Spectral self-interference microscopy for low-signal nanoscale axial imaging

Brynmor J. Davis,[1,*] Anna K. Swan,[2] M. Selim Ünlü,[2,3] William C. Karl,[2,3] Bennett B. Goldberg,[2,3,4] John C. Schotland,[5] and P. Scott Carney[1,6]

[1]*Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, 405 North Mathews Avenue, Urbana, Illinois 61801, USA*

[2]*Department of Electrical and Computer Engineering, Boston University, 8 Saint Mary's Street, Boston, Massachusetts 02215, USA*

[3]*Department of Biomedical Engineering, Boston University, 44 Cummington Street, Boston, Massachusetts 02215, USA*

[4]*Department of Physics, Boston University, 590 Commonwealth Avenue, Boston, Massachusetts 02215, USA*

[5]*Department of Bioengineering, University of Pennsylvania, 210 South 33rd Street, Philadelphia, Pennsylvania 19104, USA*

[6]*Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, 1406 West Green Street, Urbana, Illinois 61801, USA*

[*]*Corresponding author: bryn@uiuc.edu*

Received June 8, 2007; accepted August 14, 2007;
posted September 20, 2007 (Doc. ID 83935); published October 25, 2007

A theoretical and numerical analysis of spectral self-interference microscopy (SSM) is presented with the goal of expanding the realm of SSM applications. In particular, this work is intended to enable SSM imaging in low-signal applications such as single-molecule studies. A comprehensive electromagnetic model for SSM is presented, allowing arbitrary forms of the excitation field, detection optics, and tensor sample response. An evanescently excited SSM system, analogous to total internal reflection microscopy, is proposed and investigated through Monte Carlo simulations. Nanometer-scale axial localization for single-emitter objects is demonstrated, even in low-signal environments. The capabilities of SSM in imaging more general objects are also considered—specifically, imaging arbitrary fluorophore distributions and two-emitter objects. A data-processing method is presented that makes SSM robust to noise and uncertainties in the detected spectral envelope.

OCIS codes: 100.3190, 110.4280, 180.2520, 180.3170.

1. INTRODUCTION

Fluorescence microscopy is an important tool in biological and biomedical research because it provides *in vivo* imaging, and a large range of fluorescent dyes are available for use as high-specificity molecular markers. Structures can be imaged using three-dimensional techniques such as deconvolution microscopy [1] or confocal microscopy [2]; however the resolution of fluorescence microscopes is limited by diffraction and constraints on the wavelengths usable for nondestructive imaging. At length scales below the diffraction limit, nonlinear techniques such as stimulated emission depletion microscopy have made impressive strides in biological imaging (e.g., [3]); however, they have yet to become standard laboratory tools.

Less complicated instruments are routinely used to collect subdiffraction-limit fluorescence information but require specific imaging scenarios. For example, total internal reflection microscopy (TIRM) [4] uses an evanescent field to excite fluorophores in a thin volume along the boundary of the sample. Since the resolution in the lateral dimensions (parallel to the sample boundary) is still diffraction limited, the main advantage of TIRM is the elimination of stray signals from outside the excitation volume. Alternatively, single-molecule microscopy [5,6] achieves subdiffraction localization of single fluorophores when it is known that only one molecule is present in the diffraction-limited focal volume (see [7] for an important example). The distinction between localization and resolution is important—given that a single fluorophore is present, it can be localized with a precision that is limited only by the signal-to-noise ratio (SNR) [8]. Resolution refers to the minimum feature size that can be imaged in an arbitrary object.

Spectral self-interference microscopy (SSM) [9] is a fluorescence technique that allows the inference of axial sample structure. The instrumentation required to realize this technique is relatively uncomplicated, and SSM has been shown to be useful in biological imaging applications [10]. In SSM, a mirror is placed behind the sample, and the spectrum of the total signal, that is, the signal from the directly propagated and reflected fields together, is collected. The distance from the mirror to the sample is chosen so that the sample-to-mirror optical path, measured with respect to the wavelength of the emission, varies by several multiples of 2π across the spectrum of the fluorophore. The light emitted from the fluorophore traverses two paths to the detector, producing interference on arrival. Since the optical path length, measured in units of wavelength, varies significantly across the fluorophore emission band, the interference

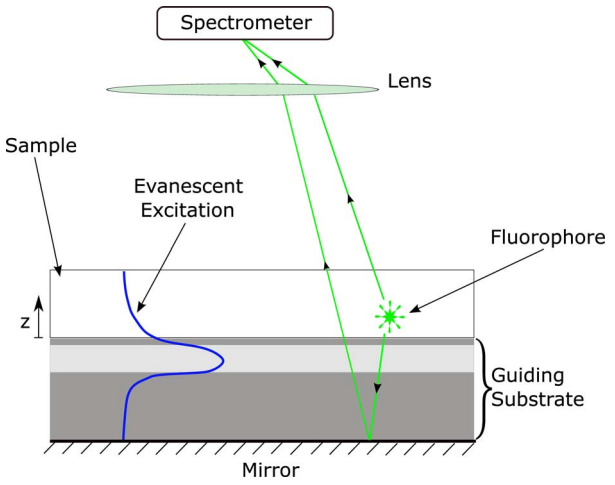

Fig. 1. (Color online) Evanescently excited spectral self-interference imaging system. Light is launched into the guiding substrate and results in an evanescent field in the sample. Sample fluorophores are excited in the same manner as in total internal reflectance microscopy, but the resulting emissions can travel two paths to the spectrometer. This leads to self-interference at the detector, which varies with wavelength.

cycles through constructive and destructive cases, producing oscillations in the spectrum. These oscillations encode the fluorophore-to-mirror distance and can be used to estimate the position of the fluorophore. The initial excitation of the fluorophore may be accomplished in a number of ways, including excitation by evanescent fields as shown in Fig. 1, a case considered in this paper. The SSM localization precision, when estimating the axial position of a single fluorescent layer, has been shown to be on the scale of a nanometer [9]. This level of precision is comparable with the lateral precision achieved by single-molecule imaging. The three-dimensional (i.e., both lateral and axial) localization of single fluorophores can also provide important biological information (e.g., [11]).

This work describes a number of advances intended to facilitate the use of SSM in a broader class of applications, including single-molecule studies, where SSM provides a means of inferring otherwise unobserved axial structure. An electromagnetic model of SSM is presented that encompasses a wide variety of instrument realizations and objects and generalizes previous SSM characterizations [12]. The electromagnetic nature of the model allows vectorial effects such as polarization dependence and molecule orientation to be taken into account. An evanescently excited SSM system, analogous to TIRM, is proposed. Evanescent excitation can be important in low-signal applications, where the limited excitation volume results in greatly reduced contributions from background fluorescence and hence an increase in the SNR. Numerical simulations are used to investigate the performance of SSM in low-signal applications. Algorithms are developed to allow the SSM imaging of objects with multiple emitters and to account for uncertainties in the spectral emission profile of the object. These algorithms are also investigated numerically.

The work is organized as follows. In Subsection 2.A a general model is developed relating an anisotropic fluorescent object to measured data for arbitrary systems of illumination and detection. In Subsection 2.B it is shown that a scalar model emerges under certain conditions, in Subsection 2.C an example is considered, and in Subsection 2.D the model is discretized. In Subsection 3.A reconstruction of arbitrary, continuous distributions of fluorophores is considered. In Subsection 3.B a prior assumption on the number of emitting fluorophores is used to condition the inverse problem. A dramatic improvement is observed, suggesting that SSM is well suited to dilute or single-molecule imaging. In Subsection 3.C evanescently excited SSM is compared with TIRM, and it is demonstrated that the additional reflection present in SSM makes a significant difference in the information available in the experiment. In Subsection 3.D uncertainty in the fluorophore spectrum is addressed, and SSM image reconstruction is shown to be robust with respect to variations in the spectral envelope. Finally, the results are summarized and placed in context in Section 4.

2. SSM FORWARD MODEL

An accurate forward model of the SSM system must be constructed to allow physically meaningful reconstructions of the object to be obtained. The model is related to the evanescently excited SSM system seen in Fig. 1; however, the framework encompasses a more general instrument. The tensor operators representing the excitation and detection optics can be chosen to represent a variety of other instrument geometries. The general mathematical framework is developed before special cases are considered.

A. General Model

The object to be imaged is illuminated through a linear optical system, characterized by a tensor $f_{uv}(\mathbf{r})$, so that illumination by a plane wave E_u^i normally incident to the entrance plane of the system produces a field distribution $f_{vu}(\mathbf{r})E_u^i$ in the region of the fluorophore. Here the subscripted symbols index the spatial dimensions, and the Einstein summation convention is used. The illumination field is assumed to be quasi-monochromatic with wavenumber k^i. This general representation of the illumination optics can include focusing by a lens (as in previous implementations of SSM), evanescent excitation from a waveguide (as shown in Fig. 1), or other more sophisticated techniques such as the use of laterally structured evanescent fields [13].

The fluorophore, or collection of fluorophores, is described by a spatially varying polarizability $\alpha_{wv}(\mathbf{r},k)$. The dependence of α_{wv} on k^i is not explicitly noted, since k^i is assumed to be fixed. The resulting source $S_w(\mathbf{r},k)$ is a function of position and frequency $\omega=ck$ (k being generally different from k^i) and is given by the expression

$$S_w(\mathbf{r},k) = (k^i)^2 \alpha_{wv}(\mathbf{r},k) f_{vu}(\mathbf{r}) E_u^i. \tag{1}$$

This source produces a field that may be computed by the method of Green's functions. The radiated field is given by the expression

$$\mathbf{E}^r(\mathbf{r},k) = \int \mathrm{d}^3 r' \bar{\mathbf{G}}(\mathbf{r},\mathbf{r}',k)\mathbf{S}(\mathbf{r}',k), \tag{2}$$

where $\bar{\mathbf{G}}$ is the dyadic Green's function satisfying the appropriate boundary conditions at the planar interface of the sample and substrate. Throughout this work bold type is used to denote a vector, bold type with an overbar is used to denote a tensor (with two or more indices), and individual components of vectors or tensors are given using subscripts and without bold type [as seen in Eq. (1)]. The Green's function used in Eq. (2) may be expressed in a plane-wave decomposition [14],

$$\bar{\mathbf{G}}(\mathbf{r},\mathbf{r}') = \frac{i}{2\pi}\int \mathrm{d}^2 k_\parallel e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')}\frac{1}{k_z}[\bar{\mathbf{D}}(\mathbf{k}) + e^{2ik_z(z'-z_m)}\bar{\mathbf{R}}(\mathbf{k})], \tag{3}$$

where it is assumed that the z axis is normal to the substrate–sample boundary, that $z=z_m$ is the mirror plane, and that $z>z'$ (i.e., the observation point is above the fluorescent object). The vector $\mathbf{k}=\mathbf{k}_\parallel+k_z\hat{\mathbf{z}}$ (where a hat $\hat{}$ is used to indicate a unit vector) and

$$k_z = \begin{cases} \sqrt{k^2-k_\parallel^2} & k \geq k_\parallel \\ i\sqrt{k_\parallel^2-k^2} & k < k_\parallel \end{cases}. \tag{4}$$

The dyads $\bar{\mathbf{D}}$ and $\bar{\mathbf{R}}$ enforce the transversality (with respect to $\mathbf{k}$) of the plane waves directly propagated to the detector and of those reflected from the substrate. The dyad $\bar{\mathbf{D}}$ is simply the identity on vectors transverse to $\mathbf{k}$ and the dyad $\bar{\mathbf{R}}$ contains the Fresnel coefficients. The phase acquired by the reflected wave on propagation to the mirror and back to the source plane is made explicit in the factor $e^{2ik_z(z'-z_m)}$ multiplying $\bar{\mathbf{R}}$. The dyads could, for instance, be written in terms of the transverse electric (TE) and transverse magnetic (TM) basis vectors, $\hat{\mathbf{u}}_{te/tm}$, relative to $\mathbf{k}$ and the normal to the interface $\mathbf{n}$. These unit vectors may be constructed,

$$\hat{\mathbf{u}}_{te}(\mathbf{k}) = \frac{\mathbf{k}\times\mathbf{n}}{|\mathbf{k}\times\mathbf{n}|}, \tag{5}$$

$$\hat{\mathbf{u}}_{tm}(\mathbf{k}) = \frac{\mathbf{k}\times\hat{\mathbf{u}}_{te}}{|\mathbf{k}\times\hat{\mathbf{u}}_{te}|}, \tag{6}$$

and it may be noted that $\hat{\mathbf{k}}=\hat{\mathbf{u}}_{te}\times\hat{\mathbf{u}}_{tm}=\mathbf{k}/k$ so that $\hat{\mathbf{k}}$, $\hat{\mathbf{u}}_{te}$, $\hat{\mathbf{u}}_{tm}$ form an ordered orthonormal triple, with $\hat{\mathbf{u}}_{te}$ always parallel to the interface. Then

$$\bar{\mathbf{D}}(\mathbf{k}) = \hat{\mathbf{u}}_{te}(\mathbf{k})[\hat{\mathbf{u}}_{te}(\mathbf{k})]^\dagger + \hat{\mathbf{u}}_{tm}(\mathbf{k})[\hat{\mathbf{u}}_{tm}(\mathbf{k})]^\dagger, \tag{7}$$

$$\bar{\mathbf{R}}(\mathbf{k}) = \hat{\mathbf{u}}_{te}(\mathbf{k})r_{te}[\hat{\mathbf{u}}_{te}(\tilde{\mathbf{k}})]^\dagger + \hat{\mathbf{u}}_{tm}(\mathbf{k})r_{tm}[\hat{\mathbf{u}}_{tm}(\tilde{\mathbf{k}})]^\dagger, \tag{8}$$

where † denotes the Hermitian conjugate, r_{te} and r_{tm} are the Fresnel reflection coefficients, and the vector $\tilde{\mathbf{k}}$ is the reflection of $\mathbf{k}$ through the $z=0$ plane, i.e., $\tilde{k}_z=-k_z$.

It may be seen that for large values of kr, the Green's function takes the asymptotic form

$$\bar{\mathbf{G}}(r\hat{\boldsymbol{\sigma}},\mathbf{r}') \sim \frac{e^{ikr}}{r}e^{-ik\hat{\boldsymbol{\sigma}}\cdot\mathbf{r}'}[\bar{\mathbf{D}}(k\hat{\boldsymbol{\sigma}}) + e^{2ik\hat{\sigma}_z(z'-z_m)}\bar{\mathbf{R}}(k\hat{\boldsymbol{\sigma}})] \tag{9}$$

in directions specified by the unit vector $\hat{\boldsymbol{\sigma}}$. This spherical wave propagates to an optical system that focuses the light to a detector. The optical system is formally characterized by the dyad $\bar{\mathbf{L}}(\hat{\boldsymbol{\sigma}})$. The propagated field is considered on a reference hemisphere of constant $r=r_0$ with $z>0$, and $\bar{\mathbf{L}}(\hat{\boldsymbol{\sigma}})$ maps the field at position $\hat{\boldsymbol{\sigma}}r_0$ to the resulting field at the detector. For the case that the imaging system is a lens, $\bar{\mathbf{L}}$ is nonzero on the region of the reference sphere falling within the lens aperture and may be computed by standard means [15]. Any polarization preference or efficiency factor for the detector may also be included in $\bar{\mathbf{L}}$. It is useful to then define a new dyadic function $\bar{\mathbf{g}}$ that takes into account the free-space propagation and the effects of the detection system, including the summation over the reference sphere, namely,

$$\begin{aligned}\bar{\mathbf{g}}(\mathbf{r}',k) &= \int_{\hat{\sigma}_x^2+\hat{\sigma}_y^2\leq 1} \mathrm{d}\hat{\sigma}_x \mathrm{d}\hat{\sigma}_y \bar{\mathbf{L}}(\hat{\boldsymbol{\sigma}})\bar{\mathbf{G}}(r_0\hat{\boldsymbol{\sigma}},\mathbf{r}'), \\ &= e^{ikr_0}\int_{\hat{\sigma}_x^2+\hat{\sigma}_y^2\leq 1} \mathrm{d}\hat{\sigma}_x \mathrm{d}\hat{\sigma}_y \frac{\bar{\mathbf{L}}(\hat{\boldsymbol{\sigma}})}{r_0} \\ &\quad \times e^{-ik\hat{\boldsymbol{\sigma}}\cdot\mathbf{r}'}[\bar{\mathbf{D}}(k\hat{\boldsymbol{\sigma}}) + e^{2ik\hat{\sigma}_z(z'-z_m)}\bar{\mathbf{R}}(k\hat{\boldsymbol{\sigma}})], \\ &= \bar{\mathbf{g}}^{\mathrm{Direct}}(\mathbf{r}',k) + \bar{\mathbf{g}}^{\mathrm{Reflected}}(\mathbf{r}',k).\end{aligned} \tag{10}$$

The field at the detector can then be found by propagating the source $\mathbf{S}$ [see Eq. (1)] to the detector:

$$\mathbf{E}^d(k) = \int \mathrm{d}^3 r' \bar{\mathbf{g}}(\mathbf{r}',k)\mathbf{S}(\mathbf{r}',k). \tag{11}$$

The detector is assumed to perform a time average of the intensity. All stochastic processes are assumed to be stationary and ergodic in the wide sense, so that the time-averaged spectrometer measurements and ensemble averages (denoted by angle brakets $\langle\cdot\rangle$) in the frequency domain are interchangeable. The detector measurement, $d(k)$, may then be written as

$$\begin{aligned} d(k) &= \mathrm{Tr}\{\langle \mathbf{E}^d(k)[\mathbf{E}^d(k)]^\dagger\rangle\}, \\ &= \mathrm{Tr}\left\{\int \mathrm{d}^3 r \int \mathrm{d}^3 r' \right. \\ &\quad \left. \times \bar{\mathbf{g}}(\mathbf{r},k)\langle \mathbf{S}(\mathbf{r},k)\mathbf{S}^\dagger(\mathbf{r}',k)\rangle \bar{\mathbf{g}}^\dagger(\mathbf{r}',k)\right\}, \end{aligned} \tag{12}$$

where $\mathrm{Tr}\{\cdot\}$ is the trace. The coherence matrix of the incident field may be expressed as

$$\langle E_u^i (E_{u'}^i)^*\rangle = A^i \mu_{uu'}, \tag{13}$$

where A^i is given by the expression $A^i=\mathrm{Tr}\{\langle \mathbf{E}^i(\mathbf{E}^i)^\dagger\rangle\}$. If the field is unpolarized, $\mu_{ij}=\delta_{ij}/2$, where δ is the Kronecker tensor. If the field is fully polarized, $\mu_{ij}=\hat{u}_i\hat{u}_j^*$, where $\hat{\mathbf{u}}$ is a unit vector describing the polarized illumi-

nation field in a plane parallel to the entrance plane of the system.

The emissions of individual fluorophores are uncorrelated, and so

$$\langle \alpha_{wv}(\mathbf{r},k)\alpha^{*}_{w'v'}(\mathbf{r}',k)\rangle = \rho_{wvw'v'}(\mathbf{r},k)\delta(\mathbf{r}-\mathbf{r}'), \quad (14)$$

where $\bar{\boldsymbol{\rho}}$ is a tensor mapping the polarization state of the excitation field to the resulting emitted-field polarization state. Additionally, the incident field and the polarizability are assumed to be incoherent because the fluouresence is spontaneous. As a result, the source density correlation function then takes the form

$$\langle S_w(\mathbf{r},k)S^{*}_{w'}(\mathbf{r}',k)\rangle = (k^i)^4 A^i f_{vu}(\mathbf{r})\mu_{uu'}f^{*}_{v'u'}(\mathbf{r}') \times \rho_{wvw'v'}(\mathbf{r},k)\delta(\mathbf{r}-\mathbf{r}'), \quad (15)$$

so that the detector measurement is

$$d(k) = (k^i)^4 \int \mathrm{d}^3 r\, g_{lw}(\mathbf{r},k)A^i f_{vu}(\mathbf{r})\mu_{uu'}f^{*}_{v'u'}(\mathbf{r}) \times \rho_{wvw'v'}(\mathbf{r},k)g^{*}_{lw'}(\mathbf{r},k). \quad (16)$$

In many cases of interest the spectral dependence of the fluorophore is separable from the spatial and polarization dependences. In those cases, it is convenient to write

$$A^i\rho_{wvw'v'}(\mathbf{r},k) = A(k)\rho_{wvw'v'}(\mathbf{r}). \quad (17)$$

The detector measurement is then

$$d(k) = (k^i)^4 A(k) \int \mathrm{d}^3 r\, g_{lw}(\mathbf{r},k)f_{vu}(\mathbf{r})\mu_{uu'}f^{*}_{v'u'}(\mathbf{r}) \times \rho_{wvw'v'}(\mathbf{r})g^{*}_{lw'}(\mathbf{r},k). \quad (18)$$

From this general model it may be seen that the SSM data depend on the illumination polarization through $\mu_{uu'}$, the illumination optics through $f_{vu}(\mathbf{r})$, the object through $\rho_{wvw'v'}(\mathbf{r})$, and the detection optics through $g_{lw}(\mathbf{r},k)$. The retention of a full electromagnetic description allows the modeling of effects such as orientation-dependent emission from a fluorescent molecule (see [16] for a discussion of measurement schemes). Indeed, the model of Eq. (18) can be used to describe methods such as dipole-orientation imaging [17] or chirality imaging (e.g., [18]). Techniques such as these rely on polarization effects to obtain information about the nonscalar object. However, in many situations it is desirable to be able to define a simplified scalar model, as is discussed in the next subsection.

B. Scalar Model

It is often possible to simplify the analysis of optical systems by the use of a scalar model. The full vector model of Eq. (18) can be reduced in this manner for systems in which the excitation and detectable fields exhibit spatially invariant polarization. That is, the dyads describing the excitation and detection have the following separable forms:

$$\bar{\mathbf{f}}(\mathbf{r}) = f(\mathbf{r})\bar{\boldsymbol{\nu}}, \quad (19)$$

$$\bar{\mathbf{g}}(\mathbf{r},k) = g(\mathbf{r},k)\bar{\boldsymbol{\xi}}. \quad (20)$$

For example, such a situation arises for plane-wave illumination or detection; is approximately valid for focused detection with low numerical aperture (NA), where there is not a significant variation in polarization state; and is invalid for high-NA focusing where the polarization state varies significantly within the focal region.

Assuming that the forms of Eqs. (19) and (20) are justified, a scalar object $s(\mathbf{r})$ can be defined as

$$s(\mathbf{r}) = \xi_{l,w}\nu_{vu}\mu_{uu'}\nu^{*}_{v'u'}\rho_{wvw'v'}(\mathbf{r})\xi^{*}_{lw'}. \quad (21)$$

Here the illumination polarization state and the microscope optics combine to operate on the second-order correlations of the polarizability. This determines which components of $\bar{\boldsymbol{\alpha}}$ are visible in the data. If the entire problem is treated within a scalar approximation, that is, if Eq. (1) becomes $S=\alpha f(\mathbf{r})E^i$ and all propagation is treated within the context of a scalar (Helmholtz) wave equation, then it may be seen that $s(\mathbf{r})$ is the square magnitude of the polarizability $\alpha(\mathbf{r})$, all other factors being associated with polarization effects.

The relationship in Eq. (21) can be used to rewrite Eq. (18) as a scalar model,

$$d(k) = (k^i)^4 A(k) \int \mathrm{d}^3 r |g(\mathbf{r},k)|^2 |f(\mathbf{r})|^2 s(\mathbf{r}). \quad (22)$$

Here the scalar object $s(\mathbf{r})$ is illuminated by light with an intensity pattern of $|f(\mathbf{r})|^2$. The detection optics result in an intensity sensitivity pattern $|g(\mathbf{r},k)|^2$, and the spectral response of the fluorophore is described by $A(k)$. Such a treatment is often used in the modeling of microscopy systems [19].

C. Example System

In this section an example system is described to illustrate the application of the theory developed. The resulting model is used in subsequent numerical simulations. As shown in Fig. 1, an evanescent excitation field is considered. A low-NA objective lens is used, and so a scalar model can be employed. The detection system does not have a preferred polarization orientation. For simplicity, the substrate below the sample is modeled as a perfect mirror at the plane $z=z_m$, where $z=0$ is taken to be the substrate–sample boundary.

The illumination is assumed to be a TE polarized field, evanescent in the sample region. Thus the illumination optics are described by the expression

$$\bar{\mathbf{f}}(\mathbf{r}) = \{\hat{\mathbf{u}}_{te}(\mathbf{k}^i)[\hat{\mathbf{u}}_{te}(\mathbf{k}^i)]^{\dagger} + \hat{\mathbf{u}}_{tm}(\mathbf{k}^i)[\hat{\mathbf{u}}_{tm}(\mathbf{k}^i)]^{\dagger}\}e^{i\mathbf{k}^i\cdot\mathbf{r}}, \quad (23)$$

and the illumination polarization is given by

$$\hat{\mathbf{u}} = \hat{\mathbf{u}}_{te}(\mathbf{k}^i), \quad (24)$$

where $\mathbf{k}^i=(k^i_x,k^i_y,i|k^i_z|)^T$ is the wave vector for the evanescent field (with ^{T} representing the transpose operator). It can be seen that $\bar{\mathbf{f}}$ is of the form given in Eq. (19).

The optical axis of the lens coincides with the z axis, so the low-NA approximation means that only fields in the

$x-y$ plane are detected. Assuming that the detection optics do not change the $x-y$ polarization on the way to the detector,

$$\bar{\mathbf{g}}(\mathbf{r},k) = g(\mathbf{r},k)[\hat{\mathbf{x}}\hat{\mathbf{x}}^{\dagger} + \hat{\mathbf{y}}\hat{\mathbf{y}}^{\dagger}], \tag{25}$$

where $g(\mathbf{r},k)$ is the scalar lens response (including reflections from the mirror).

In many cases of interest, the fluorophore is part of a long, oriented molecule and so there is a strong preference to accept and radiate only one polarization state. Then, $\bar{\boldsymbol{\alpha}}$ may be expressed as the dyadic product of a vector and its conjugate, $\alpha_{wv}=p_w p_v^*$, so that

$$\rho_{wvw'v'} = p_w p_v^* p_{w'}^* p_{v'}. \tag{26}$$

The scalar object may be evaluated to obtain

$$s(\mathbf{r}) = |[\hat{\mathbf{u}}_{te}(\mathbf{k}^i)]^{\dagger}\mathbf{p}|^2 |\mathbf{p}\times\hat{\mathbf{z}}|^2. \tag{27}$$

The factor $|[\hat{\mathbf{u}}_{te}(\mathbf{k}^i)]^{\dagger}\mathbf{p}|^2$ accounts for the portion of the dipole parallel to the excitation field, while the factor $|\mathbf{p}\times\hat{\mathbf{z}}|^2$ accounts for the portion of the dipole that is visible to the detection optics. Susbstituting the excitation model into Eq. (22), the scalar observation model is then

$$d(k) = (k^i)^4 A(k) \int \mathrm{d}^3 r |g(\mathbf{r},k)|^2 e^{-2|k_z^i|z} s(\mathbf{r}). \tag{28}$$

The detection pattern $g(\mathbf{r},k)$ can be further specified by picking a focal point for the focused detection. Since a perfect mirror is assumed, the reflected detection pattern is the negative of the direct detection pattern reflected through $z=z_m$, i.e.,

$$g^{\text{Reflected}}(x,y,z,k) = -g^{\text{Direct}}(x,y,-(z-2z_m),k). \tag{29}$$

For the particular case that the focal plane and mirror plane coincide, the detection pattern may be expressed in a symmetric form. Let $l(\mathbf{r},k)$ be the scalar field resulting from a lens focused to the origin without a mirror present. The detection pattern can then be written as a function of this field,

$$\begin{aligned} g(\mathbf{r},k) &= g^{\text{Direct}}(\mathbf{r},k) + g^{\text{Reflected}}(\mathbf{r},k), \\ &= l(x,y,z-z_m,k) - l(x,y,-(z-z_m),k). \end{aligned} \tag{30}$$

It is worth reiterating that $l(\mathbf{r},k)$ may be computed by standard methods; e.g., the focused field may be modeled as a Gaussian beam [20]. The approach taken in this work is to first use a full vectorial focusing model [15] with the scalar $l(\mathbf{r},k)$ taken to be the component in the same polarization direction as the illuminating field $\mathbf{E}^i$ (with a small NA the other focused-field components are negligible.) It may be seen on calculation that $|l(x,y,z,k)|=|l(x,y,-z,k)|$, so that the two terms composing $g(\mathbf{r},k)$ are of equal magnitude and differ only by a phase that varies with position and wavelength. Thus the interference is strongly visible, and the spectrum is modulated as discussed in the introduction.

D. Discrete Forward Model and Multiple Illuminations

For many computational approaches, it is necessary to have a discrete forward model. The system from Subsection 2.C can be sampled at discrete k and $\mathbf{r}$ values at rates fine enough to give negligible departure from the continuous model. A precise treatment of this sort of approximation can be found in [21].

As shown in Subsection 2.B, a scalar model can be used in certain circumstances. A scalar model will be used to demonstrate image reconstruction, allowing the discretized model to be written as a simple matrix equation. Let $\mathbf{s}$ be the discrete scalar object arranged into a one-dimensional vector. Similarly, let $\mathbf{d}$ be the data arranged in a vector. A matrix $\mathbf{C}$ relates the two,

$$\mathbf{d} = \mathbf{C}\mathbf{s}. \tag{31}$$

The nth column of $\mathbf{C}$ represents the data produced by a unit-amplitude entry in the nth position of the object vector $\mathbf{s}$. The data that are actually observed are similar to $\mathbf{d}$ but corrupted by some amount of noise.

A multiexcitation system can also be modeled. Several sets of data can be taken from the same object by exciting it with a number of different fields. This allows the possibility of gathering more information and improving the reconstructed image. In each case the illumination $|f(\mathbf{r})|^2$ is different, which results in a different $\mathbf{C}$ matrix and a different set of data. These multiple models can be combined into a single augmented model by vertically concatenating the $\mathbf{d}$ vectors and the $\mathbf{C}$ matrices. This results in a model of the same form as Eq. (31) but describing the multiexcitation observation.

For evanescent excitation, as considered in Subsection 2.C, the differing illumination profiles are chosen to have different decay rates (given by k_z^i) while retaining the same TE direction. This ensures that the augmented model can be represented in a scalar framework. If TM illumination is used and the magnitude k^i maintained, the TM direction necessarily changes with k_z^i, a situation not compatible with an augmented scalar model.

3. IMAGE RECONSTRUCTION

The forward model for the imaging system can be inverted by various procedures. That is, knowledge of how the object relates to the data can be used to infer properties of the object from measured data. In the following subsections, the inverse problem is investigated with various algorithms and various constraints on the object.

A. Reconstructions of Arbitrary Distributions

1. Maximum Likelihood Estimation

In the context developed here, image reconstruction is the process of using the measured data to find a vector $\check{\mathbf{s}}$ that is close to the unknown object $\mathbf{s}$. This is done by ensuring that $\mathbf{C}\check{\mathbf{s}}$, the data expected from the reconstructed image, is close (in terms of some criterion or norm) to the measured data. There are various ways of measuring this closeness, but one of the most popular estimation rules is maximum likelihood (ML). The conditional probability density function of observing the data $\mathbf{d}$ given the object $\mathbf{s}$ is $P(\mathbf{d}|\mathbf{s})$, and the ML estimator is defined as

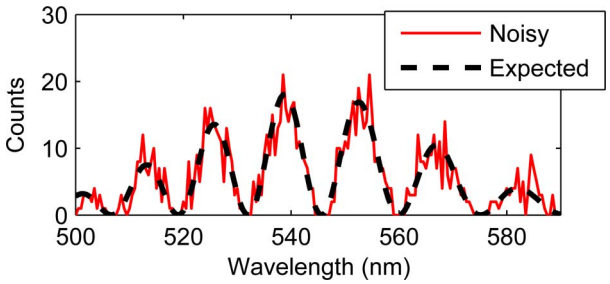

Fig. 2. (Color online) Example data from an evanescently excited fluorophore positioned at $z=20$ nm. The deterministic expected-value of the signal [calculated by using Eqs. (28) and (30)] is shown along with an example realization of the Poisson noise. The signal level is set so that the data have a total expected count of 1000 (corresponding to a SNR of 10 dB). The fluorophore is excited at a free-space wavelength of 488 nm, and the emission is measured in increments of 0.5 nm between 500 and 590 nm. A Gaussian form for the spectral envelope $A(k)$ is assumed. The fluorophore is employed to lie in water, and a perfect mirror is placed at a position z_m that corresponds to an optical path length of 8 μm in water. The resulting reflected path can be seen to produce interferometric oscillations in the spectrum. The excitation is modeled as the evanescent field resulting from a TE polarized plane wave striking a water–silicon nitride boundary at an angle of 45° (the critical angle is 40.7°). A lens of numerical aperture 0.1 is used to collect the emissions, and the fluorophore is assumed to lie on the optic axis of this lens.

$$\check{\mathbf{s}} = \underset{\mathbf{s}}{\operatorname{argmax}}\, P(\mathbf{d}|\mathbf{s}). \tag{32}$$

An ML estimator is optimal with respect to the Cramér–Rao bound [22].

It is assumed that the data collected by the SSM system are consistent with photon counting and therefore obey Poisson statistics. This means that $\mathbf{d}$ in Eq. (31) gives the expected value of the data. It is also assumed that measurements at any two distinct k values are statistically independent. An example of the data produced by such a model can be seen in Fig. 2. By using the Poisson model and Eq. (32), the following condition on the ML estimate $\check{\mathbf{s}}$ can be found [23]:

$$\mathbf{C}^T\mathbf{\Lambda}^{-1}(\check{\mathbf{s}})\mathbf{C}\check{\mathbf{s}} = \mathbf{C}^T\mathbf{\Lambda}^{-1}(\check{\mathbf{s}})\mathbf{d}, \tag{33}$$

where $\mathbf{\Lambda}(\check{\mathbf{s}})$ is a matrix that is zero except on the main diagonal, where it has the elements of the vector $\mathbf{C}\check{\mathbf{s}}$,

$$\mathbf{\Lambda}(\check{\mathbf{s}}) = \mathrm{diag}[\mathbf{C}\check{\mathbf{s}}]. \tag{34}$$

The Poisson statistical model gives a noise variance that is proportional to the expected signal at each wavelength measured. The matrix $\mathbf{\Lambda}^{-1}(\check{\mathbf{s}})$ in Eq. (33) can be interpreted as a fit-to-data weighting that reflects the Poisson noise model—the fit at lower-noise regions of the signal is weighted more heavily.

In general, the ML condition of Eq. (33) is insufficient to uniquely define an image $\check{\mathbf{s}}$. Should $\mathbf{C}$ have a nonempty nullspace, then any component of $\check{\mathbf{s}}$ lying within this nullspace would not affect Eq. (33) and would therefore be unconstrained. Even if the nullspace of $\mathbf{C}$ were empty, the ML problem is strongly ill conditioned below some finite spatial scale. This results in instability in the presence of even a small amount of noise. These issues are well understood in image reconstruction and have resulted in the development of regularization techniques for inverse problems [24].

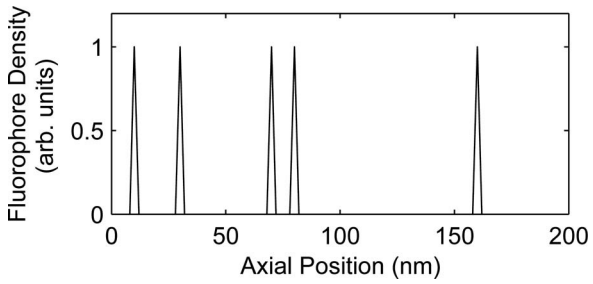

Fig. 3. Test object considered for RL reconstruction. Five fluorophores lie at varying positions along the optical axis of the detection lens. The axial position is sampled at 2 nm intervals.

2. Richardson–Lucy Reconstruction

The Richardson–Lucy (RL) algorithm [25,26] can be used to iteratively find a solution to Eq. (33). The RL algorithm is generally terminated before convergence in order to attain the predictable behavior associated with a well-regularized inverse method. Additionally, the RL algorithm is known to preferentially reconstruct pointlike objects [27], such as spatially distinct fluorescing molecules. The behavior of the RL algorithm in an SSM environment will be investigated through numerical simulation. A test object is shown in Fig. 3. This object has a variety of spacings between fluorescing molecules. The level of detail resolved determines the performance of the imaging system.

Two different excitation fields are considered in the simulations—the first is modeled by the evanescent field produced as described in Fig. 2, i.e., that resulting from a plane wave incident at 45° on a water–silicon nitride boundary. The second excitation field is produced by increasing the angle of incidence of the illuminating plane wave to 55°. The axial intensity profiles of these two excitations are shown in Fig. 4. A single-excitation system consisting of only the 45° excitation is considered, along with a two-excitation system that measures both the 45° and 55° cases. The other system parameters are the same as those described in Fig. 2.

The evolution of a RL reconstruction with iteration number is shown in Fig. 5. It can be seen that the RL estimate is smooth to start with and becomes progressively

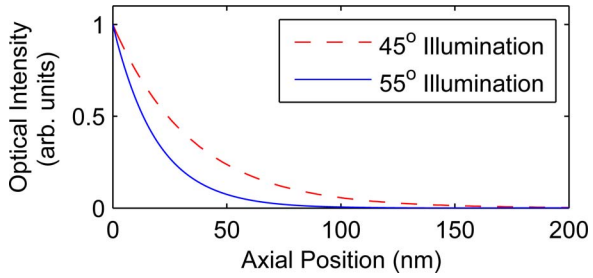

Fig. 4. (Color online) Axial intensity profiles corresponding to the two excitation regimes considered. The excitation is modeled as the evanescent field produced at a water–silicon nitride boundary when a plane wave is incident, from the silicon nitride substrate, at an angle greater than the critical angle. Incident angles of 45° and 55° are considered.

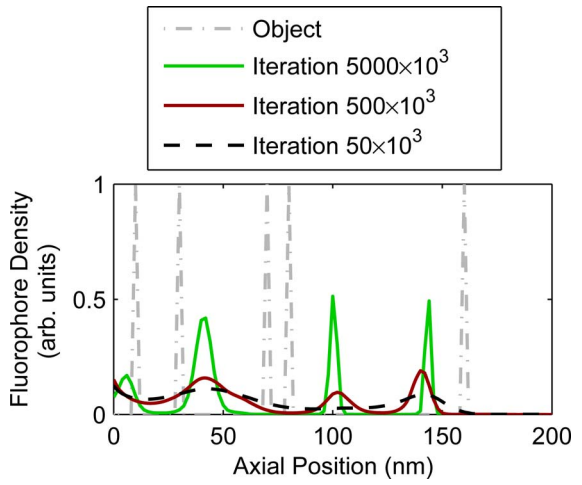

Fig. 5. (Color online) RL reconstructions at various iteration numbers for data from the test object of Fig. 3. Two excitation patterns were used, and the total expected count is 50×10^3 (for a SNR of approximately 23 dB).

more peaked with successive iterations. While the reconstruction does become more pointlike, it does not necessarily match the object well at these fine spatial scales.

The RL algorithm's behavior at different noise levels and with different excitation schemes is shown in Fig. 6.

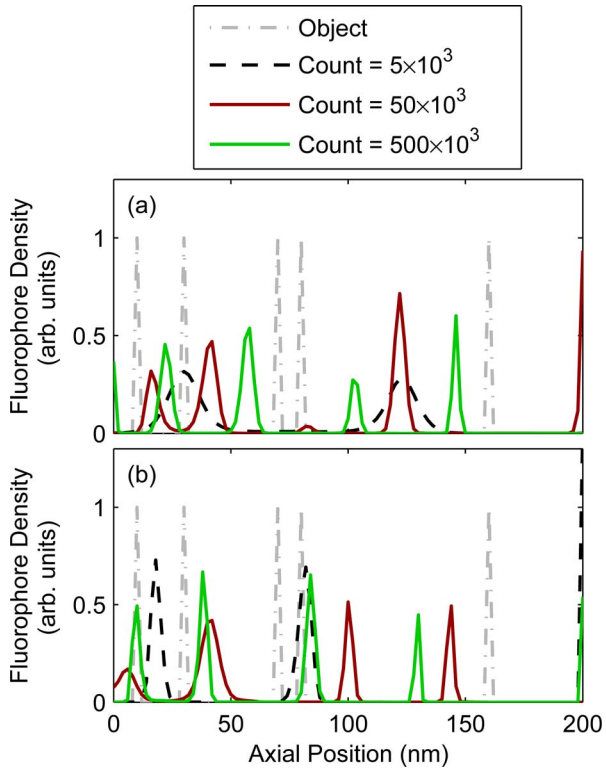

Fig. 6. (Color online) RL reconstructions for total expected counts of 5×10^3, 50×10^3, and 500×10^3 in the data. (a) A single-excitation and (b) a two-excitation case are considered. The corresponding SNRs are approximately 16, 26, and 36 dB for the single-excitation case and 13, 23, and 33 dB for the two-excitation case. In the single-excitation case the iteration numbers of the reconstructions are 1×10^5, 2×10^6, and 20×10^6 for the data of count 5×10^3, 50×10^3, and 500×10^3, respectively. The corresponding iteration numbers for the two-excitation case are 2×10^5, 5×10^6, and 50×10^6.

A representative iteration number has been chosen for each case. The algorithm converges faster for more noisy data, so this number typically increases with the photon count. The iteration number was chosen by picking a point at which the estimate was well localized. Since the RL algorithm preserves the integral of the estimate between iterations and because the magnitude of the reconstruction at a point cannot increase once it reaches zero, it is believed that the estimates will not change significantly beyond the iteration number shown.

Some general trends can be observed in Fig. 6—the two-excitation reconstructions are generally of a higher quality than the single-excitation reconstructions. The higher-signal reconstructions also tend to be better than their counterparts from noisy data. The erroneously high peaks toward the right of the plots can be explained by noting that at $z=200$ nm the excitation intensity is reduced by approximately 2 orders of magnitude compared with the value at $z=0$. Object features at higher z values therefore contribute minimally to the measured data, and the algorithm can place high-density features in that region without affecting the fidelity of the estimate to the measured data.

For the 500×10^3 count data and two excitations, the RL algorithm reasonably reconstructs the object: the first two peaks (starting from the left) are distinguished and well localized; the second two peaks are not resolved, but a single peak is reconstructed in the correct vicinity; while the final peak is not placed correctly. The poor reconstruction of the final peak is to be expected, as it is weakly illuminated and thus does not contribute strongly to the data. Unfortunately, the 500×10^3 count signal level is unrealistic for a five-fluorophore observation, and the estimate does not behave predictably as the signal is lowered. For example, for the 500×10^3 count data a peak is reconstructed at approximately 80 nm. When the signal is lowered to 50×10^3, the peak shifts but then returns at the 5000 count level. Behavior such as this leads to the conclusion that the RL algorithm cannot be used to reliably reconstruct SSM images on a tens-of-nanometers spatial scale.

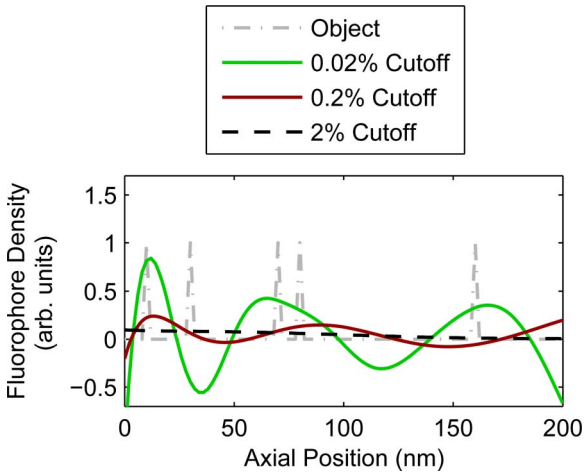

Fig. 7. (Color online) Truncated-singular-value-decomposition reconstructions of the object from two-excitation data with a total expected count of 50×10^3 (a SNR of approximately 23 dB). The singular value threshold is shown as a percentage of the maximum singular value.

3. Truncated Singular Value Decomposition Reconstruction

Other methods of reconstructing the object can be used, but there is no reason to expect they will achieve better performance than the RL algorithm. For example, Fig. 7 contains reconstructions found by using the truncated singular-value-decomposition approach [28]. This algorithm works by using the singular value decomposition to construct an approximate inverse to the operator $\mathbf{C}$. In order to achieve regularization, singular components passed with a strength below a user-defined threshold are not reconstructed. As this threshold is decreased, Fig. 7 shows that the solution moves from stable and smooth to oscillatory. This simple approach is not well matched to the problem here, as negative density estimates are given and peaked objects are not preferred. As a result, the reconstructions found are not particularly impressive. The truncated singular-value-decomposition approach does provide a linear noniterative reconstruction method and is given here as a point of comparison to the RL algorithm.

The mediocre performance of SSM for imaging an arbitrary object stands in contrast to its demonstrated nanometer-scale precision when localizing fluorophores known to reside in a single axial plane [10,12]. This difference, between the localization of a single molecule and the resolution when imaging a general distribution, has been thoroughly explored [8]. It has been shown that nanometer-scale localization precision can be expected from instruments that have standard resolution limitations when imaging an arbitrary object. High-precision localization has recently been experimentally demonstrated in a novel SSM instrument [29] that was also proved to be subject to standard resolution criteria.

When considering multiple molecules, the ability to distinguish two closely spaced fluorophores (when it is known that only two are present) is dependent only on the measurement noise rather than on standard criteria such as the Rayleigh resolution [30]. With this idea in mind, the ability of SSM to determine the axial positions and relative strengths of a strictly limited number of fluorophores will now be investigated.

B. *N*-Emitter Object

1. Theory

Under the correct circumstances it may be known *a priori* that not more than N fluorophores are present in the region from which emissions are collected. In this case the object has no more than N nonzero points. The object is therefore defined by $2N$ parameters—the N positions of the emitters and N fluorophore densities (one for each emitter). This complicates the inverse problem, as these $2N$ parameters are related to the object in a nonlinear manner.

To solve the inverse problem an exhaustive approach is taken. If the discretized object is defined on a spatial grid of M points, then there are $M!/[N!(M-N)!]$ possible sets of positions for the object. Given a set of N positions, the data are linearly related to the object densities. These N object densities can then be estimated by using standard methods for linear observations. For example, the RL algorithm can be applied as in the previous section. However, since the densities must be estimated for each possible set of positions, it is desirable to have a single-step, closed-form procedure to speed up the computation. This can be achieved by making an approximation to the Poisson noise model.

For reasonable photon counts the Poisson noise can be approximated with a Gaussian model. If the Gaussian model for the data noise has a covariance matrix of $\tilde{\mathbf{\Lambda}}$, then the maximum likelihood condition reduces to minimizing the exponent in the Gaussian density,

$$\check{\mathbf{s}} = \underset{\mathbf{s}}{\operatorname{argmin}}[\mathbf{d}-\mathbf{C}\mathbf{s}]^T\tilde{\mathbf{\Lambda}}^{-1}[\mathbf{d}-\mathbf{C}\mathbf{s}]. \tag{35}$$

The function to be minimized represents a weighted-squares cost function. The inverse weighting matrix $\tilde{\mathbf{\Lambda}}$ is the covariance of the data. This matrix is nonzero only on the main diagonal (since the noise in distinct pairs of data points is assumed independent) and, according to the Poisson model, equal to the expected data on the main diagonal. The expected data values are not known precisely, but, as can be seen from Fig. 2, the general shape is evident in the measured data. With this idea in mind, the main diagonal of $\tilde{\mathbf{\Lambda}}$ is defined by applying a low-pass filter to the measured data. Any values below a certain threshold may be set to that threshold to ensure that the weighting (which is determined by the inverse of $\tilde{\mathbf{\Lambda}}$) does not become overly large or negative.

The N-dimensional position space is scanned exhaustively (i.e., every possible combination of position values is investigated), and at each point Eq. (35) is solved. The Poissonian ML condition of Eq. (33) becomes the following within the accuracy of the Gaussian approximation:

$$\mathbf{C}^T\tilde{\mathbf{\Lambda}}^{-1}\mathbf{C}\check{\mathbf{s}} = \mathbf{C}^T\tilde{\mathbf{\Lambda}}^{-1}\mathbf{d}. \tag{36}$$

For the N-emitter problem, at a given set of positions, $\mathbf{C}$ has only N columns. Provided that N is relatively small, the minimization problem is overdetermined, and regularization need not be applied. This results in the following form for $\check{\mathbf{s}}$:

$$\check{\mathbf{s}} = [\mathbf{C}^T\tilde{\mathbf{\Lambda}}^{-1}\mathbf{C}]^{-1}\mathbf{C}^T\tilde{\mathbf{\Lambda}}^{-1}\mathbf{d}. \tag{37}$$

Thus for a given set of positions, the densities can be found by using the equation above. A cost [seen in Eq. (35)] can also be associated with this estimate. Comparing this cost across all sets of positions allows a minimum cost to be found. The positions and densities associated with this cost constitute the estimate of the object.

The N-emitter reconstruction algorithm described is guaranteed to find the ML estimate under the Gaussian noise approximation, since the entire allowable position-density space is searched in finding $\check{\mathbf{s}}$. The position variables are exhaustively searched, which can become computationally expensive for large M and/or a value of N greater than a few. However, the goal is to reconstruct the object only within the evanescent excitation volume. This limits the axial range considered and hence bounds M. The value of N used will depend on the application. It should be noted that N is actually an upper limit on the number of emitters, as samples containing fewer than N

emitters can be represented in this framework by having one or more of the object densities equal to zero. It may also be possible to improve the efficiency of the N-emitter reconstruction by using cost minimization algorithms that are more sophisticated than a simple exhaustive search [31].

2. Single Emitter

Estimating the axial position of a single fluorescent layer was the original application of SSM [9]. This section investigates the performance of $N=1$ estimation for a variety of noise levels. The position space is discretized into 0.5 nm steps, and for each position the optimal fluorophore density is calculated according to the method outlined in Subsection 3.B.1. The position and density corresponding to the least weighted-squares cost gives the fluorophore position and density estimate. This process was repeated for many realizations of the Poisson noise random variable, and the results are shown in Fig. 8. The results of Fig. 8 verify that nanometer-scale localization precision can be expected at both high and low signal levels. The fluorophore density is also estimated to within about 1%. It should also be noted that the signal levels investigated are consistent with single-molecule experiments. For example, a single Rhodamine 6G fluorescent molecule can be expected to emit of the order of 10^5 photons before photobleaching, when optimally excited [32]. If the total detection efficiency of the SSM instrumentation is 1%, then the resulting signal level is consistent with the 1000-count plots shown in Fig. 8. The signal level can be expected to be significantly higher for more photostable dyes or quantum dots.

3. Two Emitters

The next step is to evaluate the performance for two emitters. In this case the reconstruction process estimates four parameters—two positions and two fluorophore densities. The reconstructed positions will be investigated first. Estimated position pairs for the single- and two-excitation systems are shown in Fig. 9. Two positions are estimated in the reconstruction (one for each emitter), and this pair of numbers can represent a point on a two-dimensional plane. In Fig. 9 the intensity of a given point in the plane represents the number of times that pair of positions is estimated. As such, a well-localized plot represents a precise reconstruction, while a broadly scattered plot represents significant variability of the reconstructed emitter positions.

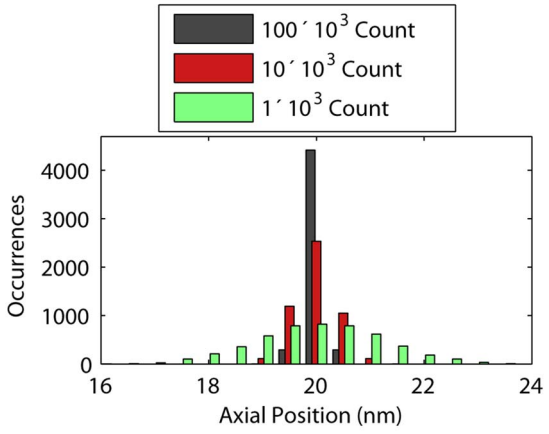

Fig. 8. (Color online) Histogram showing position estimates for a single fluorophore. Data with total expected counts of 1×10^3, 10×10^3, and 100×10^3 (SNRs of 10, 20, and 30 dB) are considered. In each case 5000 simulations were run, each with a different realization of the noise random variable. The true position of the emitter is at $z=20$ nm. The fluorophore was excited with one evanescent field only.

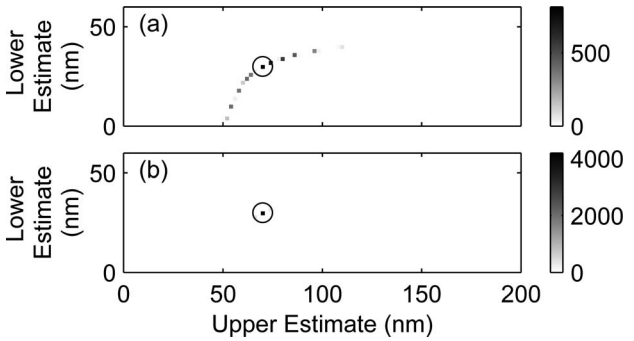

Fig. 9. Two-dimensional histograms showing the estimated position pairs from data corresponding to equal strength emitters at $z=30$ nm and $z=70$ nm. The pixel intensity at a given point represents the number of times that position pair was estimated. The ideal position pair estimate is marked with a circle. (a) Single-excitation and (b) two-excitation cases are considered, and the total expected count is 200×10^3 (giving SNRs of approximately 33 dB in the single-excitation data and 30 dB in the two-excitation data). For each case 5000 simulations were run, each with a different realization of the Poisson noise random variable.

Several conclusions can be drawn from the results seen in Fig. 9. The first is that the two-excitation case (i.e., using illumination plane waves at both 45° and 55°) performs much better than the single-excitation case. For this reason, only the two-excitation case will be considered for the remainder of this work. It can also be seen that the error in the position estimates is highly correlated. If one estimate is too high, then the other is also too high (this is evidenced by the fact that all estimated points are either on the upper right or lower left of the true point). Additionally, the estimated positions are never both on the same side of the midpoint—the lower estimate (i.e., the smaller of the two z estimates) never exceeds $z=50$ nm, and the upper estimate never falls below this point. Most important, it can be seen that for the two-excitation case, at this noise level, the reconstruction process accurately estimates the positions of the fluorophores.

Results for different noise levels and different emitter spacings are shown in Fig. 10. Similar forms of the position estimates can be seen. For poorly reconstructed positions, the estimate pairs are distributed along a narrow curve. This curve indicates that one estimate point is always above the midpoint and one below. The upper-left-hand plot appears dim, as most of the estimates are at extreme pixel points with either a lower estimate of 0 or an upper estimate of 200. For the noise levels considered, it can be seen that points separated by 80 nm are reliably estimated, while points separated by 10 nm are not well imaged. Estimates of emitter positions with spacings between these two extremes are dependent on noise level. To quantify this performance, the probability of the recon-

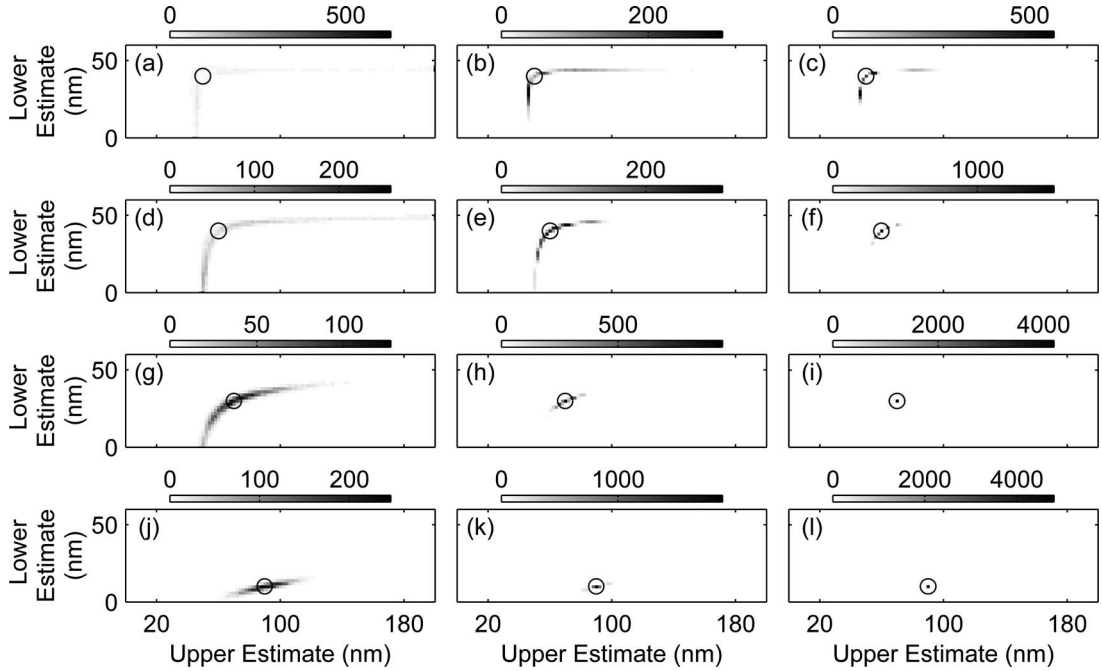

Fig. 10. Two-dimensional histograms of estimated pairs of positions, where a circle marks the true position pair. Various fluorophore spacings (constant across a row) and signal levels (constant down a column) are considered. The column containing plots (a), (d), (g), and (j) shows reconstruction statistics from data with a total expected count of 2×10^3; (b), (e), (h), and (k) are from data with a total expected count of 20×10^3; and (c), (f), (i), and (l) are from data with a total expected count of 200×10^3. Plots (a), (b), and (c) have a fluorophore spacing of 10 nm; (d), (e), and (f) have a spacing of 20 nm; (g), (h), and (i) have a spacing of 40 nm; and (j), (k), and (l) have a spacing of 80 nm. Because of the changes in the expected data, the SNR varies slightly with emitter spacing. However the SNRs are approximately 10, 20, and 30 dB for the data sets with count 2×10^3, 20×10^3, and 200×10^3, respectively.

structions meeting some quality criterion can be estimated by dividing the number of successful reconstructions by the total number of trials. In Fig. 11 the criterion used is that both position estimates are within ±2 nm of the true positions. This plot confirms that the reconstruction quality improves with signal level and increasing spacing between the emitters.

The plots of Figs. 9 and 10 display only the estimated positions, two of the four estimated parameters. The densities of the emitters are also estimated. The density estimates are reliable in the sense that any outlying points are estimated to have a low density. This phenomenon can be seen in Fig. 12. This figure considers emitters spaced 10 nm apart and with a total expected count of 20×10^3 (i.e., the conditions of the second plot in the first row of Fig. 10). The estimates in this scenario contain many cases where one position is estimated to be between the fluorophores and one a large distance from them. It can be seen from Fig. 12 that the outlying estimate has a low density, while the density estimate corresponding to the position estimate between the true positions is approximately 2 (accounting for two emitters of density 1). At the correct estimates of 40 and 50 nm, the density is estimated as approximately 1 for both points. This is in fact a general property of the estimate—the sum of the densities for the 5000 trials has a mean of 1.999 and a standard deviation of 0.021. This shows that the total density is estimated with high accuracy and precision re-

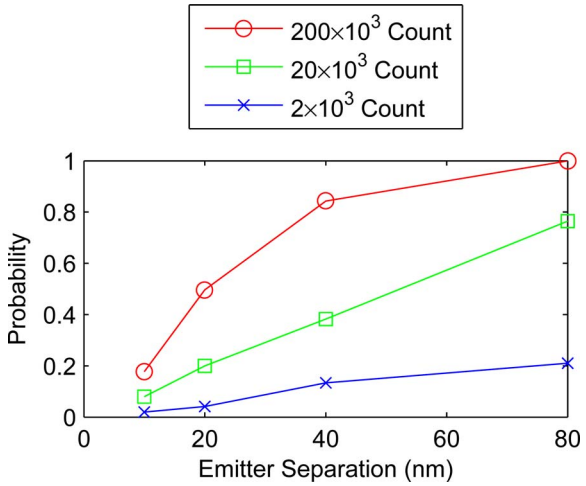

Fig. 11. (Color online) Estimated probabilities of both reconstructed positions being within ±2 nm of the true values, as a function of the spacing between emitters, and for various total expected counts.

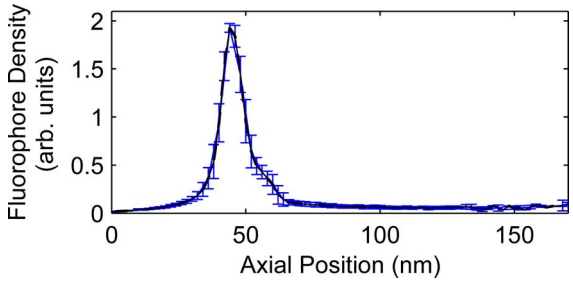

Fig. 12. (Color online) Mean fluorophore density estimate plotted as a function of the corresponding estimated position. The error bars show the standard deviation. The data are from an object with emitters at $z=40$ nm and $z=50$ nm, and the total expected count is 20×10^3 (a SNR of 20 dB).

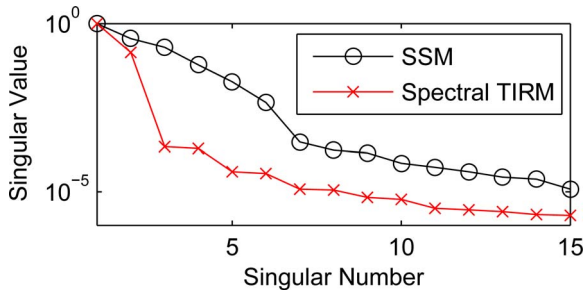

Fig. 13. (Color online) First 15 singular values for a two-excitation SSM operator and a two-excitation spectral TIRM operator plotted on a logarithmic scale. Both systems are defined on a spatial axis ranging from $z=0$ to $z=200$ nm, with a sampling period of 2 nm. Both plots are normalized to have a maximum singular value of 1.

gardless of where and how it is distributed. In short, Fig. 12, combined with the previous results, indicates that the emitters are either estimated correctly or are mistaken for a strong intermediate emitter with a weak outlier.

C. SSM and TIRM

As can be seen from Fig. 1, evanescently excited SSM is similar to a TIRM system. However, the introduction of the mirror and spectral detection has a dramatic effect on the collection of axial information from the sample. In TIRM, the amplitude of the collected signal is all that changes for different axial positions. In evanescently excited SSM, the mirror, the spectral detection, and the resulting spectral oscillations (see Fig. 2) give a much stronger axial-position dependence to the data. To illustrate the difference between SSM and TIRM, consider the one-dimensional SSM system used in the previous simulations and a similar system but without the underlying mirror. The second system is TIRM with spectral detection. The data from the TIRM system are given by an expression similar to the SSM system but absent the reflection term. Explicitly,

$$\bar{\mathbf{g}}(\mathbf{r}',k)=\bar{\mathbf{g}}^{\mathrm{Direct}}(\mathbf{r}',k), \tag{38}$$

as opposed to Eq. (10) for an SSM system. The information content of the two systems can be compared by analyzing the effective number of degrees of freedom, as measured by the number of singular values of the forward model that are above the noise floor. The singular values of these two systems are plotted in Fig. 13. It can be seen that the SSM system has larger singular values with a more gradual roll-off. These results indicate that more information is passed with SSM.

D. Estimation of the Spectral Envelope

The estimation procedures described in the previous section require the predicted data to be fitted to the measured data with great accuracy. For this reason, the estimation procedure may be expected to be sensitive to any inaccuracies in the model. In this section it is shown that a simple algorithm modification can compensate for uncertain model parameters without significant loss of image quality.

The fluorophore spectral envelope $A(k)$ may not be precisely known. Other measurement errors can also cause an effective change in $A(k)$, e.g., a nonuniform spectrometer gain. In addition, for multiple excitations, there may be a scaling difference in $A(k)$ across measurements if the illumination amplitude cannot be precisely controlled. For these reasons the spectral envelope is assumed to be unknown but is assumed to vary smoothly and relatively slowly with k.

This slow variation means that the spectral envelope can be written as a weighted sum of some relatively small number of smooth basis functions. If these basis functions are placed in the columns of a matrix $\mathbf{B}$ and the weighting coefficients placed in a vector $\boldsymbol{\kappa}$, then the discrete-k spectral envelope $\mathbf{A}$ can be written as

$$\mathbf{A}=\mathbf{B}\boldsymbol{\kappa}. \tag{39}$$

In this work an eight-member cubic-B-spline basis [33] forms the columns of $\mathbf{B}$.

A new model matrix $\mathbf{C}'$ can be calculated by leaving the effects of the spectral envelope out of the model. This results in the following observation equation:

$$\mathbf{d}=\mathrm{diag}[\mathbf{A}]\mathbf{C}'\mathbf{s}=\mathrm{diag}[\mathbf{B}\boldsymbol{\kappa}]\mathbf{C}'\mathbf{s}. \tag{40}$$

In this model both the envelope coefficients $\boldsymbol{\kappa}$ and the object vector $\mathbf{s}$ are unknown. The inverse problem is solved by iterating between estimating $\mathbf{s}$ with $\boldsymbol{\kappa}$ fixed and estimating $\boldsymbol{\kappa}$ with $\mathbf{s}$ fixed. This is a coordinate descent procedure [31].

Estimating the object with the envelope fixed can be done using the procedures mentioned earlier in this paper. The envelope coefficients can be estimated easily with the object held fixed. The problem is linear in $\boldsymbol{\kappa}$ and well conditioned due to the limited number of coefficients that need to be estimated. A weighted-least-squares approach [of the same form as Eq. (37)] can be employed. The iterative procedure is initialized by assuming a uniform envelope (i.e., every element of $\mathbf{A}$ is set to 1.)

This procedure was tested on data similar to that used earlier—the two-emitter object with fluorophores at z

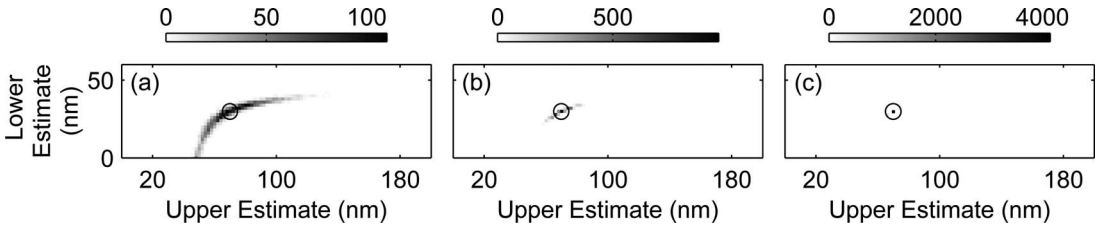

Fig. 14. Histogram of the position reconstructions for a two-emitter object with a 40 nm spacing between the objects. Total expected counts of (a) 2×10^3, (b) 20×10^3, and (c) 200×10^3 are considered. The envelope was not known *a priori* for these reconstructions but was estimated as part of the reconstruction procedure. This figure is equivalent to the third row of plots in Fig. 10 but with envelope estimation included.

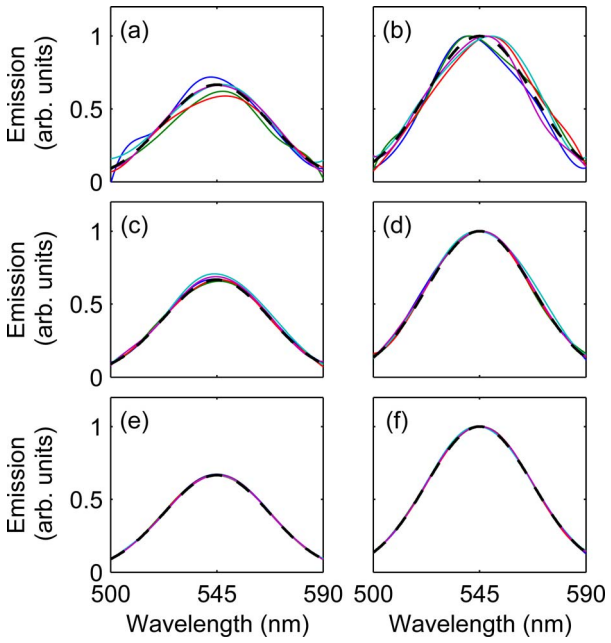

Fig. 15. (Color online) Five estimated envelopes (solid curves) and the true envelope (dashed curve) for each of the measurements in example two-excitation data sets. The first and second envelopes for data with a total expected count of 2×10^3 (SNR of 10 dB) are shown in (a) and (b) respectively; (c) and (d) are for a total expected count of 20×10^3 (SNR of 20 dB); and (e) and (f) are for a total expected count of 200×10^3 (SNR of 30 dB).

$=30$ nm and $z=70$ nm, and with a number of signal levels. An attenuation of 2/3 was applied to the first measurement to model unequal illumination strengths between the two excitations. The resulting position estimates are displayed in Fig. 14. Comparing Fig. 14 with the corresponding results in Fig. 10 (where the envelope was known *a priori*), it can be seen that there is little degradation caused by having estimated the envelope. It also shows that the unequal illumination did not cause a problem.

Some typical examples of the estimated envelopes are shown in Fig. 15. Here it can be seen that the envelopes are well estimated at the higher signal levels. For the lowest-count data there are some clear inaccuracies in the envelope estimate. This is consistent with Fig. 14, where the most visible degradation from Fig. 10 can be seen in the 2000-count data. It can be seen that the reconstruction procedure has estimated the unequal illumination strengths, as the two envelopes estimated have different scales. Again note that it was assumed only that the envelope was relatively smooth and could be represented by a fairly coarse spline basis.

4. CONCLUSIONS

This work presents a comprehensive model for spectral self-interference microscopy (SSM) that includes general expressions for the excitation and detection fields and the object to be imaged. This electromagnetic model includes the effects of anisotropic object response, the polarization state of the fields used, and an accurate representation of the detection optics. A method of reducing the model to a scalar representation is given. The model is used to characterize an evanescently excited SSM system and to perform image reconstruction from the resultant observations. The application of evanescent excitation to SSM allows the realization of the background suppression seen in TIRM, and the corresponding SNR gains.

The model was used to generate synthetic data, which were then employed in Monte Carlo simulations investigating the imaging capabilities of the system. The nanometer scale localization precision previously demonstrated with SSM was shown to be achievable at low signal levels consistent with single-molecule imaging. Investigations of a general object (i.e., with no restrictions on the number of fluorophores) showed that SSM could not provide reliable imaging on the nanoscale in this case. However, the capacity of two-excitation SSM to resolve two closely spaced emitters was demonstrated by using the prior knowledge that only two features were present. Finally, it was shown that when the emitted spectrum is not known *a priori*, it can be robustly estimated in conjunction with the object.

The results presented here verify that SSM is quite robust to variations in the fluorophore spectrum, and under the correct circumstances SSM can be expected to give good results in low-signal environments. Furthermore, the analysis of evanescent wave illumination suggests that a marriage of SSM and TIRM offers advantages over both existing modalities. Finally, the improved performance of the reconstruction algorithms when it is known *a priori* that only N emitters are present suggests that SSM is particularly well suited to dilute or single-molecule imaging, a previously unexplored application.

The image reconstruction methods and simulations presented were one dimensional and scalar. The one-dimensional treatment implicitly assumes that the axial and lateral spatial dimensions are separable, while the assumptions necessary for a scalar model were given in Subsection 2.B. While such simplifications are suitable for plane-wave illumination and low-NA detection, more complex systems can be analyzed by using the model presented in Subsection 2.A. For example, two counterpropagating evanescent waves could be used to create a lateral standing wave in the excitation pattern [13]. Such techniques have been shown to allow improved lateral resolution [34]. As discussed in Subsection 2.A, the model is also applicable to modalities that image a nonscalar object, but appropriate image reconstruction algorithms are needed for data interpretation. Topics such as structured illumination and nonscalar measurements in SSM are possible future directions for this technology.

ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation (grant DBI 0138425, award EEC-9987821 of the Engineering Research Centers Program, and grant DMR 0425780 to J. C. Schotland), the Air Force Office of Scientific Research (grant MURI F-49620-03-1-0379, grant FA9550-07-1-0096 to J. C. Schotland, and grant FA9550-07-1-0295 to W.C. Karl), and the National Institutes of Health (National Institute of Biomedical Imaging and BioEngineering grant 5R01 EB00 756-03).

REFERENCES

1. D. A. Agard, Y. Hiraoka, P. Shaw, and J. W. Sedat, "Fluorescence microscopy in three dimensions," Methods Cell Biol. **30**, 353–377 (1989).
2. J. Pawley, ed., *Handbook of Biological Confocal Microscopy* (Plenum, 1990).
3. K. I. Willig, S. O. Rizzoli, V. Westphal, J. Reinhard, and S. W. Hell, "STED microscopy reveals that synaptotagmin remains clustered after synaptic vesicle exocytosis," Nature **440**, 935–939 (2006).
4. D. Axelrod, N. L. Thompson, and T. P. Burghart, "Total internal reflection fluorescence microscopy," J. Microsc. **129**, 19–28 (1982).
5. T. Basché, W. E. Moerner, M. Orrit, and U. P. Wild, eds., *Single-Molecule Optical Detection, Imaging and Spectroscopy* (VCH, 1997).
6. S. Weiss, "Fluorescence spectroscopy of single biomolecules," Science **283**, 1676–1683 (1999).
7. A. Yildiz, J. N. Forkey, S. A. McKinney, T. Ha, Y. E. Goldman, and P. R. Selvin, "Myosin V walks hand-over-hand: single fluorophore imaging with 1.5-nm localization," Science **300**, 2061–2065 (2003).
8. R. J. Ober, S. Ram, and E. S. Ward, "Localization accuracy in single-molecule microscopy," Biophys. J. **86**, 1185–1200 (2004).
9. A. K. Swan, L. A. Moiseev, C. R. Cantor, B. Davis, S. B. Ippolito, W. C. Karl, B. B. Goldberg, and M. S. Ünlü, "Toward nanometer-scale resolution in fluorescence microscopy using spectral self-interference," IEEE J. Sel. Top. Quantum Electron. **9**, 294–300 (2003).
10. L. Moiseev, M. S. Ünlü, A. K. Swan, B. B. Goldberg, and C. R. Cantor, "DNA conformation on surfaces measured by fluorescence self-interference," Proc. Natl. Acad. Sci. U.S.A. **103**, 2623–2628 (2006).
11. M. Y. Ali, S. Uemura, K. Adachi, H. Itoh, K. Kinosita Jr., and S. Ishiwata, "Myosin V is a left-handed spiral motor on the right-handed actin helix," Nat. Struct. Biol. **9**, 464–467 (2002).
12. L. Moiseev, C. R. Cantor, M. I. Aksun, M. Dogan, B. B. Goldberg, A. K. Swan, and M. S. Ünlü, "Spectral self-interference fluorescence microscopy," J. Appl. Phys. **96**, 5311–5315 (2004).
13. P. S. Carney, V. Markel, and J. C. Schotland, "Near-field tomography without phase retrieval," Phys. Rev. Lett. **86**, 5874–5877 (2001).
14. C.-T. Tai, *Dyadic Green Functions in Electromagnetic Theory*, 2nd ed. (IEEE, 1993), Chaps. 11–7, pp. 251–254.
15. B. Richards and E. Wolf, "Electromagnetic diffraction in optical systems. II. Structure of the image field in an aplanatic system," Proc. R. Soc. London, Ser. A **253**, 358–379 (1959).
16. L. Novotny and B. Hecht, *Principles of Nano-Optics* (Cambridge U. Press, 2006), Chap. 3.6, pp. 61–66.
17. M. Böhmer and J. Enderlein, "Orientation imaging of single molecules by wide-field epifluorescence microscopy," J. Opt. Soc. Am. B **20**, 554–559 (2003).
18. R. Hassey, E. J. Swain, N. I. Hammer, D. Venkataraman, and M. D. Barnes, "Probing the chiroptical response of a single molecule," Science **314**, 1437–1439 (2006).
19. M. Nagorni and S. W. Hell, "Coherent use of opposing lenses for axial resolution increase in fluorescence microscopy. I. Comparative study of concepts," J. Opt. Soc. Am. A **18**, 36–48 (2001).
20. B. E. A. Saleh and M. C. Teich, *Fundamentals of Photonics* (Wiley, 1991), pp. 80–107.
21. L. M. Delves and J. L. Mohamed, *Computational Methods for Integral Equations* (Cambridge U. Press, 1985).
22. S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory* (Prentice Hall, 1993).
23. J. Markham and J.-A. Conchello, "Fast maximum-likelihood image-restoration algorithms for three-dimensional fluorescence microscopy," J. Opt. Soc. Am. A **18**, 1062–1071 (2001).
24. W. C. Karl, *Handbook of Image and Video Processing* (Academic, 2000), pp. 141–161.
25. W. H. Richardson, "Bayesian-based iterative method of image restoration," J. Opt. Soc. Am. **62**, 55–59 (1972).
26. L. B. Lucy, "An iterative technique for the rectification of observed distributions," Astron. J. **79**, 745–754 (1974).
27. R. Vio, J. Bardsley, and W. Wamsteker, "Least-squares methods with Poissonian noise: analysis and comparison with the Richardson–Lucy algorithm," Astron. Astrophys. **436**, 741–755 (2005).
28. P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems* (SIAM, 1998), pp. 45–68.
29. B. J. Davis, M. Dogan, B. B. Goldberg, W. C. Karl, M. S. Ünlü, and A. K. Swan, "4Pi spectral self-interference microscopy," J. Opt. Soc. Am. A (to be published).
30. S. Ram, E. Ward, and R. Ober, "Beyond Rayleigh's criterion: a resolution measure with application to single-molecule microscopy," Proc. Natl. Acad. Sci. U.S.A. **103**, 4457–4462 (2006).
31. D. P. Bertsekas, *Non-Linear Programming*, 2nd ed. (Athena Scientific, 1999).
32. C. Eggeling, J. Widengren, R. Rigler, and C. A. M. Seidel, "Photobleaching of fluorescent dyes under conditions used for single-molecule detection: evidence of two-step photolysis," Anal. Chem. **70**, 2651–2659 (1998).
33. C. De Boor, *A Practical Guide to Splines* (Springer, 2001).
34. M. G. L. Gustafsson, "Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy," J. Microsc. **198**, 82–87 (2000).