$$

Because the  $P$  measurements are statistically independent, the joint PDF of all of the holographic intensity data takes a particularly simple form:

$$\mathcal{P}_{\text{total}}(\tilde{\mathbf{c}}_0, \dots, \tilde{\mathbf{c}}_{P-1}) = \prod_{j=0}^{P-1} \mathcal{P}_{\tilde{\mathbf{c}}_j}(\tilde{\mathbf{c}}_j). \quad (23)$$

The PDF in Eq. (23) can be simplified if we redefine the holographic photocount vector as  $\tilde{\mathbf{c}} = \{\tilde{c}_{mn}, 0 \leq m, n \leq M-1\}$ , where  $\tilde{\mathbf{c}}_{mn} \equiv \{\tilde{c}_{j,mn}, 0 \leq j \leq P-1\}$ . Upon substituting Eq. (20) into Eq. (23) and taking the logarithm, we find that

$$\ln \mathcal{P}_{\text{total}}(\tilde{\mathbf{c}}_0, \dots, \tilde{\mathbf{c}}_{P-1}) = \sum_{m=0}^{M-1} \sum_{n=0}^{M-1} \ln \mathcal{P}_{mn}(\tilde{\mathbf{c}}_{mn}), \quad (24)$$

where

$$\ln \mathcal{P}_{mn}(\tilde{\mathbf{c}}_{mn}) = -(\tilde{\mathbf{c}}_{mn} - \overline{\tilde{\mathbf{c}}_{mn}})^H \Sigma^H \Sigma (\tilde{\mathbf{c}}_{mn} - \overline{\tilde{\mathbf{c}}_{mn}}) + \text{const.}, \quad (25)$$

where the  $P \times P$ , real, diagonal matrix  $\Sigma$  is defined by

$$\Sigma \equiv M \begin{bmatrix} 1/\sigma_0 & & & \\ & \ddots & & \\ & & \ddots & \\ & & & 1/\sigma_{P-1} \end{bmatrix}. \quad (26)$$

Under this arrangement, the  $PM^2 \times PM^2$  noise covariance matrix for the holographic intensity measurements (in the DFT domain) is given by placing  $P$  copies of  $\Sigma^{-2}$  along the diagonal [see Eq. (22)], i.e.,

$$\overline{(\tilde{\mathbf{c}} - \bar{\tilde{\mathbf{c}}})(\tilde{\mathbf{c}} - \bar{\tilde{\mathbf{c}}})^H} = \begin{bmatrix} \Sigma^{-2} & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \Sigma^{-2} \end{bmatrix}. \quad (27)$$

Because  $\tilde{\mathbf{c}}$  is circular,  $\overline{(\tilde{\mathbf{c}} - \bar{\tilde{\mathbf{c}}})(\tilde{\mathbf{c}} - \bar{\tilde{\mathbf{c}}})^T} = \mathbf{0}$ .

### C. Maximum-Likelihood and Maximum *a Posteriori* Estimates

For each spatial frequency  $\mathbf{q}_{mn}$ , Eq. (16) takes the form

$$\tilde{\mathbf{c}}_{mn} = L^{-2} \mathbf{W} \mathbf{H}_{mn} \tilde{\eta}_{mn}, \quad (28)$$

where  $\mathbf{H}_{mn} \equiv \mathbf{H}(\mathbf{q}_{mn})$  and  $\tilde{\eta}_{mn} \equiv \tilde{\eta}(\mathbf{q}_{mn})$ . It should be noted that the factor  $L^{-2}$  in Eq. (28) arises due to the fact that  $\tilde{c}_{mn}$  is the DFT of  $c_{kl}$ , whereas  $\tilde{\eta}_{mn}$  are sample values of the continuous Fourier transform  $\tilde{\eta}$ . Using Eq. (28), Eq. (25) takes the form

$$\begin{aligned} \ln \mathcal{P}_{mn}(\tilde{\mathbf{c}}_{mn}) &= -(\tilde{\mathbf{c}}_{mn} - L^{-2} \mathbf{W} \mathbf{H}_{mn} \tilde{\eta}_{mn})^H \Sigma^H \Sigma \\ &\times (\tilde{\mathbf{c}}_{mn} - L^{-2} \mathbf{W} \mathbf{H}_{mn} \tilde{\eta}_{mn}) + \text{const.} \end{aligned} \quad (29)$$

The value  $\tilde{\eta}$  must be estimated from the measurements, and the PDF in Eq. (29) allows determination of the best possible estimate given certain criteria. In this situation (Gaussian PDF and no meaningful prior distribution for the unknown object), the maximum-likelihood estimate (MLE) is a popular choice, and it is also the least-squares solution [29]. Because  $\mathbf{H}$  is a singular operator, the inversion is underdetermined and the MLE is nonunique. The MLE is determined by maximizing  $\ln \mathcal{P}_{mn}(\tilde{\mathbf{c}})$  with respect to  $\tilde{\eta}$ , i.e., by solving the equation

$$\frac{\partial \ln \mathcal{P}_{mn}(\tilde{\mathbf{c}})}{\partial \tilde{\boldsymbol{\eta}}} = 0. \quad (30)$$

At this point it is useful to define the weighted measurement operator  $\mathbf{K} \equiv \Sigma \mathbf{H}$ , so the MLE,  $\hat{\boldsymbol{\eta}}$ , satisfies the algebraic equation

$$L^{-2} \mathbf{W} \mathbf{K}^H \mathbf{K} \hat{\boldsymbol{\eta}} = \mathbf{K}^H \Sigma \tilde{\mathbf{c}}. \quad (31)$$

As mentioned above, Eq. (31) does not uniquely specify  $\hat{\boldsymbol{\eta}}$  due to the singular nature of  $\mathbf{K}$  (which has the same null space as  $\mathbf{H}$ ). However, the minimum-norm solution to Eq. (31), known as the maximum-likelihood minimum-norm (MLMN) estimate, is unique, and it is given by  $\hat{\boldsymbol{\eta}}^+ = L^2 \mathbf{W}^{-1} \mathbf{K}^+ \Sigma \tilde{\mathbf{c}}$ , where  $\mathbf{K}^+$  represents the pseudoinverse of  $\mathbf{K}$ . The pseudoinverse may be calculated via the singular value decomposition (SVD) of the forward operator  $\mathbf{K}$  (see Appendix A) and does not depend on the existence of  $(\mathbf{K}^H \mathbf{K})^{-1}$ . It has the form

$$\mathbf{K}^+ = \sum_{k=0}^{k_r-1} \frac{1}{s_k} \mathbf{f}_k(\mathbf{z}) \mathbf{g}_k^H, \quad (32)$$

where  $k_r = \text{rank}(\mathbf{K}) \leq P$  is the number of singular values of  $\mathbf{K}$ . The number of singular values represents the number of degrees of freedom in the calculation of the estimate  $\hat{\boldsymbol{\eta}}^+$  and is a direct measure of the information content of the scattered field in the near zone. After multiplying both sides of Eq. (28) by  $\Sigma$  and taking the minimum-norm solution [see also Eq. (17)], we find that  $\tilde{\boldsymbol{\eta}}^+$  is given by

$$\tilde{\boldsymbol{\eta}}^+ = \frac{L^2}{\mathbf{W}} \mathbf{K}^+ \Sigma \tilde{\mathbf{c}}. \quad (33)$$

In some situations, the PDF of  $\tilde{\boldsymbol{\eta}}$  may be known *a priori*, meaning that even in the absence of data,  $\tilde{\boldsymbol{\eta}}$  is more likely to assume some values than others. In this case the maximum *a posteriori* estimate is preferred to the ML or MLMN estimates [29,32]. Assuming a circular Gaussian PDF, Eq. (29) generalizes to

$$\begin{aligned} \ln \mathcal{P}_{mn}(\tilde{\mathbf{c}}) = & -(\Sigma \tilde{\mathbf{c}} - L^{-2} \mathbf{W} \mathbf{K} \tilde{\boldsymbol{\eta}})^H (\Sigma \tilde{\mathbf{c}} - L^{-2} \mathbf{W} \mathbf{K} \tilde{\boldsymbol{\eta}}) \\ & - \frac{1}{2} (\tilde{\boldsymbol{\eta}} - \tilde{\boldsymbol{\eta}}_0)^H \mathbf{P} (\tilde{\boldsymbol{\eta}} - \tilde{\boldsymbol{\eta}}_0) + \text{const.}, \end{aligned} \quad (34)$$

where  $\tilde{\boldsymbol{\eta}}_0$  is the mean and  $\mathbf{P}$  is Hermitian. If  $\mathbf{P}$  is invertible (it need not be, see below), it is the inverse of the covariance. Analogous to the development of Eq. (31), one obtains the matrix equation:

$$(L^{-2} \mathbf{W} \mathbf{K}^H \mathbf{K} - \frac{1}{2} \mathbf{P}) \hat{\boldsymbol{\eta}} = \mathbf{K}^H \Sigma \tilde{\mathbf{c}} - \frac{1}{2} \mathbf{P} \tilde{\boldsymbol{\eta}}_0. \quad (35)$$

For many choices of  $\mathbf{P}$ ,  $(L^{-2} \mathbf{W} \mathbf{K}^H \mathbf{K} - \frac{1}{2} \mathbf{P})$  is not a singular matrix (otherwise  $\mathbf{P}$  and  $\mathbf{K}^H \mathbf{K}$  share some of the same null space) and, therefore, it has an inverse.

#### D. Regularized Estimate

Although the pseudoinverse yields the MLMN estimate, this estimate may be a poor representation of the unknown object because the contribution of the smallest singular values in Eq. (32) will greatly amplify noise in the data. The solution may be stabilized (or regularized) by truncating the sum in

Eq. (32), so that only those terms corresponding to values of  $s_k$  larger than some minimum value (the noise level, for instance) are included [28,32,33]. In this case, the regularized (truncated SVD) estimate is given by

$$\hat{\boldsymbol{\eta}}_{\text{reg}} = \frac{L^2}{\mathbf{W}} \mathbf{K}_{\text{reg}}^+ \Sigma \tilde{\mathbf{c}} = \sum_{k=0}^{k_t-1} d_k \mathbf{f}_k(\mathbf{z}), \quad (36)$$

where  $k_t$  is the threshold index and  $d_k = W^{-1} L^2 s_k^{-1} \mathbf{g}_k^H \Sigma \tilde{\mathbf{c}}$ . Because smaller singular values tend to be associated with singular functions that are more oscillatory, truncation in this way tends to make the solution smoother. In the presence of noise, then,  $k_t$  is a measure of the effective information content of the scattered field, because this represents the amount of scattered field data used in the estimation. As an alternative to truncation, the smaller (more sensitive) singular values can be filtered gradually by replacing  $1/s_k$  in Eq. (32) with  $s_k/(s_k^2 + \Gamma)$ , where  $\Gamma$  is a constant.

Incorporating *a priori* PDF information, as in Eq. (34), is effectively a form of regularization, as constraining the values that  $\tilde{\boldsymbol{\eta}}$  can take generally tends to stabilize the inversion procedure. One can also achieve some measure of stability by artificially introducing a term (hereafter referred to as a regularization term) of similar construct to the PDF term in Eq. (34). In this case, one chooses the matrix  $\mathbf{P}$  to constrain some aspect of the object estimate. For instance, one may choose  $\mathbf{P}$  to penalize gradients in the solution (e.g., Ref. [32]). Regardless of its form, the weighting assigned to the regularization term is critically important. The weighting essentially determines the extent to which the constraint is imposed at the expense of the least-squares fit of the forward model to the data. Excessive constraint weighting, for example, can lead to a poor (yet stable) estimate of the object. The constraint weighting is often chosen using a signal-to-noise criterion; specifically, because the data should not be fit with an accuracy greater than that afforded by the noise level, the fit error (or residual) can be assigned to the regularization term. This is implemented by multiplying the regularization term by a weighting (or regularization) parameter  $\gamma$ , which can be adjusted accordingly. As an example, consider the case for which  $\mathbf{P}$  is proportional to the identity matrix (i.e., Tikhonov regularization) and  $\tilde{\boldsymbol{\eta}}$  has a zero mean. It follows that the regularization term has the form  $\gamma \tilde{\boldsymbol{\eta}}^H \tilde{\boldsymbol{\eta}}$ , where the solution to Eq. (35) reduces to the unregularized pseudoinverse in the limit  $\gamma \rightarrow 0$  [33].

#### E. Statistics of MLMN and Regularized Estimates

Let us now calculate the mean and covariance of the estimate for the susceptibility. This analysis applies equally well to the regularized estimate  $\hat{\boldsymbol{\eta}}_{\text{reg}}$  and the MLMN estimate  $\hat{\boldsymbol{\eta}}^+$  when the regularization is done by SVD truncation, because  $\hat{\boldsymbol{\eta}}_{\text{reg}} = \hat{\boldsymbol{\eta}}^+$  when one includes all  $k_r$  terms in Eq. (36). Because the MLMN estimate  $\hat{\boldsymbol{\eta}}^+$  is a linear combination of the data  $\tilde{\mathbf{c}}$  (which have a Gaussian distribution), its PDF will also be Gaussian. Taking the expected value of Eq. (33), we find that  $\overline{\hat{\boldsymbol{\eta}}^+} = \tilde{\boldsymbol{\eta}}^+$ . From Eqs. (32) and (33) and the relation  $\overline{(\tilde{\mathbf{c}} - \tilde{\mathbf{c}})(\tilde{\mathbf{c}} - \tilde{\mathbf{c}})^H} = \Sigma^{-2}$ ,

$$\overline{(\hat{\boldsymbol{\eta}}^+ - \tilde{\boldsymbol{\eta}}^+)(\hat{\boldsymbol{\eta}}^+ - \tilde{\boldsymbol{\eta}}^+)^H} = \frac{L^4}{\mathbf{W}^2} \mathbf{K}^+ \mathbf{K}^{+H} = \frac{L^4}{\mathbf{W}^2} \sum_{k=0}^{k_r-1} \frac{1}{s_k^2} \mathbf{f}_k(\mathbf{z}_1) \mathbf{f}_k^H(\mathbf{z}_2), \quad (37)$$

where  $k_r$  is the number of nonzero eigenvalues of  $\mathbf{K}\mathbf{K}^H$ , and  $z_1$  and  $z_2$  may assume any value (independently) between 0 and the thickness  $\Delta z$ .

An important measure of the reconstruction is the projected square error (PSE), the integral square difference between  $\hat{\eta}^+$  and  $\tilde{\eta}^+$ :

$$\text{PSE} \equiv \overline{(\hat{\eta}^+ - \tilde{\eta}^+)^H (\hat{\eta}^+ - \tilde{\eta}^+)} = \frac{L^4}{W^2} \sum_{k=0}^{k_r-1} \frac{1}{s_k^2}. \quad (38)$$

This is a slight generalization of the well-known fact that the trace of a matrix is equal to the sum of its eigenvalues [34].

A related quantity is the total square error (TSE),  $\overline{(\tilde{\eta} - \hat{\eta}^+)^H (\tilde{\eta} - \hat{\eta}^+)}$ , which is the square difference between the MLMN estimate and the true susceptibility. This quantity is a measure of the accuracy to which the susceptibility can ultimately be determined and, as such, is an indicator of information content itself. While it cannot be calculated without knowledge of the true object  $\tilde{\eta}$  (by definition), the TSE is shown below to be comprised of the PSE and another term. Expanding the general expression for the TSE, one obtains

$$\text{TSE} = \overline{\tilde{\eta}^H \tilde{\eta} - 2\tilde{\eta}^H \hat{\eta}^+ + \hat{\eta}^+ H \hat{\eta}^+} = [(\tilde{\eta} - \hat{\eta}^+)^H (\tilde{\eta} - \hat{\eta}^+)] + \text{PSE}, \quad (39)$$

where the relations  $\overline{\tilde{\eta}^+} = \tilde{\eta}^+$  and  $\overline{(\hat{\eta}^+ - \tilde{\eta}^+)^H (\hat{\eta}^+ - \tilde{\eta}^+)} = \overline{\hat{\eta}^+ H \hat{\eta}^+ - \tilde{\eta}^+ H \hat{\eta}^+}$  have been used. Equation (39) has a simple interpretation: the first term is the error due to projection onto the range space of  $\mathbf{K}^+$ , and the second term is the PSE discussed above.

For the regularized case, consider the values of the PSE and TSE as a function of the cutoff index  $k_t$  in Eq. (36). It can be seen from Eq. (38) that the PSE must increase monotonically with  $k_t$ . However, the first term in Eq. (39) monotonically decreases with  $k_t$ , because the dimension of the range space of  $\mathbf{K}_{\text{reg}}^+$  increases with  $k_t$ . It follows that there is an optimal index  $k_t \leq k_r$  for which the TSE is a minimum.

It is helpful to expand the estimate error in the eigenfunctions of  $\mathbf{K}^H \mathbf{K}$ :

$$(\hat{\eta}^+ - \tilde{\eta}^+) = \sum_{l=0}^{k_r-1} (d_l - \bar{d}_l) f_k(z), \quad (40)$$

where  $d_l$  and  $\bar{d}_l$  are the expansion coefficients of the estimate and the object, respectively. Taking advantage of the orthogonality of the  $f_k(z)$  and substituting back, it is found that

$$\overline{(d_r - \bar{d}_r)(d_q - \bar{d}_q)} = \frac{L^4}{W^2} \frac{1}{s_q^2} \delta_{qr}, \quad (41)$$

where  $\delta_{qr}$  is the Kronecker delta. Some comment on this result is appropriate. Equation (41) gives the variance of the estimate in terms of the singular values of the weighted measurement operator. This result will be used below to provide a criterion for regularization. The  $(d_q - \bar{d}_q)$  are also circular random variables and statistically independent. It may now be seen that the variance of an expansion coefficient is inversely proportional to  $s_q^2$ . Because the  $d_q$  are circular, the variance of the real and imaginary parts are equal and each contribute  $L^4/(2W^2 s_q^2)$  to the total.

A semianalytic solution of the inverse problem for discrete, noisy data is given above. Error estimates and variances for

the reconstructed object function are shown. The implications of the results in terms of the number of degrees of freedom, i.e., the information content, of the data as well as the impact of various system parameters are explored numerically in the next section.

## 4. RESULTS AND CONCLUSIONS

The data available in a near-field scattering experiment were investigated by examining the form of the SVD of the weighted measurement operator  $\mathbf{K} = \Sigma \mathbf{H}$ . This allows determination of the singular value spectrum (and hence the number of degrees of freedom of the scattered field) at a particular transverse spatial frequency. In practice, the number of degrees of freedom is effectively smaller, as measurement noise limits the amount of scattered field data that can be reliably incorporated into the estimation algorithm and necessitates the regularization of the pseudoinverse by truncation of the SVD or other means, as discussed previously. The effects of regularization were also investigated by calculating the point response of the imaging system under different SVD truncation criteria.

Numerical simulations were performed for several measurement configurations, labeled A1, A2, B1, and B2, which are listed in Table 1. These configurations differ only in estimated sample thickness ( $\Delta z$ ) and measurement plane height ( $z_d$ ). Other common parameters on which the simulations depend include the incident wavelength, the number of distinct incident waves ( $P$ ), their transverse wave vectors and polarizations ( $\{\mathbf{q}_j^i\}$  and  $\{\mathbf{e}_j^i\}$ ), and the prism index of refraction ( $n$ ). For this study, the wavelength and prism index of refraction were taken to be  $\lambda = 1 \mu\text{m}$  and  $n = 2.0$ , respectively.

A uniform distribution of  $P = 151$  incident wave vectors was chosen, subject to the condition  $|\mathbf{q}^i| \leq nk_0$ . The inequality  $|\mathbf{q}^i| \leq nk_0$  represents a fundamental restriction of PSTM. In practice, the wave vector spectrum will be further limited by the experimental apparatus. Computations were made for both transverse magnetic (TM) and transverse electric (TE) incident polarizations. Unlike the results presented in paper I, however, the results for the TM and TE polarizations are quite similar. This is attributable to the high degree of symmetry in the present case, as opposed to the case in paper I, in which only two incident beams were considered. Consequently, only the TM results are shown here.

Figures 1 through 4 refer to cases A1 and A2 in Table 1. Figures 1(a) and 1(c) show plots, for cases A1 and A2, respectively, of the singular value spectrum of the weighted measurement operator  $\mathbf{K}$  as a function of index  $j$  for three transverse spatial frequencies,  $|\mathbf{q}| = 0.5k_0$ ,  $3k_0$  and  $9k_0$ . It can be seen from the plots that the singular value spectrum, and consequently the axial information content, is similar for the two smallest spatial frequencies as the distance to the measurement plane is increased, whereas the spectrum for

**Table 1. Parameters for the Numerical Examples Discussed in the Text<sup>a</sup>**

Case	Sample Thickness ( $\Delta z/\lambda$ )	Detector Height ( $z_d/\lambda$ )
A1	0.5	0.51
A2	0.5	1.0
B1	0.05	0.051
B2	0.05	1.0

<sup>a</sup>Note that the sample lies in  $0 \leq z \leq \Delta z$ , and the detection of the field is in the plane  $z = z_d$ .

the largest spatial frequency falls off progressively faster with increasing distance (case A2). This behavior is expected because the two smallest spatial frequencies correspond to propagating homogeneous waves.

Figures 1(b) and 1(d) show plots, for cases A1 and A2, respectively, of the log of the first-, third-, and ninth-largest singular values of the weighted measurement operator  $\mathbf{K}$  as a function of the magnitude of the transverse spatial frequency  $|\mathbf{q}|$ . Consistent with Fig. 1(a), we see that the greatest bandwidth is associated with case A1, and narrows with increasing measurement distance  $z_d$  (case A2). In addition, in case A1, due to the close proximity of the detection plane to  $z = \Delta z$ , there are one or two large singular values even at very large spatial frequencies, well beyond the plot range. This is because homogeneous or weakly evanescent waves may be converted to evanescent waves near  $z = \Delta z$  and be seen at  $z = z_d$ . In case A2, most of these waves are not seen because the measurement plane lies at a greater distance from the sample.

Figure 2 shows plots of the absolute values of the singular functions  $|f^+(z)|$  for the first-, third-, and ninth-largest singular values of the weighted measurement operator  $\mathbf{K}$  for case A1 and spatial frequencies (a)  $|\mathbf{q}| = 0.5k_0$  and (b)  $|\mathbf{q}| = 9k_0$ , and case A2 and spatial frequencies (c)  $|\mathbf{q}| = 0.5k_0$  and (d)  $|\mathbf{q}| = 9k_0$ . At low spatial frequencies (corresponding to homogeneous waves), the sample volume is interrogated fairly evenly, while at higher frequencies, the back face is preferentially interrogated. Also, it is seen that the singular functions do not change quickly with  $z_d$ .

Figures 3 and 4 illustrate the point response for cases A1 and A2. In both cases, the point response was calculated for three different axial locations of the source: the front face ( $z = 0$ ), the middle ( $z = \Delta z/2 = 0.25\lambda$ ), and the back face ( $z = \Delta z = 0.5\lambda$ ). The point scatterer was modeled as a sphere of radius  $a = 0.01\Delta z$  and dielectric susceptibility  $\eta_s = 1.0$ . Its transverse Fourier transform is given by

$$\tilde{\eta}(\mathbf{q}, z) = \eta_s \frac{\sqrt{2za - z^2} J_1(|\mathbf{q}| \sqrt{2za - z^2})}{|\mathbf{q}|}, \quad (42)$$

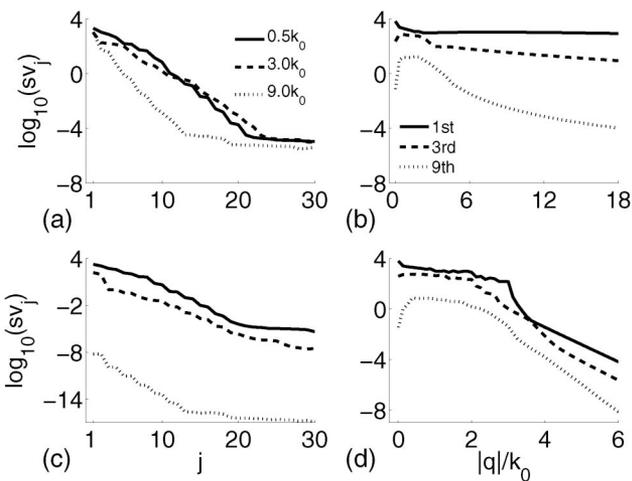


Fig. 1. Log of the singular values of the weighted measurement operator  $\mathbf{K}$  as a function of index  $j$  for cases (a) A1 and (c) A2 is shown. In both cases, the singular value spectra are plotted for three spatial frequencies,  $|\mathbf{q}| = 0.5k_0$ ,  $3k_0$ , and  $9k_0$ . The log of the first-, third-, and ninth-largest singular values of the weighted measurement operator  $\mathbf{K}$  as a function of spatial frequency  $|\mathbf{q}|$  for cases (b) A1 and (d) A2 is also shown.

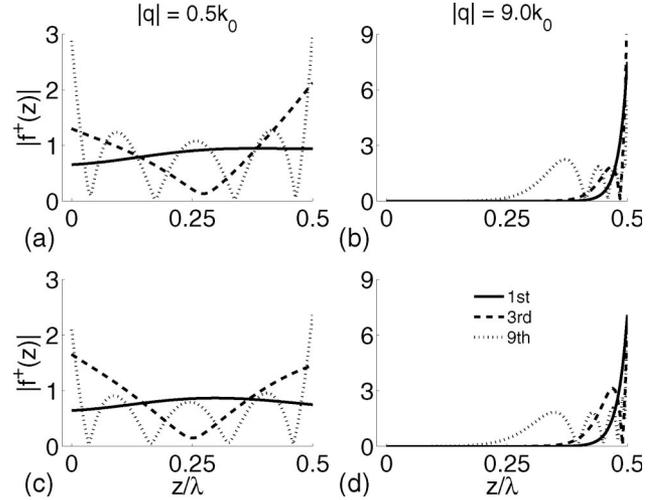


Fig. 2. Absolute values of the singular functions  $|f^+(z)|$  for the first-, third-, and ninth-largest singular values of the weighted measurement operator  $\mathbf{K}$  for case A1 and spatial frequencies (a)  $|\mathbf{q}| = 0.5k_0$  and (b)  $|\mathbf{q}| = 9k_0$ , and case A2 and spatial frequencies (c)  $|\mathbf{q}| = 0.5k_0$  and (d)  $|\mathbf{q}| = 9k_0$ .

where  $J_1$  is the Bessel function of the first kind of order one [35]. The scattered field due to the point scatterer was calculated for each incident wave using Eqs. (10) and (11), and noise was added to the simulated data. An estimate of the susceptibility was made using the truncated SVD [see Eq. (36)], using one of two truncation criteria. These truncation criteria were based on the signal-to-noise ratio (SNR) of the data.

As discussed in Subsection 3.B, it is assumed that the dominant noise source is photon noise. Under the weak hologram approximation, the standard deviation of the noise (and hence the scattered field) is given by  $\sigma_j \approx \sqrt{N_{0,j}}$ . Assuming an incident intensity of  $(A^r)^2 = 53.5$  mW, an effective “pixel” area (including quantum efficiency) of  $0.8 \times (0.05 \mu\text{m})^2$ , and an integration time of  $\Delta t = 1.0$  s, the standard deviation of the noise is  $\sigma_j = 2.31 \times 10^4 |\exp[i(k_z(\mathbf{q}_j^r) - k_z^*(\mathbf{q}_j^r))z_d/2]|$ . We see immediately that the noise statistics are independent of the pixel position (spatial frequency) and depend only on the incident wave vector and the height of the measurement plane. Consequently, for a given measurement (i.e., incident wave/detector pixel combination), the standard deviation of the noise is fixed and the SNR is determined by the transverse spatial-frequency structure of the scattered field.

For our purposes, it is most convenient to formulate a SVD truncation criteria in the object domain. Estimates of the expansion coefficients appearing in Eq. (36) may be found by projection of the object onto the singular functions:

$$\bar{d}_j(\mathbf{q}_{mn}) = f_j^H(z) \tilde{\eta}(z). \quad (43)$$

Let us define the SNR of the  $j$ th expansion coefficient as  $\text{SNR}_j = |\bar{d}_j|/\Delta d_j$ , where  $\Delta d_j$  (the standard deviation of  $d_j$ ) is given by Eq. (41). Substituting Eqs. (41) and (43) into the expression for  $\text{SNR}_j$ , and making use of Eq. (42), we find that the SNR of the  $j$ th expansion coefficient is

$$\text{SNR}_j(\mathbf{q}_{mn}) = \frac{WMs_j\eta_s}{L^2|\mathbf{q}_{mn}|} \left| \int_0^{\Delta z} dz [f_j^{+*}(\mathbf{q}_{mn}, z) + f_j^{-*}(\mathbf{q}_{mn}, z)] \sqrt{2za - z^2} J_1(|\mathbf{q}_{mn}| \sqrt{2za - z^2}) \right|. \quad (44)$$

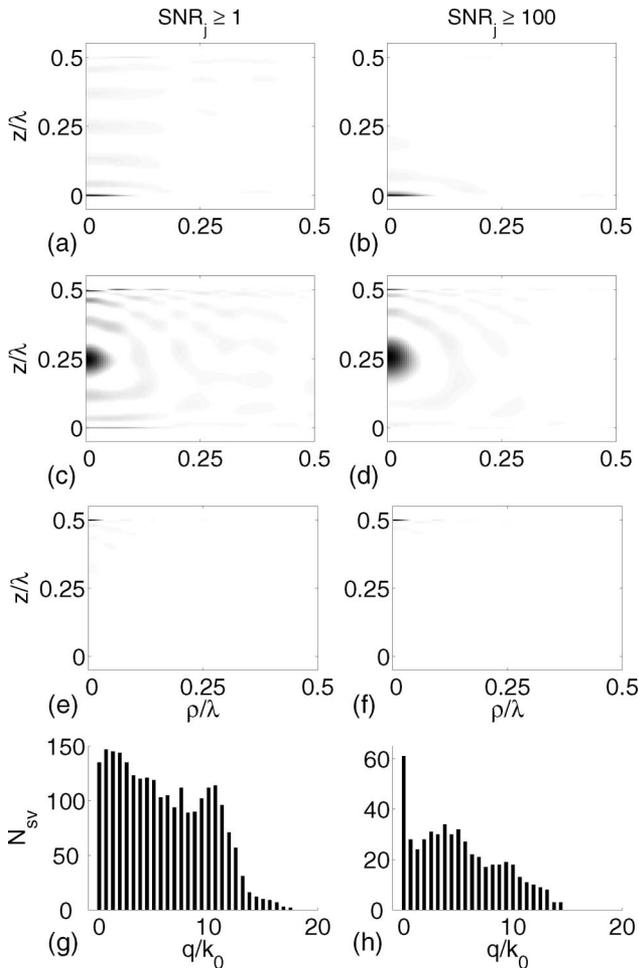


Fig. 3. Absolute value of the  $\delta$ -function response for case A1 and two values of the truncation threshold,  $\text{SNR}_j \geq 1$  and  $\text{SNR}_j \geq 100$ . The left-hand column illustrates the point response for  $\text{SNR}_j \geq 1$  at three different locations in the sample: (a) lower,  $z/\lambda = 0$ ; (c) middle,  $z/\lambda = 0.25$ ; and (e) upper,  $z/\lambda = 0.5$ , as well as (g) the number of singular values used in the reconstruction as a function of spatial frequency. The right-hand column illustrates the corresponding quantities for the case  $\text{SNR}_j \geq 100$ .

It should be pointed out that, while the (field) noise does not explicitly appear in Eq. (44), it is nonetheless accounted for by  $s_j$  and  $f_j(\mathbf{q}_{mn}, z)$ , which are calculated from the weighted measurement operator  $\mathbf{K}$ .

We considered two SVD truncation criteria. First, we included only those terms in the SVD that satisfied the inequality  $\text{SNR}_j > 1$ . This was the less restrictive of the two criteria, retains more terms, and incorporates more noisy data into the estimate of the susceptibility. The second SVD truncation criteria included only those terms that satisfied  $\text{SNR}_j > 100$ , limiting the amount of noisy data used to estimate the susceptibility. In doing so, the second criterion also limits the resolution of the system.

Figures 3 and 4 illustrate the point response for cases A1 and A2, respectively, for the two SVD truncation criteria. We see that the vertical localization is much better at the two faces than in the middle of the sample. This is due to the fact that we have *a priori* knowledge about the exact location of the faces. We also see that the resolution (both vertical and transverse) is better at the top face than at the bottom, because the top face is much closer to the measure-

ment plane. In addition, as one would expect, there is less spurious noise for  $\text{SNR}_j > 100$ , but at the cost of reduced resolution. We also illustrate, in each figure, the number of terms that are retained in the SVD for each case. We see that the resolution of the system worsens with increasing measurement distance (case A2). Consistent with this fact, we also see that fewer terms are retained in the SVD at each transverse spatial frequency (for a given SNR truncation point) as the distance to the measurement plane is increased.

Figures 5 through 8 refer to cases B1 and B2 in Table 1. In these cases, the sample is an order of magnitude thinner than in cases A1 and A2. Figures 5(a) and 5(c) are plots, for cases B1 and B2, respectively, of the singular value spectrum of the weighted measurement operator  $\mathbf{K}$  as a function of index  $j$ . In each case, the singular value spectrum is plotted for three spatial frequencies,  $|\mathbf{q}| = 0.5k_0$ ,  $3k_0$ , and  $9k_0$ . In case B1, we see that the singular value spectrum and, hence, the axial information content, is similar for all three spatial frequencies. In case B2, the amplitudes of the singular value spectrum are significantly smaller, as the measurement distance is much larger. The width of the singular value spectrum is smaller for cases B1 and B2 as compared to cases A1 and A2. This results in

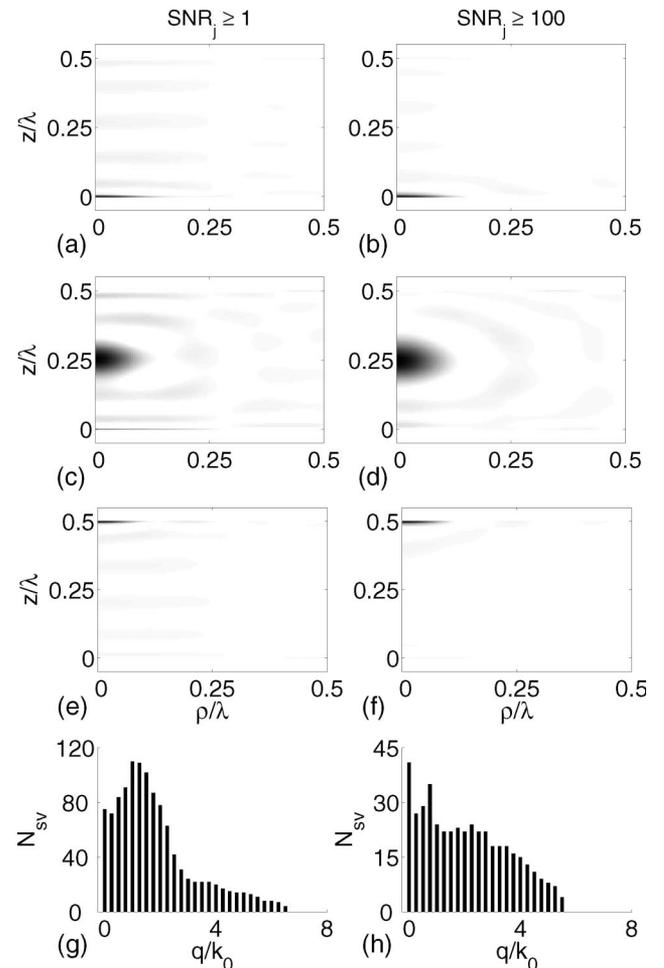


Fig. 4. Absolute value of the  $\delta$ -function response for case A2 and two values of the truncation threshold,  $\text{SNR}_j \geq 1$  and  $\text{SNR}_j \geq 100$ . The left-hand column illustrates the point response for  $\text{SNR}_j \geq 1$  at three different locations in the sample: (a) lower,  $z/\lambda = 0$ ; (c) middle,  $z/\lambda = 0.25$ ; and (e) upper,  $z/\lambda = 0.5$ , as well as (g) the number of singular values used in the reconstruction as a function of spatial frequency.

fewer resolvable axial sections in cases B1 and B2, although the absolute axial resolution in case B1 is superior to that in case A1, as expected.

Figures 5(b) and 5(d) show plots, for cases B1 and B2, respectively, of the log of the first-, third-, and ninth-largest singular values of the weighted measurement operator  $\mathbf{K}$  as a function of  $|\mathbf{q}|$ . For case B1, we see that the singular values have a relatively weak dependence on the spatial frequency, especially for smaller frequencies. This is consistent with Fig. 5(a). We also see that the spatial bandwidth (transverse information content) is largest for case B1, and decreases (as expected) with measurement distance (case B2). Also, the bandwidth in case B1 is much larger than the bandwidth in case A1, respectively, due to the fact that the measurement planes are much closer. The bandwidth in case B2 is comparable to that in case A2, however, because the measurement plane is roughly at the same distance.

Figure 6 shows plots of the absolute values of the singular functions  $|f^+(z)|$  for the first-, third-, and ninth-largest singular values of the weighted measurement operator  $\mathbf{K}$  for case B1 and spatial frequencies (a)  $|\mathbf{q}| = 0.5k_0$  and (b)  $|\mathbf{q}| = 9k_0$  and case B2 and spatial frequencies (c)  $|\mathbf{q}| = 0.5k_0$  and (d)  $|\mathbf{q}| = 9k_0$ . In contrast to cases A1 and A2, we see that the singular functions are essentially independent of both the transverse spatial frequency and measurement plane distance mainly because the sample is very thin.

Figures 7 and 8 illustrate the point response for cases B1 and B2, respectively, for the two SVD truncation criteria. As in cases A1 and A2, we see that the vertical localization is much better at the two faces than at the middle of the sample and better at the top face than at the bottom. Likewise, spurious noise in the reconstruction decreases with increasing the SNR threshold, but at the cost of poorer resolution. We also illustrate, in each figure, the number of terms that are retained in the SVD for each case. We see that the resolution is quite good in case B1 and much better than that in case A1 due to the close proximity of the measurement plane to the back face of the sample. In case B2, however, the resolution is comparable to that in case A2 because the measurement distance

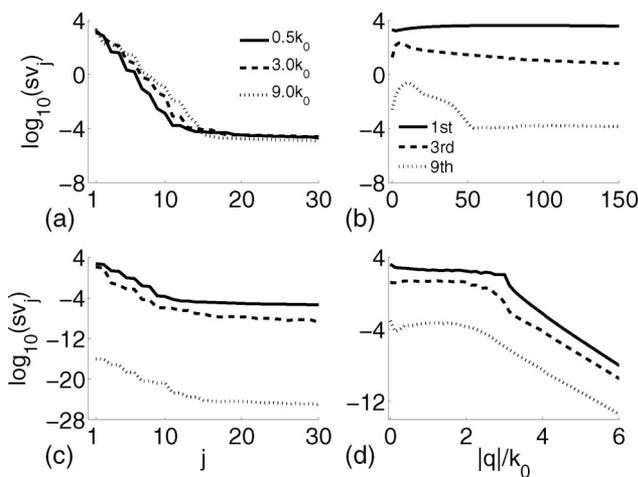


Fig. 5. Log of the singular values of the weighted measurement operator  $\mathbf{K}$  as a function of index  $j$  for cases (a) B1 and (c) B2 is shown. In both cases, the singular value spectra are plotted for three spatial frequencies,  $|\mathbf{q}| = 0.5k_0$ ,  $3k_0$ , and  $9k_0$ . The log of the first-, third-, and ninth-largest singular values of the weighted measurement operator  $\mathbf{K}$  as a function of spatial frequency  $|\mathbf{q}|$  for cases (b) B1 and (d) B2 is also shown.

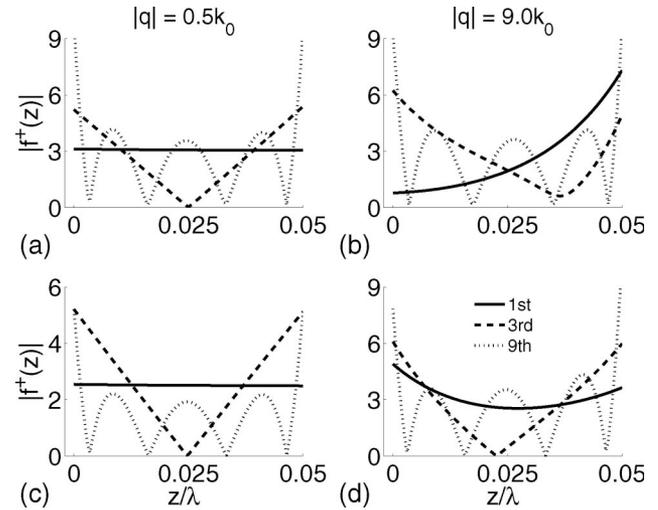


Fig. 6. Absolute values of the singular functions  $|f^+(z)|$  for the first-, third-, and ninth-largest singular values of the weighted measurement operator  $\mathbf{K}$  for case B1 and spatial frequencies (a)  $|\mathbf{q}| = 0.5k_0$  and (b)  $|\mathbf{q}| = 9k_0$ , and case B2 and spatial frequencies (c)  $|\mathbf{q}| = 0.5k_0$  and (d)  $|\mathbf{q}| = 9k_0$ .

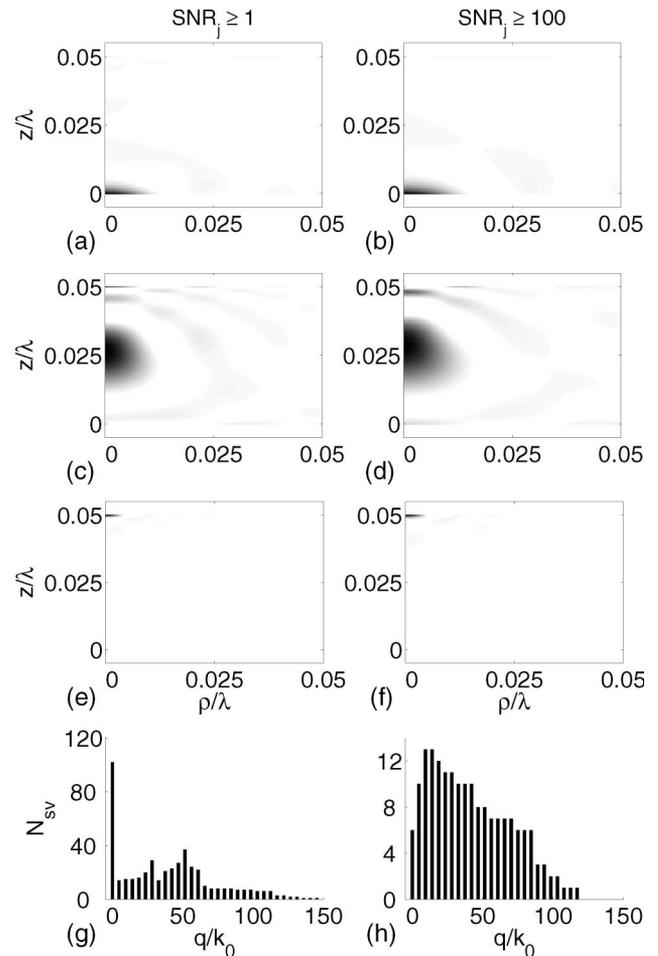


Fig. 7. Absolute value of the point response for case B1 and two values of the truncation threshold,  $\text{SNR}_j \geq 1$  and  $\text{SNR}_j \geq 100$ . The left-hand column illustrates the point response for  $\text{SNR}_j \geq 1$  at three different locations in the sample: (a) lower,  $z/\lambda = 0$ ; (c) middle,  $z/\lambda = 0.25$ ; and (e) upper,  $z/\lambda = 0.5$ , as well as (g) the number of singular values used in the reconstruction as a function of spatial frequency. The right-hand column illustrates the corresponding quantities for the case  $\text{SNR}_j \geq 100$ .

is of the same order, as we have seen above. In addition, paradoxically, the number of singular values retained for a given SNR threshold is less in case B1 than case A1, leading to reduced information content and poorer local resolution (fewer resolvable sections), despite much better global (or absolute) resolution.

In conclusion, we see that prior information, here in the form of a constraint on total sample thickness, can have a profound effect on the resolution while simultaneously reducing the effective information content of the measurements, defined by the number of significant singular values. Characteristic of near-field measurements, high spatial-frequency information is lost as the measurement plane is moved farther from the object. Because of the exponential behavior of the evanescent plane waves accessible in the near field, noise is exponentially amplified on reconstruction. We have provided SNR-based regularization criteria and examined the effects of the SNR requirements. As the required SNR of the reconstruction is increased, the available information, i.e., the number of expansion coefficients above the threshold, decreases. This results in effectively more data being used per degree of freedom in the reconstruction.

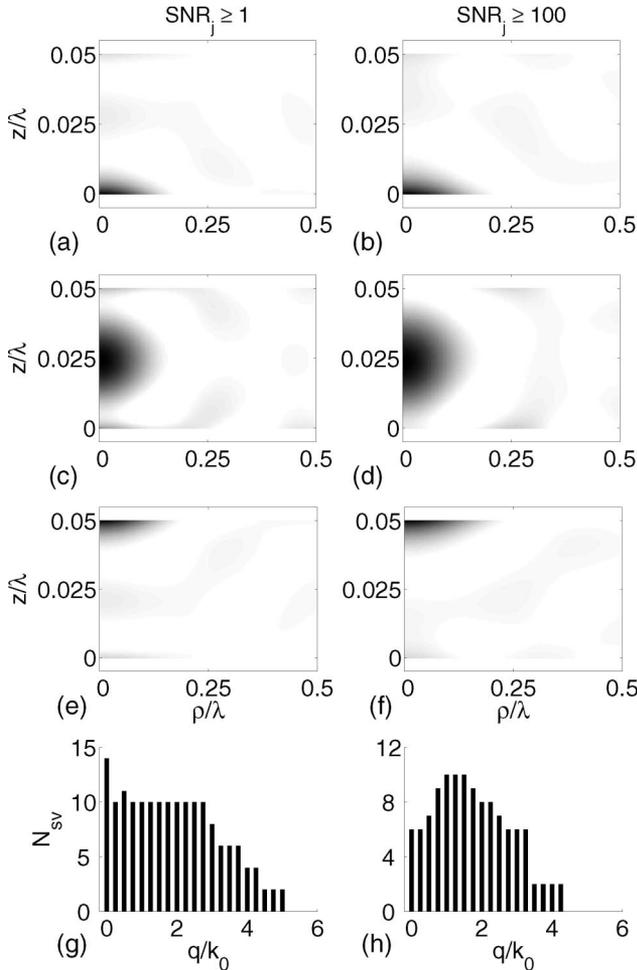


Fig. 8. Absolute value of the point response for case B2 and two values of the truncation threshold,  $\text{SNR}_j \geq 1$  and  $\text{SNR}_j \geq 100$ . The left-hand column illustrates the point response for  $\text{SNR}_j \geq 1$  at three different locations in the sample: (a) lower,  $z/\lambda = 0$ ; (c) middle,  $z/\lambda = 0.25$ ; and (e) upper,  $z/\lambda = 0.5$ , as well as (g) the number of singular values used in the reconstruction as a function of spatial frequency.

The model employed here included a weakly interacting probe, a linearized scattering model, and for the specific examples, Gabor-type holographic measurements. Future work will include nonlinear reconstructions for strongly scattering samples [19] and strongly interacting probes [36,37].

## APPENDIX A

### 1. Green's Tensor

The half-space Green's tensor has been discussed previously [1,13]. For the near-field case considered in this paper, the tensor  $G_{\alpha\beta}$  in Eq. (11) is given by

$$G_{\alpha\beta}(\mathbf{r}, \mathbf{r}') = \int \frac{d^2\mathbf{q}}{k_z(\mathbf{q})} h_{\alpha\beta}(\mathbf{q}, z') \exp[i\mathbf{k}(\mathbf{q}) \cdot (\mathbf{r} - \mathbf{r}')]. \quad (\text{A1})$$

Equation (A1) differs from the standard plane wave decomposition by a factor  $i/2\pi$  (which was included in Eq. (11) instead):

$$h_{\alpha\beta}(\mathbf{q}, z') = \frac{1}{|\mathbf{q}|^2} \begin{bmatrix} q_x^2 p_{xx} + q_y^2 p_{yy} & q_x q_y (p_{xx} - p_{yy}) & |\mathbf{q}| q_x p_{xz} \\ q_x q_y (p_{xx} - p_{yy}) & q_y^2 p_{xx} + q_x^2 p_{yy} & |\mathbf{q}| q_y p_{xz} \\ |\mathbf{q}| q_x p_{zx} & |\mathbf{q}| q_y p_{zx} & |\mathbf{q}|^2 p_{zz} \end{bmatrix}, \quad (\text{A2})$$

where  $\mathbf{q} = (q_x, q_y)$  and

$$\begin{aligned} p_{xx} &= \frac{k_z^2(\mathbf{q})}{k_0^2} [1 + R_2 \exp(2ik_z(\mathbf{q})z')], \\ p_{xz} &= \frac{-|\mathbf{q}|k_z(\mathbf{q})}{k_0^2} [1 - R_2 \exp(2ik_z(\mathbf{q})z')], \\ p_{yy} &= 1 + R_1 \exp(2ik_z(\mathbf{q})z'), \\ p_{zx} &= \frac{-|\mathbf{q}|k_z(\mathbf{q})}{k_0^2} [1 + R_2 \exp(2ik_z(\mathbf{q})z')], \\ p_{zz} &= \frac{|\mathbf{q}|^2}{k_0^2} [1 - R_2 \exp(2ik_z(\mathbf{q})z')]. \end{aligned} \quad (\text{A3})$$

The functions  $R_1(\mathbf{q})$  and  $R_2(\mathbf{q})$  are the reflection coefficients given by

$$R_1(\mathbf{q}) = \frac{k_z(\mathbf{q}) - k'_z(\mathbf{q})}{k_z(\mathbf{q}) + k'_z(\mathbf{q})}, \quad (\text{A4})$$

$$R_2(\mathbf{q}) = \frac{k'_z(\mathbf{q}) - n^2 k_z(\mathbf{q})}{k'_z(\mathbf{q}) + n^2 k_z(\mathbf{q})}, \quad (\text{A5})$$

with  $k'_z(\mathbf{q}) = \sqrt{n^2 k_0^2 - |\mathbf{q}|^2}$ .

### 2. Singular Value Decomposition of $\mathbf{K}$

Following Natterer [38], consider the SVD of  $\mathbf{K}$ :

$$\mathbf{K} = \mathbf{\Sigma} \mathbf{H} = \sum_{k=0}^{P-1} s_k \mathbf{g}_k \mathbf{f}_k(z)^H, \quad (\text{A6})$$

where  $\{s_k\}$  are the singular values (arranged in order from largest to smallest),  $\{\mathbf{g}_k\}$  is a set of singular (column) vectors that form a complete, orthonormal basis in  $\mathcal{C}^P$ , and

$$\mathbf{f}_k(\mathbf{z}) = \begin{bmatrix} \mathbf{f}_k^+(\mathbf{z}) \\ \mathbf{f}_k^-(\mathbf{z}) \end{bmatrix}$$

is a set of  $P$  of singular functions that form an incomplete, orthonormal basis in  $\mathcal{L}_2(0, \Delta z) \times \mathcal{L}_2(0, \Delta z)$ .

The  $P$  pairs of singular values and singular vectors  $\{(s_k^2, \mathbf{g}_k)\}$  are the eigenvalues and eigenvectors of the  $P \times P$  matrix  $\mathbf{K}\mathbf{K}^H$ , that may be determined via standard numerical algorithms (e.g., Ref. [34]). The singular functions obey the generalized eigenvalue equations,  $s_k \mathbf{g}_k = \mathbf{K}\mathbf{f}_k(\mathbf{z})$ , and  $s_k \mathbf{f}_k(\mathbf{z}) = \mathbf{K}^H \mathbf{g}_k$ .

The elements of the  $P \times P$  matrix  $\mathbf{K}\mathbf{K}^H$  are given by

$$\begin{aligned} (\mathbf{K}\mathbf{K}^H)_{jk} &= \frac{M^2}{\sigma_j \sigma_k} \int_0^{\Delta z} [H(\mathbf{q} - \mathbf{q}_j^i + \mathbf{q}_j^r, \mathbf{q}_j^r, \mathbf{q}_j^i, z) H^*(\mathbf{q} - \mathbf{q}_k^i \\ &+ \mathbf{q}_k^r, \mathbf{q}_k^r, \mathbf{q}_k^i, z) + H^*(-\mathbf{q} - \mathbf{q}_j^i + \mathbf{q}_j^r, \mathbf{q}_j^r, \mathbf{q}_j^i, z) \\ &\times H(-\mathbf{q} - \mathbf{q}_k^i + \mathbf{q}_k^r, \mathbf{q}_k^r, \mathbf{q}_k^i, z)] dz. \end{aligned} \quad (\text{A7})$$

The integral in Eq. (A7) may be performed analytically.

## ACKNOWLEDGMENTS

We would like to thank John C. Schotland, Brian J. Williams, and Farzad Kamalabadi for useful discussions, P. S. Carney would like to acknowledge support under United States Air Force Multidisciplinary Research Initiative grant F49620-03-1-0379 and National Aeronautics and Space Administration (NASA) grant NAG3-2764.

## REFERENCES

- J.-J. Greffet and R. Carminati, "Image formation in near-field optics," *Prog. Surf. Sci.* **56**, 133–237 (1997).
- D. Courjon, K. Sarayeddine, and M. Spajer, "Scanning tunneling optical microscopy," *Opt. Commun.* **71**, 23–28 (1989).
- M. Born and E. Wolf, *Principles of Optics*, 7th (expanded) ed. (Cambridge University, 1999).
- E. Synge, "A suggested method for extending microscopic resolution into the ultra-microscopic region," *Phil. Mag.* **6**, 356–362 (1928).
- E. Ash and G. Nicholls, "Super-resolution aperture scanning microscope," *Nature* **237**, 510–512 (1972).
- N. Garcia and M. Nieto-Vesperinas, "Direct solution to the inverse scattering problem for surfaces from near-field intensities without phase retrieval," *Opt. Lett.* **20**, 949–951 (1995).
- A. Lewis, M. Isaacson, A. Harootunian, and A. Muray, "Development of a 500 Å spatial resolution light microscope. I. light is efficiently transmitted through  $\lambda/16$  diameter apertures," *Ultramicroscopy* **13**, 227–231 (1984).
- E. Betzig and J. K. Trautman, "Near-field optics: microscopy, spectroscopy, and surface modification beyond the diffraction limit," *Science* **257**, 189–195 (1992).
- D. G. Fischer, "Sub-wavelength depth resolution in near-field microscopy," *Opt. Lett.* **25**, 1529–1531 (2000).
- P. S. Carney and J. C. Schotland, "Inverse scattering for near-field microscopy," *Appl. Phys. Lett.* **77**, 2798–2800 (2000).
- P. S. Carney and J. C. Schotland, "Three-dimensional total internal reflection microscopy," *Opt. Lett.* **26**, 1072–1074 (2001).
- P. S. Carney and J. C. Schotland, "Determination of three-dimensional structure in photon scanning tunneling microscopy," *J. Opt. A: Pure Appl. Opt.* **4**, S140–S144 (2002).
- P. S. Carney and J. C. Schotland, "Theory of total-internal-reflection tomography," *J. Opt. Soc. Am. A* **20**, 542–547 (2003).
- D. L. Marks and P. S. Carney, "Near-field diffractive elements," *Opt. Lett.* **30**, 1870–1872 (2005).
- P. S. Carney, R. A. Frazin, S. Bozhevolnyi, V. S. Volkov, A. Boltasseva, and J. C. Schotland, "A computational lens for the near field," *Phys. Rev. Lett.* **92**, 163903 (2004).
- A. Sentenac, C.-A. Guérin, P. C. Chaumet, F. Drsek, H. Giovannini, and N. Bertaux, "Influence of multiple scattering on the resolution of an imaging system: a Cramer–Rao analysis," *Opt. Express* **15**, 1340–1347 (2007).
- P. C. Chaumet, K. Belkebir, and A. Sentenac, "Superresolution of three-dimensional optical imaging by use of evanescent waves," *Opt. Lett.* **29**, 2740–2742 (2004).
- K. Belkebir, P. C. Chaumet, and A. Sentenac, "Superresolution in total internal reflection tomography," *J. Opt. Soc. Am. A* **22**, 1889–1897 (2005).
- G. Y. Panasyuk, V. A. Markel, P. S. Carney, and J. C. Schotland, "Nonlinear inverse scattering and three-dimensional near-field optical imaging," *Appl. Phys. Lett.* **89**, 221116 (2006).
- A. Sentenac, P. C. Chaumet, and K. Belkebir, "Beyond the Rayleigh criterion: grating assisted far-field optical diffraction tomography," *Phys. Rev. Lett.* **97**, 243901 (2006).
- R. A. Frazin, D. G. Fischer, and P. S. Carney, "Information content of the near field: two-dimensional samples," *J. Opt. Soc. Am. A* **21**, 1050–1057 (2004).
- S. Sotthivirat and J. A. Fessler, "Penalized-likelihood reconstruction for digital holography," *J. Opt. Soc. Am. A* **21**, 737–750 (2004).
- R. Solimene, G. Leone, and R. Pierri, "Resolution in two-dimensional tomographic reconstructions in the Fresnel zone from Born scattered fields," *J. Opt. A: Pure Appl. Opt.* **6**, 529–536 (2004).
- P. Guo and A. J. Devaney, "Digital microscopy using phase-shifting digital holography with two reference waves," *Opt. Lett.* **29**, 857–859 (2004).
- R. Pierri, A. Lisenò, F. Soldovieri, and R. Solimene, "In-depth resolution for a strip source in the Fresnel zone," *J. Opt. Soc. Am. A* **18**, 352–359 (2001).
- D. G. Fischer, "The information content of weakly scattered fields: implications for near-field imaging of three-dimensional structures," *J. Mod. Opt.* **47**, 1359–1374 (2000).
- A. Brancaccio, G. Leone, and R. Pierri, "Information content of Born scattered fields: results in the circular cylindrical case," *J. Opt. Soc. Am. A* **15**, 1909–1917 (1998).
- A. K. Louis, "A unified approach to regularization methods for linear ill-posed problems," *Inverse Probl.* **15**, 489–498 (1999).
- T. K. Moon and W. C. Sterling, *Mathematical Methods and Algorithms for Signal Processing* (Prentice-Hall, 2000).
- J. W. Goodman, *Statistical Optics* (Wiley, 1985).
- C. W. Helstrom, *Elements of Signal Detection & Estimation* (Prentice Hall, 1995).
- G. Demoment, "Image reconstruction and restoration: overview of common estimation structures and problems," *IEEE Trans. Acoust. Speech Signal Process.* **37**, 2024–2036 (1989).
- P. C. Hansen, "Numerical tools for analysis and solution of Fredholm integral equations of the first kind," *Inverse Probl.* **8**, 849–872 (1992).
- G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. (Johns Hopkins University, 1996).
- R. N. Bracewell, *The Fourier Transform and Its Applications*, 2nd revised ed. (McGraw-Hill, 1986).
- J. Sun, J. C. Schotland, and P. S. Carney, "Strong probe effects in near-field optics," *J. Appl. Phys.* **102**, 103103 (2007).
- J. Sun, J. C. Schotland, R. Hillenbrand, and P. S. Carney, "Nanoscale optical tomography based on volume-scanning near-field microscopy," *Appl. Phys. Lett.* **95**, 121108 (2009).
- F. Natterer, *The Mathematics of Computerized Tomography* (SIAM, 2001).