Journal of
Microscopy

Journal of Microscopy, Vol. 247, Pt 2 2012, pp. 147–160
Received 14 January 2012; accepted 30 March 2012

doi: 10.1111/j.1365-2818.2012.03625.x

Fluorescence excitation and imaging of single molecules near dielectric-coated and bare surfaces: a theoretical study

DANIEL AXELROD
Departments of Physics, Biophysics, and Pharmacology, University of Michigan, Ann Arbor, Michigan, U.S.A.

Key words. Fluorescence microscopy, high aperture, molecular orientation, polarization, point spread function, thin films, total internal reflection.

Summary

Microscopic fluorescent samples of interest to cell and molecular biology are commonly embedded in an aqueous medium near a solid surface that is coated with a thin film such as a lipid multilayer, collagen, acrylamide, or a cell wall. Both excitation and emission of fluorescent single molecules near film-coated surfaces are strongly affected by the proximity of the coated surface, the film thickness, its refractive index and the fluorophore's orientation. For total internal reflection excitation, multiple reflections in the film can lead to resonance peaks in the evanescent intensity versus incidence angle curve. For emission, multiple reflections arising from the fluorophore's near field emission can create a distinct intensity pattern in both the back focal plane and the image plane of a high aperture objective. This theoretical analysis discusses how these features can be used to report film thickness and refractive index, and fluorophore axial position and orientation.

Introduction

Microscopic fluorescent samples of interest to cell and molecular biology are almost always in an aqueous medium near a solid surface. Frequently, that surface is coated with a thin film such as: a lipid monolayer, bilayer, or multilayer; a collagen or agarose layer deposited before cell plating or deposited by the cells themselves; acrylamide gel to immobilize beads or single molecules (such as in commercial preparations for nucleic acid sequencing); or a cell wall interposed between the substrate and cellular organelle. The optical properties of fluorescent single molecules near film-coated surfaces for both excitation and emission are strongly affected by the proximity of the coated surface, the film thickness, its refractive index and the molecular orientation. Conversely, these optical properties potentially can be used to calculate distances to the surface, 3D molecular orientation, and thickness, irregularity and optical density of the film. This paper provides a theoretical basis and practical guide to such calculations.

The mathematics of fluorescence excitation and emission are independent of each other. However, some of the effects discussed here conceptually connect the two. When the incidence angle (relative to the substrate normal) of excitation light propagating through the substrate is increased beyond the critical angle for total internal reflection (TIR), the transmitted field in the aqueous medium becomes evanescent (exponentially decaying). Likewise, when emitted light energy from a single molecule interacts with a nearby substrate, the only contribution to the intensity transmitted into the substrate at large angles to the normal is the 'near field' of the emitting fluorophore, which is also evanescent and decays with distance from the fluorophore. In both cases, these evanescent fields are strongly affected by the presence of a film coating; the effect of the film coating on the optics is a major theme of this paper. Excitation and emission are discussed separately, while pointing out the analogies where they exist.

In the case of TIR excitation, the intensity of the evanescent field in the aqueous medium can be enhanced or decreased by the presence of an intermediate film. Enhancement occurs for a metallic film, due to surface plasmon resonance (Ekgasit *et al.*, 2004 and 2005a,b; Le Moala *et al.*, 2007; Wood *et al.*, 2006; Acher *et al.*, 2000; Feng *et al.*, 2005). But even for a dielectric film, resonant enhancement can occur if the film (or some layers in a multilayer structure) acts as a waveguide (Chiu *et al.*, 2008; Nesnidal & Walker, 1996; Challener *et al.*, 2000; Kaiser *et al.*, 1994; Ke *et al.*, 1997). In this paper, the effect of a single dielectric film sandwiched between glass and water is emphasized, with an intermediate refractive index as is typical for the practical cases listed above.

In the case of emission from a fluorophore and its imaging in a microscope, the simplest and least accurate approach is to completely ignore molecular orientation and the presence of any surfaces, and just assume that a fluorophore emits light

Correspondence to: Daniel Axelrod, Department of Physics, University of Michigan, Ann Arbor, MI 48109. Tel: +530-938-1859; e-mail: daxelrod@umich.edu

in a spherically isotropic manner. This crude approximation, most appropriate only for low apertures or for groups of fluorophores with an isotropic orientational distribution, leads to the standard 'Airy disk' intensity pattern for the point spread function (PSF) in the image plane (Gu, 2000). The Airy disk intensity pattern (which involves a Bessel function of the radial distance from the optical axis) will never be seen through a high-aperture objective from a real single oriented molecule.

One step up in theoretical accuracy is to treat the fluorophore as a classical dipole emitter, either isolated or bathed in the back-reflection of its own field from an interface. These approaches give information about molecular orientation (Foreman *et al.*, 2008; Cyphersmith *et al.*, 2011; Fourkas, 2001; Patra *et al.*, 2004 ; Ishitobi *et al.*, 2010; Holbein & Hübner; 2008). But they do not account for the near field interaction with a (possibly) coated substrate nor the additional information the interaction contains about substrate distance and film thickness and refractive index.

Numerous previous works have considered various aspects of the effect on emission of nearby dielectric surfaces in which near field effects are included. The classic papers of Chance *et al.* (1978) and Ford & Weber (1984) examine the total power transfer from a fluorophore (modelled as an oscillating dipole) into a film-coated surface, with particular emphasis on metal films. Fattinger & Lukosz (1984) and Burghardt & Thompson (1984) considered emission from fluorophores located within a thin intermediate layer sandwiched between two dielectrics. Subsequently, Hellen & Axelrod (1987) considered the angular emission pattern of a fluorophore in the low-density medium above a higher density substrate coated by a film of metal or intermediate refractive index dielectric. Mertz (2000) used a somewhat different theoretical method for the case of bare glass without film. Courtois *et al.* (1996) approached the problem with a rigorous quantum mechanical analysis, rather than the more usual classical electromagnetic radiating dipole analysis, and showed the relationship between the two. A general theory for fluorophores embedded anywhere in a stack of dielectric layers of arbitrary number and thickness has been presented by Polerecky *et al.* (2000). None of these papers considered the resulting theoretical light pattern that would be seen at the back focal plane (BFP) or image plane of a high aperture objective. Enderlein *et al.* (1999) also considered emission at bare glass, but proposed using a high-aperture parabolic mirror collector rather than a high aperture objective.

With specific attention to high aperture microscopy (including near field effects), Dickson *et al.* (1998) compared theoretical PSF images with actual observations of oriented molecules immobilized at a surface. Aguet *et al.* (2009) derived a procedure for fitting PSF image patterns to derive 3D orientations of fluorophores on bare glass. Intensity patterns at the objective's BFP rather than the image plane were examined by Lieb *et al.* (2004) for a fluorophore on bare (or negligibly thin film coated) glass, by Burghardt & Ajtai (2009) and by Matthe*yse*s & Axelrod (2005) for metal (but not dielectric) coated glass, by Axelrod (2001) specifically for wide-field (non-single molecule) imaging and by Ruckstuhl & Verdes (2004) for confocal scanning imaging. Bohmer & Enderlein (2003) predicted the pattern (including near field effects) at images that are slightly defocused to more vividly show fluorophore orientation. Burghardt (2011) subsequently extended the analysis to prediction and fitting of the pattern at the image plane and applied it to oriented myosin in muscle crossbridges. Sikorski & Davis (2008) described a method of electrooptically modifying the light phase at the BFP for fluorophores near bare glass to produce images that are fluorophore orientation sensitive.

This theoretical paper emphasizes the effect of a thin dielectric film coated on the glass coverslip upon a single molecule's emission phase and intensity at the microscope's BFP and image plane (where the intensity forms the point spread function), including the effects of the fluorophore's near field. These intensity patterns can be used to retrieve information about local film thickness and refractive index, and about fluorophore axial position and orientation. The zero-thickness film case (i.e. bare glass) is a natural subset of the results.

TIR excitation through a dielectric film-coated substrate

Basic theory

Consider a laterally infinite planar stack of three dielectric materials where the z-axis is in the direction normal to the planes, with the intermediate layer of (non-negative) thickness h. The x-axis is parallel to the interfaces and in the plane of incidence containing the incident, reflected and refracted rays; the y-axis is also parallel to the interfaces but is perpendicular to the plane of incidence. Refractive index n_3 (presumably glass) occupies the semi-infinite z-range $(-\infty,-h)$; refractive index n_2 (presumably a film coating the glass) occupies the z-range $(-h,0)$; and refractive index n_1 (presumably water) occupies the z-range $(0, \infty)$. The identification of the first medium traversed as 3 rather than 1 is to assure consistency with the notation of Hellen & Axelrod (1987), where emitted light travels in the reverse direction, i.e. from a fluorophore in medium 1 toward medium 3. For this section, the excitation light (a monochromatic plane wave of vacuum wavelength l) travels in the x–z plane from medium 3 toward medium 1 at an incidence angle of θ_3 with respect to the normal as measured in medium 3.

For the common circumstance where $n_3 > n_2 > n_1$ (see Fig. 1), the system *as a whole* will support total internal reflection somewhere between media 3 and 1 only if $\theta_3 > \theta_{31}^c$ where

$$\theta_{31}^c = \sin^{-1}(n_1/n_3). \quad (1)$$

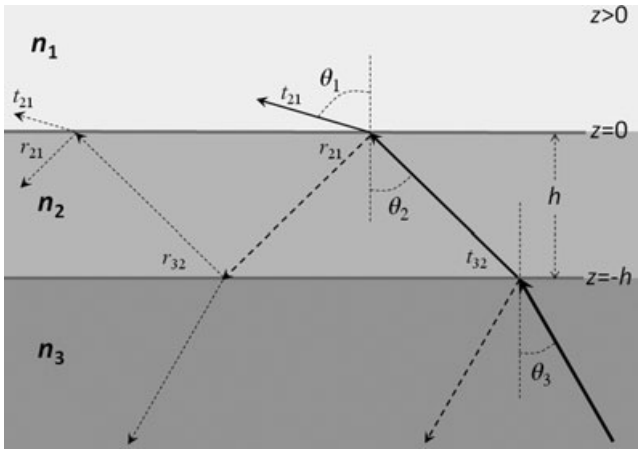

Fig. 1. Ray diagram for excitation light traversing a three layer system, including multiple reflections, showing the definition of variables. Fresnel coefficients for reflection (r) and transmission (t) are defined in Eq. (3). The circumstance where $n_3 > n_2 > n_1$ and no TIR occur at either surface is depicted. For larger incidence angles than shown, the system as a whole will support total internal reflection only for $\theta_3 > \theta_{31}^c$. If $\theta_{32}^c > \theta_3 > \theta_{32}^c$ TIR occurs at the 2:1 interface and $\cos\theta_1$ becomes imaginary. If $\theta_3 > \theta_{32}^c$ TIR occurs at the 3:2 interface and both $\cos\theta_2$ and $\cos\theta_1$ become imaginary.

TIR will occur at the *first* encountered interface (3:2) if $\theta_3 > \theta_{32}^c$ where

$$\theta_{32}^c = \sin^{-1}(n_2/n_3). \tag{2}$$

TIR will occur at the *second* encountered interface (2:1) if θ_3 is in the intermediate range $\theta_{32}^c > \theta_3 > \theta_{31}^c$. This θ_3 range between the two critical angles produces nontrivial effects on intensity in both media 1 and 2 because total reflection of light back into medium 2 at the 2:1 interface coupled with subsequent partial reflection of that light back again into medium 2 at the 2:3 interface leads to interference effects in the film. The film acts like a 'leaky' waveguide (not a full waveguide in which light is trapped by TIR at both surfaces), but it still exhibits lossy resonances at particular θ_3 angles depending on film thickness h and refractive index n_2.

To calculate the intensities in media 2 and 1 as a function of θ_3, expressions are written for the electric field strengths based on the well-known Fresnel coefficients (Born & Wolf, 1975, Chap. 1] for transmission $t_{ij}^{p,s}$and reflection $r_{ij}^{p,s}$for a plane wave traversing once through a single interface *from* medium i *to* medium j for *p*-pol or *s*-pol light. (Polarizations are defined as the orientation of the electric field relative to the plane of incidence which contains the incident, reflected and refracted rays: *s*-pol and *p*-pol are perpendicular and parallel, respectively, to the plane of incidence.)

$$t_{ij}^{p} = \frac{2n_i\cos\theta_i}{n_i\cos\theta_j + n_j\cos\theta_i}$$

$$t_{ij}^{s} = \frac{2n_i\cos\theta_i}{n_i\cos\theta_i + n_j\cos\theta_j}$$

$$r_{ij}^{p} = \frac{n_j\cos\theta_i - n_i\cos\theta_j}{n_i\cos\theta_j + n_j\cos\theta_i} \tag{3}$$

$$r_{ij}^{s} = \frac{n_i\cos\theta_i - n_j\cos\theta_j}{n_i\cos\theta_i + n_j\cos\theta_j}$$

where, from Snell's Law,

$$\cos\theta_i = \sqrt{1-\sin^2\theta_i} = \sqrt{1-(n_3/n_i)^2\sin^2\theta_3}. \tag{4}$$

For $n_3 > n_2 > n_1$, $\cos q_i$ is imaginary if θ_3 is larger than the TIR critical angles given by Eqs (1) or (2), respectively.

In the simple one-interface case in going from medium i to medium j, the x, y and z components of the incident complex electric field are multiplied by the appropriate Fresnel coefficients to obtain each component of the complex transmitted and reflected electric fields (Born & Wolf, 1975). But in the leaky-waveguide case considered here, the incident plane wave necessarily experiences multiple reflections in medium 2. Each reflection at (2:3) and (2:1) introduces a multiplicative factor of a Fresnel coefficient and each subsequent traversal through medium 2 introduces a propagation phase factor depending on the thickness and refractive index of medium 2. The result is a convergent geometric series describing the total electric field for medium 2 and for medium 1). Each series can be summed into a closed form and separated into *p*-pol components E_{ix} and E_{iz} and *s*-pol component E_{iy} as follows (see Born & Wolf, 1975, but with reversal of medium 1 and 3 nomenclature).

For $-h < z < 0$ (i.e. within medium 2):

$$E_{2x}(z,\theta_3) = \frac{t_{32}^{p}e^{i\beta_2 h}\left(e^{i\beta_2 z} - r_{21}^{p}e^{-i\beta_2 z}\right)\cos\theta_2}{1 - r_{21}^{p}r_{23}^{p}e^{2i\beta_2 h}}$$

$$E_{2y}(z,\theta_3) = \frac{t_{32}^{s}e^{i\beta_2 h}\left(e^{i\beta_2 z} + r_{21}^{s}e^{-i\beta_2 z}\right)}{1 - r_{21}^{s}r_{23}^{s}e^{2i\beta_2 h}} \tag{5}$$

$$E_{2z}(z,\theta_3) = \frac{t_{32}^{p}e^{i\beta_2 h}\left(e^{i\beta_2 z} + r_{21}^{p}e^{-i\beta_2 z}\right)\sin\theta_2}{1 - r_{21}^{p}r_{23}^{p}e^{2i\beta_2 h}}.$$

For $z > 0$ (i.e. within medium 1),

$$E_{1x}(z,\theta_3) = \frac{t_{32}^{p}t_{21}^{p}(e^{i\beta_2 h}e^{i\beta_1 z})\cos\theta_1}{1 - r_{21}^{p}r_{23}^{p}e^{2i\beta_2 h}}$$

$$E_{1y}(z,\theta_3) = \frac{t_{32}^{s}t_{21}^{s}e^{i\beta_2 h}e^{i\beta_1 z}}{1 - r_{21}^{s}r_{23}^{s}e^{2i\beta_2 h}} \tag{6}$$

$$E_{1z}(z,\theta_3) = \frac{t_{32}^{p}t_{21}^{p}(e^{i\beta_2 h}e^{i\beta_1 z})\sin\theta_1}{1 - r_{21}^{p}r_{23}^{p}e^{2i\beta_2 h}}.$$

Parameters b_1 and b_2 are the phase lag angle per unit distance along z incurred by propagation through medium 1 and 2,

respectively.

$$\beta_1 = (2\pi n_1 \cos\theta_1)/\lambda$$
$$\beta_2 = (2\pi n_2 \cos\theta_2)/\lambda. \quad (7)$$

If either $\cos\theta_1$ or $\cos\theta_2$ is imaginary (corresponding to evanescent rather than propagating light in the corresponding medium), the geometrical picture of propagation does not apply, but the mathematical formalism still does.

The evanescent intensities I in medium i can be reported in the two polarizations p and s:

$$I_{ip}(z,\theta_3) = |E_{ix}|^2 + |E_{iz}|^2$$
$$I_{is}(z,\theta_3) = |E_{iy}|^2. \quad (8)$$

For any $\theta_3 > \theta_{31}^c$, the evanescent intensity in medium 1 (i.e. the zone for which $z > 0$) exponentially decays in z with a well-known decay rate that increases with θ_3 but does not depend at all upon film thickness h or refractive index n_2:

$$I_{1p,s}(z,\theta_3) = I_{1p,s}(0,\theta_3)e^{-(\frac{4\pi z}{\lambda})\sqrt{n_3^2\sin^2\theta_3 - n_1^2}}. \quad (9)$$

Intensity at the film surface

The intensity $I_{1p,s}(0,\theta_3)$ of the exponential decay (Eq. 9) at the immediate surface of the film in medium 1 is particularly interesting in the intermediate zone $\theta_{32}^c > \theta_3 > \theta_{31}^c$, in which the incident light propagates within the film and is evanescent in medium 1. Fig. (2) shows $I_{1s}(0,\theta_3)$ (normalized to the incident intensity) versus θ_3 for various h, for a typical set of refractive indices that might be encountered in biochemical or cell biological samples. (The p-pol $I_{1p}(0,\theta_3)$ is not shown but is qualitatively similar).

For the case of no film ($h = 0$), the evanescent intensity slowly decreases monotonically with increasing θ_3, reaching exactly zero at $\theta_3 = 90^o$. But in the presence of a film ($h > 0$), the evanescent intensity $I_{1s}(0,\theta_3)$ shows a strong dependence on h, reaching maxima at particular θ_3 values in the intermediate zone that can exceed the evanescent intensity in the $h = 0$ (no-film) case by a factor of more than 2. This resonance-like behaviour results from the multiple reflections in the film. It is analogous to the classic effects of oil film on water or antireflection coating on glass, except that the amplitude of an evanescent field rather than a propagating reflection or transmission is examined here. The exact conditions for constructive or destructive interference are complicated because the phase shifts at the medium 2: medium 1 TIR interface are neither exactly zero nor exactly 180^o but instead are intermediate depending upon θ_2, n_1 and polarization.

Several practical consequences arise from the effects depicted in Fig. (2), as follows:

(a) The existence of a film (which otherwise might not have been known) should be obvious from the shape of $I_{1s}(0,\theta_3)$ vs. θ_3 within the $\theta_{32}^c > \theta_3 > \theta_{31}^c$ range, as might be reported by emission from a fluorophore in medium 1 adsorbed to the medium 2 surface at $z = 0$. (At any fixed fluorophore position z away from the interface in medium 1, the excitation intensity and the observed fluorescence intensity are exactly proportional to each other in their θ_3 dependence, but the z-dependences are different because of near field emission effects discussed later.) For thin films (e.g. with $h = 0.1\lambda$), $I_{1s}(0,\theta_3)$ monotonically decreases but significantly more rapidly than for the no-film case. For thick films ($h > 0.6\lambda$), the $I_{1s}(0,\theta_3)$ curves are no longer monotonic decays and show at least one clear maximum in the $\theta_{32}^c > \theta_3 > \theta_{31}^c$ range, which is qualitatively different from the no-film case.

(b) For thicker films ($h > 0.6\lambda$), the thickness range should be readable from the number of maxima, and (if film

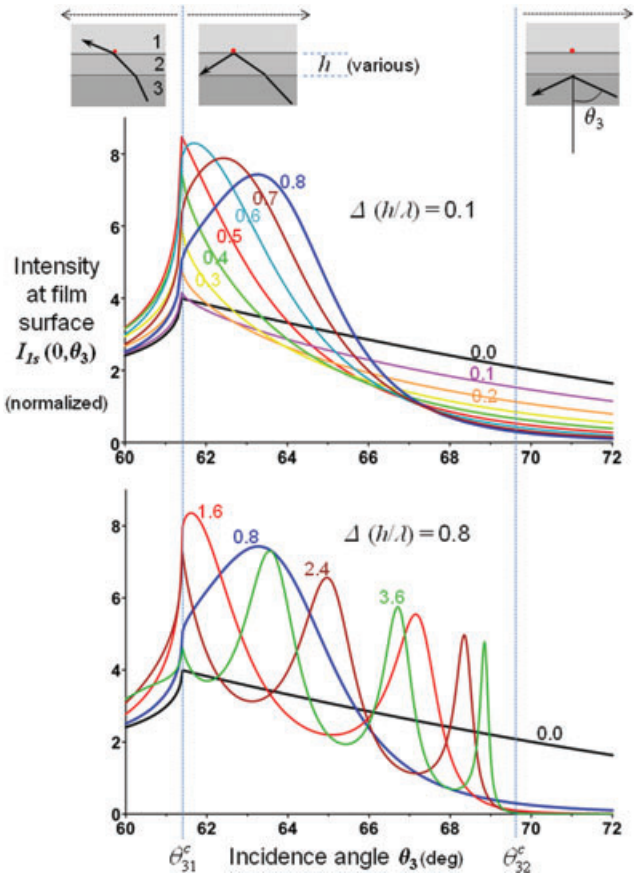

Fig. 2. $I_{1s}(0,\theta_3)$ (normalized to the incident intensity) versus θ_3 for various h, for a typical set of refractive indices that might be encountered in biochemical or cell biological samples ($n_1 = 1.33; n_2 = 1.42; n_3 = 1.515$). The three zones (from left to right) demarked by dashed lines along the θ_3 axis correspond to $\theta_3 < \theta_{31}^c$ (no TIR); $\theta_{32}^c > \theta_3 > \theta_{31}^c$ (TIR at 2:1) and $\theta_3 > \theta_{32}^c$ (TIR at 3:2). The upper and lower graphs correspond to finer and courser intervals in h/λ, respectively. Data for all graphs and images in all figures were generated by a custom program written in interactive data language (IDL). [Correction added after online publication 13 July 2012: Curve labelled 3.6 and curve labelled 2.4 were connected]

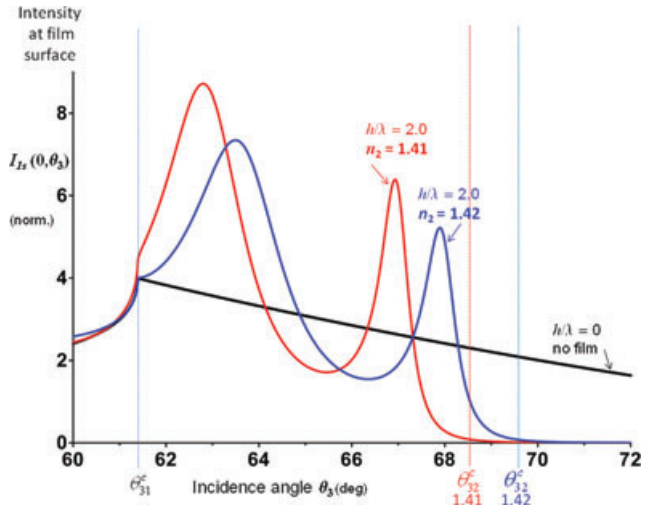

Fig. 3. $I_{1s}(0, \theta_3)$ (normalized to the incident intensity) vs θ_3 for two (slightly) different film refractive indices, $n_2 = 1.41$ (red) and $n_2 = 1.42$ (blue). In both these cases, $h = 2\lambda$, $n_1 = 1.33$ and $n_3 = 1.515$. The no-film case ($h = 0$, black) is also shown.

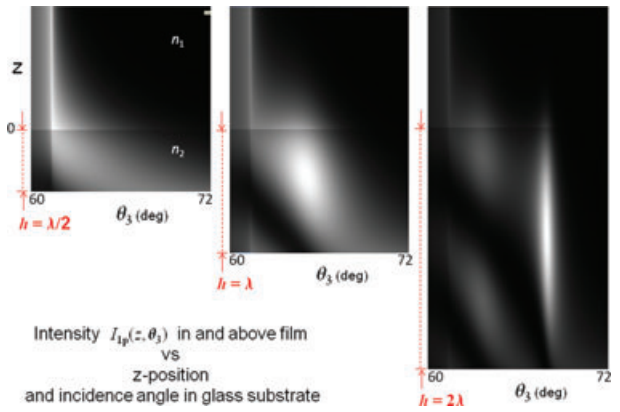

Fig. 4. The intensity I_{1p} inside ($z < 0$) and above ($z > 0$) the film, shown in greyscale as a function of θ_3 (abscissa) and z (ordinate) for three different film thicknesses ($h = \lambda/2$, λ, and 2λ). In all cases, $h = 2\lambda$, $n_1 = 1.33$, $n_2 = 1.42$ and $n_3 = 1.515$. [Correction added after online publication 13 July 2012: I_{1s} has been changed to I_{1p}.]

refractive index is known) the exact thickness can be deduced from the location of maxima in $I_{1s}(0, \theta_3)$.

(c) A film can boost the evanescent intensity in medium 1 by at least a factor of two over its no-film value at the same incidence angle, by a judicious choice of θ_3. This evanescent enhancement may be useful where laser power is marginal or where more rapid fluorophore photobleaching is desired.

(d) The effects are qualitatively similar for both *p*-pol and *s*-pol incident light (only *s*-pol is shown in Fig. 2). The only difference is a very slightly lower contrast between maxima and minima and a very slight shift ($\sim 0.1^\circ$) of maxima to smaller angles in the *p*-pol case. Larger-than-expected experimental deviations between $I_{1s}(0, \theta_3)$ and $I_{1p}(0, \theta_3)$ may suggest birefringence in the film.

(e) If an experimental measurement of $I_{1p,s}(0, \theta_3)$ does not match any curves typified by those in Fig. (2), one can strongly conclude that sample surface does indeed support a film and that it must be heterogeneous in thickness and/or refractive index.

(f) The $I_{1p,s}(0, \theta_3)$ curves are sensitive to the film refractive index n_2, especially for higher film thicknesses in which these curves are non-monotonic. For example, decreasing n_2 by only 0.01 from 1.42 to 1.41 (at a constant $h = 2\lambda$) shifts both prominent maxima of $I_{1s}(0, \theta_3)$ to lower θ_3 values by about 1°. (Fig. 3 shows this shift in detail.) This effect is not due to a simple change in the optical path length of the film, which is proportional to the product $n_2 h$. A decrease in n_2 can be distinguished from a similar-factored increase in h; these two parameters affect $I_{1s}(0, \theta_3)$ differently. Therefore, it should be possible to curve-fit a theoretical $I_{1s}(0, \theta_3)$ to experimental data to unambiguously determine both n_2 and h simultaneously.

(g) $I_{1p,s}(0, \theta_3)$, curves are also sensitive to variations in n_1, for higher film thicknesses. This sensitivity may not be interesting in most cases because n_1 is typically known and is also homogeneous. However, if the 2:1 interface is overlaid with a heterogeneous structure (such as adsorbed cells), then different lateral locations in the sample will exhibit locally different $I_{1p,s}(0, \theta_3)$ curves. Scanning θ_3 while recording wide-field images should reveal a new contrast mechanism based on absolute n_1 rather than gradients in n_1.

(h) Several studies have appeared in which $I_{1p,s}(0, \theta_3)$ data are used to determine the *z*-location of a fluorophore in medium 1 or to deduce a concentration profile in the *z*-dimension (Oheim *et al.*, 1998; Ölveczky *et al.*, 1997). These studies assumed that there was no film present at the TIR interface. However, even a thin (perhaps cell-secreted collagen) film of $h = 0.1\ \lambda$, and clearly thicker films, will distort the curves and give rise to a possibly erroneous conclusion about fluorophore location.

(i) To determine the thickness of a dielectric film, a previously published TIRF method presents an alternative to examination of the $I_{1p,s}(0, \theta_3)$ vs. θ_3 curve as discussed here. Benešová & Tománek (1999) compare two measurements of $I_{1p,s}(0, \theta_3)$ at a fixed θ_3, one measurement at a region on the surface supporting the film and one at a region (if such conveniently exists) where the surface is bare. In that study, a proxy for $I_{1p,s}(0, \theta_3)$ at a fixed θ_3 is used: the total emission observed from TIR-excited dissolved fluorophores in the liquid.

Intensity inside the film

To show the intensity inside the film, Fig. (4) displays intensity information as brightness on a 2D (θ_3,z) plane. Boundary

conditions imposed by Maxwell's equations lead to continuity (regardless of refractive index ratio) of E_x and E_y across the 2:1 ($z = 0$) interface, and a ratio $(n_2/n_1)^2$ in E_z in traversing from medium 2 to medium 1. However, for particular θ_3, z and h/λ combinations, the brightest intensities are found deeper within the film. If h/λ and n_2 are known, then the shape of $I_{1p,s}(z < 0, \theta_3)$ could indicate the z-position of a fluorophore embedded in medium 2.

Emission through a dielectric film-coated substrate

The optics of both TIR excitation of a surface-proximal fluorophore and its subsequent emission involve evanescent fields and both set up a resonance in the film. Since a fluorophore, modelled as a classical dipole emitter, is much smaller than the wavelength of the emitted light, the emitted field cannot be described solely by outwardly propagating light rays (each ray representing a plane wave in a spectrum of directions). If that were the case, reversing time and tracing these plane waves backward to their source would result only in a blurry spot approximately one half the wavelength in radius, not a sharp infinitesimal spot. A 'near field' accounts for the actual sharpness of the spot: the sum of a continuum of evanescent fields decaying exponentially with distance from the fluorophore and with closely spaced wavefronts perpendicular to the radial direction. Some of these evanescent fields originating at the fluorophore interact with the film-coated surface nearby and convert into propagating light, which can then set up resonances in the film, much like excitation light.

The emission theory here assumes a fluorophore is located in a refractive index n_1 medium, at a distance z from a film coated surface. As in the theory for excitation, the film thickness is h with refractive index n_2. The underlying substrate, with refractive index n_3 extends all the way to a high aperture objective. (In practice, the substrate is a glass coverslip with index-matched immersion oil in the space between the bottom of the coverslip and the objective.) As before, assume the common case where $n_3 > n_2 > n_1$. The objective is located millions of wavelengths away from the fluorophore so it only captures propagating light. However, some of that captured light originates as an evanescent near field that was subsequently converted into propagating form by the interfaces. For simplicity, only a fluorophore on the optical axis is considered.

In terms of ray optics, the objective 'sees' emission light rays emanating approximately radially from the fluorophore and refracts those rays into a direction parallel to the optical axis (for an infinity-corrected objective and dipole in the sample plane). These rays travel through the BFP and encounter the tube lens, which reconverges them to the image plane. The expected emission electric field and intensity patterns at the BFP and at the image plane are derived below, with variables as depicted in Fig. (5)

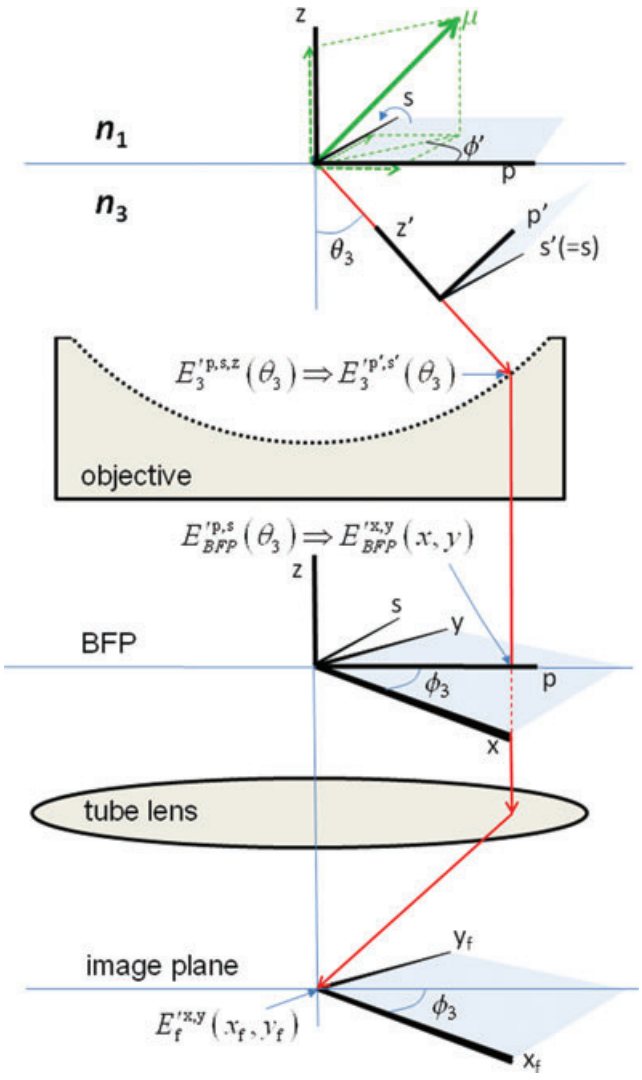

Fig. 5. Schematic diagram of the coordinate systems and variables used for transforming the expression for electric field emitted from the dipole $\boldsymbol{\mu}$ to the expressions for electric field at the back focal plane (BFP) and the image plane. The plane of the diagram is the 'meridional plane' (also referred to as the 'plane of observation' in Hellen and Axelrod, 1987) in the pre-objective space) containing the dipole, the z-axis, and the path of a particular ray through the objective and BFP and tube lens to the image point. That plane is 'ray-fixed', but varies in azimuthal angle ϕ_3 relative to the lab-fixed x-z and x_f–z planes. Neither the objective's actual shape nor the path of the ray through it is shown literally. The drawing is intended to show only that the electric field on a spherical surface (with centre at the dipole object) maps onto the planar BFP according to the sine condition for aplanatic systems.

BFP electric field

From the viewpoint of phase optics, the function of the objective lens is to convert the emitted electric field (amplitude and phase) as it appears on a convex spherical surface (with some fixed radius R) in medium 3 centred on the sample into an electric field on the flat surface of the BFP. E'_3 on that

spherical surface (with the prime accent here to distinguish emission from excitation) can be derived by first expressing the dipole emission pattern as an angular spectrum of plane waves in medium 1 (in which the dipole resides). Some of those plane waves, including evanescent ones, interact with the film-coated interface, reflecting or refracting at interfaces 2:1 and 3:2 according to Snell's Law and Fresnel coefficients. The resulting fields in all three media are integrals over the angular spectrum.

To specifically calculate the components of the vector field E'_3 that propagates into medium 3, define a coordinate system (p, s, z) fixed to each ray emanating from the dipole $\boldsymbol{\mu}$ located on the optical axis in the sample plane (see Fig. 5). Unit vector $\hat{\mathbf{z}}$ lies along the optical axis of the objective, which is normal to the interface and positive directed into medium 1. The direction of a particular ray in medium 3 heading toward the objective, combined with $\hat{\mathbf{z}}$ forms a 'plane of propagation' for that ray. Unit vector $\hat{\mathbf{p}}$ lies in that plane and parallel to the interface. Unit vector $\hat{\mathbf{s}}$ is perpendicular to the plane of propagation and also parallel to the interface. In this coordinate system, the polar angle is θ_3 measured from the - $\hat{\mathbf{z}}$ direction and the azimuthal angle is ϕ_3 measured from the + $\hat{\mathbf{p}}$ direction. The orientation of the dipole vector $\boldsymbol{\mu}$ is (θ', ϕ'), where θ' is the polar angle with respect to the $+\hat{\mathbf{z}}$ direction and ϕ' is the azimuthal angle measured from an arbitrary x-axis (to be specified later). The results for the (p, s, z) components of the electric field E'_3 are given by Eqs (26–33) in Hellen & Axelrod (1987). (Note some notational modifications that deviate from Hellen & Axelrod, 1987: the polar angle θ_3 on the pre-objective spherical surface here is measured from the negative z-axis whereas the corresponding polar angle θ in Hellen & Axelrod (1987) is the supplement $(\pi\text{-}\theta_3)$ measured from the positive z-axis. Terms have been algebraically rearranged. The film thickness h here is called t in Hellen & Axelrod (1987). Also in Hellen & Axelrod (1987) there is a typographical sign error in Eq. 26b in which the '+' should be a '−', thereby affecting the phase but not the amplitude nor the resulting intensities.)

$$\begin{aligned} E_3'^{\mathrm{p}} &= Ct^{\mathrm{p}} \cos\theta_3 \left(\frac{\mu_{\mathrm{p}}}{n_3} + \frac{\mu_{\mathrm{z}} \sin\theta_3}{\cos\theta_1} \right) \\ E_3'^{\mathrm{s}} &= Ct^{\mathrm{s}} \left(\frac{\mu_{\mathrm{s}}}{n_3 \cos\theta_1} \right) \\ E_3'^{\mathrm{z}} &= Ct^{\mathrm{p}} \sin\theta_3 \left(\frac{\mu_{\mathrm{p}}}{n_3} + \frac{\mu_{\mathrm{z}} \sin\theta_3}{\cos\theta_1} \right) \end{aligned} \tag{10}$$

where

$$C \equiv \frac{k_3^2 e^{ik_3 R} \cos\theta_3}{R n_1} e^{-ik_3 h \cos\theta_3} e^{ik_1 \cos\theta_1 z}. \tag{11}$$

Factors $t^{\mathrm{p,s}}$ are Fresnel transmission coefficients for our three layer system, given by

$$t^{\mathrm{p,s}} = \frac{t_{12}^{\mathrm{p,s}} t_{23}^{\mathrm{p,s}} e^{ik_2 h \cos\theta_2}}{1 + r_{12}^{\mathrm{p,s}} r_{23}^{\mathrm{p,s}} e^{2ik_2 h \cos\theta_2}} \tag{12}$$

with $\cos\theta_{1,2}$ defined as in Eq. (4). The components μ_p, μ_s, μ_z of the dipole along the p-, s-, and z-axes are

$$\begin{aligned} \mu_{\mathrm{p}} &= \mu \sin\theta' \cos(\phi' - \phi_3) \\ \mu_{\mathrm{s}} &= \mu \sin\theta' \sin(\phi' - \phi_3) \\ \mu_{\mathrm{z}} &= \mu \cos\theta'. \end{aligned} \tag{13}$$

Parameters $k_{1,2,3}$ are defined as

$$k_{1,2,3} \equiv \left(\frac{2\pi}{\lambda} \right) n_{1,2,3} \tag{14}$$

where λ is the emission wavelength in vacuum, which is a redefinition from its earlier use as the excitation wavelength.

The $1/R$ dependence in Eq. (11) is expected from the fact that only 'far field' light propagating from a point source enters the objective. However, as previously discussed, some of the 'near field' emission from the dipole is captured by the interfaces and converted to propagating light; its effect appears in the amplitude and phase of C at polar angles θ_3 large enough for $\cos\theta_1$ to be imaginary.

The expressions for $E_3'^{\mathrm{p,s,z}}$ apply only to a spherical surface in the pre-objective zone and are based on a coordinate system that depends on the azimuthal angle of the plane of propagation. Therefore, this coordinate system is differently rotated for every 'ray' emanating from the dipole. To calculate the field in a lab-fixed frame at the flat BFP in the post-objective zone, several successive coordinate system transformations are necessary (see Fig. 5).

The first transformation takes into account how the direction of linear polarization is affected by passage through the objective. The rule is that the angle made by the electric field vector with the 'meridional plane' (the plane containing the ray and the z-axis) remains the same on the pre- and post-sides of the objective. [The meridional plane is identical to the 'plane of observation' in the pre-objective space, in the terminology of Hellen & Axelrod (1987)]. Since the electric field is transverse to the propagation direction, we rotate the (p,s,z) system around the s-axis by an angle of θ_3 to a coordinate system (p',s',z') such that z' lies along the ray's path in medium 3. Then the angle of the pre-objective ray's polarization with respect to the meridional plane is completely defined by the ratio of $E_3'^{\mathrm{p}'}$ to $E_3'^{\mathrm{s}'}$. The rotation gives

$$\begin{aligned} E_3'^{\mathrm{p}'} &= E_3'^{\mathrm{p}} \cos\theta_3 + E_3'^{\mathrm{z}} \sin\theta_3 \\ E_3'^{\mathrm{s}'} &= E_3'^{\mathrm{s}} \\ E_3'^{\mathrm{z}'} &= 0. \end{aligned} \tag{15}$$

In travelling through the objective, the ray is redirected to propagate parallel to the z-axis, with the transverse directions along the original p- and s-axes. Conservation of the polarization with respect to the meridional plane implies that the electric field components in the post-objective zone $E_{\mathrm{BFP}}'^{\mathrm{p,s}}$ are relatively the same as those in the pre-objective zone $E_3'^{\mathrm{p}',\mathrm{s}'}$ (neglecting losses by reflection at surfaces in the objective).

The relative phases of the two polarization components are left unaltered, but the amplitudes are affected by the apodization factor $(\cos\theta_3)^{-1/2}$ which describes the amount of area on the pre-objective spherical surface that is projected onto a unit area of the BFP.

$$\begin{aligned} E'^{\mathrm{p}}_{\mathrm{BFP}} &= E'^{\mathrm{p}'}_{3}/\sqrt{\cos\theta_3} \\ E'^{\mathrm{s}}_{\mathrm{BFP}} &= E'^{\mathrm{s}'}_{3}/\sqrt{\cos\theta_3} \\ E'^{\mathrm{z}}_{\mathrm{BFP}} &= 0. \end{aligned} \tag{16}$$

These relationships preserve the polarization angles with respect to the meridional plane.

Equation 16 describes the electric field at the single point where the ray crosses the BFP, with the p- and s-directions pointing differently for different rays. The second transformation converts these θ_3-based components into a lab-fixed planar system (x, y) at the BFP, with z the optical axis and the x-axis chosen (arbitrarily) to be in the $\phi_3 = 0$ direction. (If emission polarizers are used, the direction of x and y would be chosen to lie along the two orthogonal directions of those polarizers.) For the electric field at each BFP location (x, y), the following relations give the correct (θ_3,ϕ_3) to substitute into Eqs (10–16), based on the Abbe sine condition for aplanatic systems (Born & Wolf, 1975, Chapter 4):

$$\begin{aligned} \theta_3 &= \sin^{-1}\left(\frac{\sqrt{x^2+y^2}}{r}\right) \\ \phi_3 &= \cos^{-1}\left(\frac{x}{r}\right) = \sin^{-1}\left(\frac{y}{r}\right). \end{aligned} \tag{17}$$

The final transformation, a rotation of $(-\phi_3)$ around the z-axis, expresses the electric field components at the BFP along the lab-fixed x-, y-, z-axes rather than the ray-fixed p-, s-, z-axes

$$\begin{aligned} E'^{\mathrm{x}}_{\mathrm{BFP}}(x, y) &= \left(\frac{x}{r}\right) E'^{\mathrm{p}}_{\mathrm{BFP}}(x, y) - \left(\frac{y}{r}\right) E'^{\mathrm{s}}_{\mathrm{BFP}}(x, y) \\ E'^{\mathrm{y}}_{\mathrm{BFP}}(x, y) &= \left(\frac{y}{r}\right) E'^{\mathrm{p}}_{\mathrm{BFP}}(x, y) + \left(\frac{x}{r}\right) E'^{\mathrm{s}}_{\mathrm{BFP}}(x, y). \end{aligned} \tag{18}$$

The derivation leading to Eq. (18) ignores diffraction of the propagating light in the 'infinity space' between the objective and the BFP. However, this is not likely to be significant because the spatial variations of electric field in that zone are much broader than the wavelength of light. The coordinates (x,y) may extend to infinity in a formal sense, but $E'^{\mathrm{x,y}}_{\mathrm{BFP}}(x, y)$ is set equal to zero outside the radius of the objective's aperture.

BFP intensity pattern

The intensity at the BFP is related to the electric field at the BFP according to $I'^{\mathrm{x,y}}_{\mathrm{BFP}} = |E'^{\mathrm{x,y}}_{\mathrm{BFP}}|^2$. There is a qualification: the derivation of Eq. (18) implicitly assumed that the fluorophore is a constant amplitude oscillating dipole. But for dipoles near a surface, some of the near field energy that otherwise would not propagate away is captured by the surface and converted into propagating energy. The result is that the total power radiated (in all directions) from a constant amplitude dipole [denoted as $P_{\mathrm{T}}(z,h)$ and displayed in Fig. 1 of Hellen & Axelrod (1987)] increases slightly as the dipole approaches the surface. This slight z-dependence also depends on the film thickness and dipole orientation. However, the total emission power (for a 100% quantum efficiency fluorophore) is a fixed fraction of the total power of light absorption, which in turn is proportional to the local intensity of the excitation light. [See Hellen & Axelrod (1987) for a discussion of quantum efficiencies < 100%]. Therefore, a fixed power (rather than fixed amplitude) dipole is the appropriate model, and the intensities $I'^{\mathrm{x,y}}_{BFP}$. derived above therefore must be normalized to the total power $P_{\mathrm{T}}(z,h)$. Fortunately, normalization with $P_{\mathrm{T}}(z,h)$ does not matter at all for analyzing the *relative* intensities among locations in the BFP for any fixed film thickness h and fluorophore distance z from the film.

To generate BFP patterns for display, it is still assumed here that $n_3 > n_2 > n_1$.

The maximum possible θ_3 is determined by the numerical aperture NA of the objective:

$$\theta_3^{\mathrm{MAX}} = \sin^{-1}(\mathrm{NA}/n_3) \tag{19}$$

and critical angles θ^c_{31} and θ^c_{32} are defined as previously (Eqs 1 and 2). *Propagating* rays from the dipole that almost skim along the 2:1 interface in medium 1 will refract toward the normal into medium 2 and then more so into medium 3 at the polar angle θ^c_{31}. Therefore, any light seen in medium 3 at $\theta_3 > \theta^c_{31}$ cannot possibly have originated from propagating light emitted by the dipole. Light captured by a high-aperture objective in those large polar angles arises solely from dipole's near field. Light in this zone is called 'supercritical' because it travels at polar angles in medium 3 which, if the light is going in the opposite direction, would result in TIR in medium 1. In the entire supercritical zone of $\theta_3^{\mathrm{MAX}} > \theta_3 > \theta^c_{31}$ (observable if $\mathrm{NA} > n_1$), $\cos\theta_1$ is imaginary.

The supercritical zone is discussed separately below for the $h = 0$ and $h > 0$ cases.

(a) *No-film present ($h = 0$).* As pointed out by Lieb *et al.* (2004) and Burghardt & Ajtai (2009), the BFP pattern can be used to provide a sensitive quantitative reading of both the polar and azimuthal angle of the fluorophore orientation. Fig. (6) shows a computer-calculated view of the intensity pattern (with no polarizers in the path) at the BFP for the no-film case ($h = 0$), at three different fluorophore distances from the interface ($z = 0$, $\lambda/10$, λ) and three different fluorophore orientations ($\theta' = 0$, $\pi/4$, $\pi/2$, all at $\phi' = 0$). Note that the patterns are quite distinct for different polar orientation angles, and that azimuthal angles are unambiguous, i.e. that orientation (θ', ϕ') should be quite distinguishable from $(\theta', \pi+\phi')$ if θ' is intermediate between 0 and $\pi/2$.

An intensity peak always occurs at a radius corresponding to $\theta_3 = \theta^c_{31}$. Note that light from the near field, seen in the $\theta_3 > \theta^c_{31}$ annulus, is much greater for the $z = 0$ case than the $z = \lambda$ case. This distinction suggests a direct experimental method to determine the absolute z-distance of a fluorophore

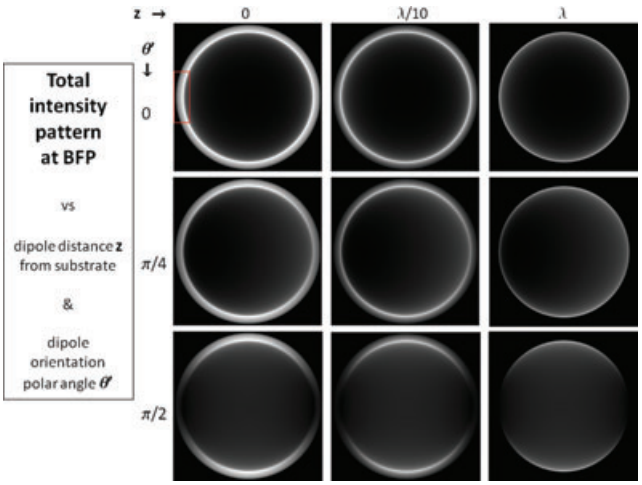

Fig. 6. Computer-calculated view of the intensity pattern at the BFP shown in greyscale for the no-film case ($h = 0$), at three different fluorophore distances from the interface: $z = 0$, 0.1λ and λ, and three different fluorophore orientations (θ', ϕ') as measured in the lab-fixed x-, y-, z-coordinate system: (0,0), ($\pi/4$,0) and ($\pi/2$,0). In each panel, lab axis +z points orthogonally into the plane of the figure, +x points to the right in the plane, and +y points up in the plane. The red rectangle indicates the region to be magnified under various conditions in Fig. (8). All patterns here and subsequently were generated by a custom program written in IDL. Parameters assumed in these calculations are: NA = 1.49, $n_1 = 1.33$; $n_3 = 1.52$ (slightly changed from 1.515 used to generate Figs. 2–4).

from the substrate: take a ratio γ of the integrated supercritical to the integrated subcritical intensities. By first reading out the correct orientation from the asymmetry of the pattern, an appropriate γ versus z can be generated. Fig. 7(a) shows γ versus z for three different fluorophore orientations. As expected, the curves monotonically decrease so given a particular γ, the corresponding absolute z can be read. The key advantage to using γ to estimate z is that it depends upon neither calibration of observed intensity nor normalization by total emitted power $P_T(z,h)$: γ is inherently normalized against changes in illumination power and emission and collection efficiencies.

For small z, most of the light gathered by a high-aperture objective comes through the supercritical zone. The 'extra' aperture, say in going from NA = 1.33 (for which there is no supercritical zone) to 1.49 more than doubles the light gathering ability, which is a strong reason for using the highest apertures available.

(b) *Film present ($h > 0$).* A film leads to be shifted and possibly multiple peaks in the intensity pattern in the BFP supercritical zone, as seen in Fig. 8. In the lower θ_3 portion of the supercritical zone defined as $\theta_3^{MAX} > \theta_{32}^c > \theta_3 > \theta_{31}^c$, energy captured from the dipole's near field by the 2:1 interface becomes propagating light in medium 2. In this lower sub-zone, cos θ_2 is real. Interference/resonance effects can occur as this propagating light multiply reflects at the $z = 0$ and $z = -h$ boundaries of medium 2 (in analogue with the interference/resonance discussed earlier for excitation light in the same angular range). In the upper θ_3 portion of the supercritical zone defined as $\theta_3^{MAX} > \theta_3 > \theta_{32}^c > \theta_{31}^c$, energy medium 2 is evanescent (decaying into the film from the 2:1 interface) rather than propagating and no interference/resonance effects occur. In this sub-zone, cos θ_2 is imaginary. For certain $h > 0$, the BFP intensity at the critical angle θ_{31}^c can actually increase by more than a factor of two over its value at $h = 0$.

For each h, the shape of the BFP intensity versus radius curve is qualitatively unique. Therefore, it should be possible to identify h by matching an experimentally observed shape to a theoretically predicted curve in a set such as those in Fig. 8. Once h is determined, then γ (the supercritical:subcritical integrated intensity ratio) can be used to infer the distance z, as it was for the $h = 0$ case. Fig. 7(b) shows γ versus z for various h. All the curves monotonically decrease with z as expected, with the rate of decrease and the maximum amplitude of γ (at $z = 0$) bearing a complicated dependence on h. To avoid ambiguities in determining z from an experimental γ, h must first be inferred from BFP intensity pattern versus radius.

For large enough h, the border between the lower sub-zone ($\theta_{32}^c > \theta_3 > \theta_{31}^c$ with its multiple peaks) and the upper sub-zone ($\theta_3 > \theta_{32}^c > \theta_{31}^c$ with no peaks) becomes quite distinct: that border occurs at $\theta_3 = \theta_{32}^c$ (as indicated in Fig. 8). From this potentially observable value of θ_{32}^c, the index of refraction of the film can be deduced via Eq. (2). For such thick films, the extranumerical aperture of the objective corresponding to angles beyond $\theta_3 = \theta_{32}^c$ is 'wasted' because almost no light is gathered there. The near-field energy that might have gone into those high angles decays in the −z-direction from the 2:1 interface and cannot 'reach across' a thick film to become propagating in the higher refractive index of medium 3.

The absence of light in the upper sub-zone for thick films corresponds to a loss of total power gathered by the objective. But the loss is only slight, and as h becomes large, the total power gathered by the objective approaches a constant value (Fig. 7(c)). Note that calculations of total power as a function of either h or z must include normalization with the total-radiated power function $P_T(z,h)$ [as derived by Ford & Weber (1984) and copied as Eq. (38) in Hellen & Axelrod (1987]. Normalization with $P_T(z,h)$ was used in the generation of Fig. 7(c).

Image plane electric field

Just as the (negligible) diffraction in the infinity space between the objective and the BFP was ignored, assume that the E field at the BFP is preserved diffraction free as it impinges upon the tube lens, which focuses the light to the image plane in an infinity-corrected system. The field $E_f^{x,y}(x_f, y_f)$. at the focal plane of the tube lens is then the Fourier transform (see

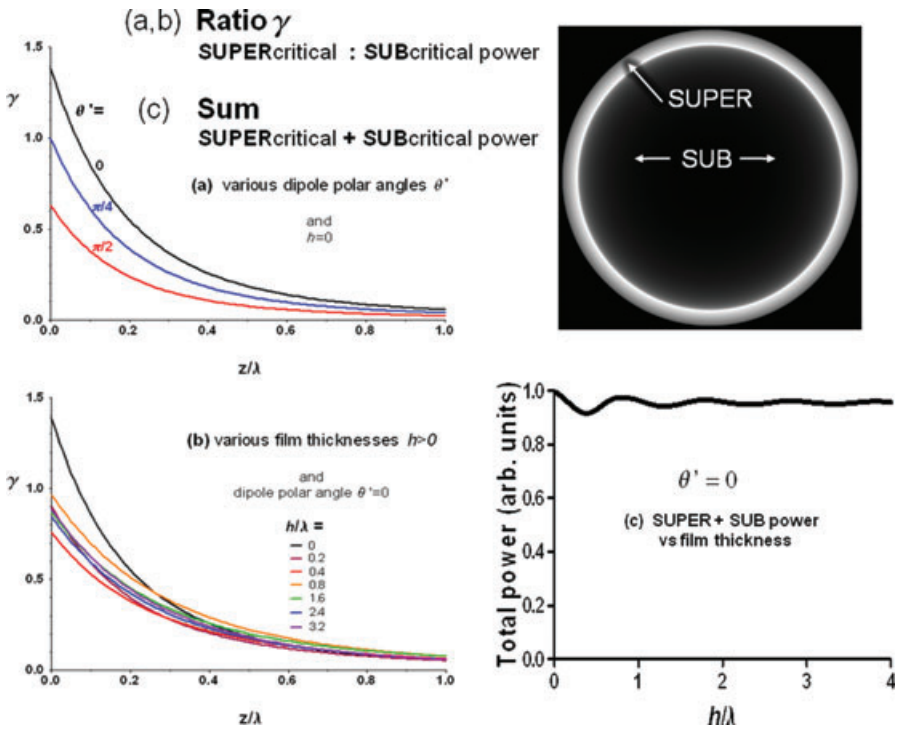

Fig. 7. Power in the back focal plane. (**a**) Ratio γ (supercritical zone power: subcritical zone power in the BFP) versus z for three different fluorophore orientations (θ', ϕ'): (0,0), $(\pi/4,0)$ and $(\pi/2,0)$ with $h = 0$; (**b**) Ratio γ for various film thicknesses $h/\lambda = 0$, 0.2, 0.4, 0.8, 1.6, 2.4, and 3.2 with $(\theta', \phi') = (0,0)$; (**c**) The total power gathered by the objective versus film thickness h/λ. This total power is the sum of the powers in the supercritical and subcritical zones, divided by the total power emitted in all directions $P_T(z,h)$ as discussed in the text. The unpolarized BFP pattern shown for illustration is the $(\theta', \phi') = (0,0)$ and $h/\lambda = 0$ case. Parameters assumed in these calculations are: NA = 1.49; $n_1 = 1.33$; $n_2 = 1.42$; $n_3 = 1.52$, and unpolarized (total) emission.

Goodman, 1968, Eqs 5–15) of $E'^{x,y}_{BFP}(x, y)$ from Eq. (18):

$$E'^{x,y}_{f}(x_f, y_f) = AB \int_{-\infty}^{\infty} E'^{x,y}_{BFP}(x, y)\, e^{-i(\frac{k}{f})(x_f x + y_f y)} dx\, dy \quad (20)$$

Variable f is the focal length of the tube lens, wavenumber k equals $2\pi/\lambda$ where λ is the emission wavelength in air and (x_f, y_f) is the position in the focal plane. Factor A is a constant amplitude, and factor B describes a spherical phase curvature in the focal plane, but neither factor will have any impact on the form of the intensity $I'^{x,y}_f = |E'^{x,y}_f|^2$ in the focal plane.

Image plane intensity: the point spread function

(a) *No-film present ($h = 0$).* Fig. 9(a–c) shows the unpolarized PSF intensity pattern $I'^{x}_f + I'^{y}_f$ for the $h = 0$, $z = 0$ case at three different fluorophore orientations corresponding to the BFP intensity patterns of Fig. 6. Note that the unpolarized PSF has a 'hole' in the middle for fluorophore orientation polar angles near $\theta' = 0$ (see Fig. 9(a)). At $\theta' = \pi/2$, the unpolarized PSF is hole-free but is slightly elongated in the direction of the dipole (see Fig. 9(c)).

In the intermediate polar angle range (e.g. $\theta' = \pi/4$), the unpolarized PSF pattern is sufficiently asymmetrical to retain information about both the polar and azimuthal dipole orientation angles (see Fig. 9(b)). However, this asymmetry is entirely due to the supercritical (i.e. near field) annulus of emission light at the BFP. If that supercritical light is blocked (in computer simulation), then the unpolarized PSF pattern for $\theta' = \pi/4$ becomes completely symmetric (see Fig. 9(k)). An important consequence of this asymmetry is that fitting this shape to a symmetrical function (such as a Gaussian) to find the centroid, as is commonly done in image reconstruction techniques for superresolution such as PALM and STORM (Rust *et al.*, 2006), will produce a positional error, on the order of 20 nm for visible fluorescence. What is really an orientational feature will be misinterpreted as a slight positional shift.

Fig. 9(d–i) also shows how the PSF's would appear as seen through a linear polarizer oriented along either the x- or y-axis. Clearly, the polarizer enhances the sensitivity to dipole orientation. In an alternative approach to determining orientation, some single molecule studies have employed a slightly defocused PSF because it inherently 'mixes in' the unambiguous azimuthal information clearly present at the BFP (Cyphersmith *et al.*, 2011; Patra *et al.*, 2004; Hohlbein & Hübner, 2008; Aguet *et al.*, 2009; Burghardt, 2011; Pavani *et al.*, 2009).

The PSF's discussed above are based on the 'complete' theory as presented above, which takes the following features

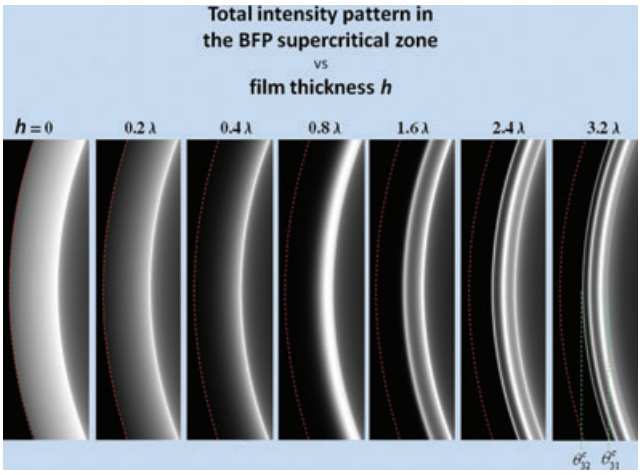

Fig. 8. Detail of BFP pattern in supercritical zone for various film thicknesses $h/\lambda = 0$, 0.2, 0.4, 0.8, 1.6, 2.4, and 3.2, all with $(\theta', \phi') = (0,0)$ and $z = 0$. The regions are magnified views corresponding to the red rectangle region marked in the upper left panel of Fig. (6). The dashed red line in each panel marks the location limiting aperture of the objective as it is projected in the BFP, corresponding to $\theta_3 = \theta_3^{MAX}$ The two green dashed lines in the rightmost panel show the radial positions in the BFP corresponding to $\theta_3 = \theta_{31}^c$ and θ_{32}^c. Parameters assumed in these calculations are: NA = 1.49, $n_1 = 1.33$; $n_2 = 1.42$; $n_3 = 1.52$.

into account: (a) refraction and reflection at the surface; (b) apodization (Eqs. 16); (c) the highly anisotropic oriented dipole emission pattern including phase information and (d)) the near-field interaction with the surface. It is interesting to compare this 'complete' model covered here against the simplest (and most commonly assumed) model: that the fluorophore emits in a completely isotropic pattern, the light from which completely and uniformly fills the entire objective aperture. This fictitious assumption leads to the classic 'Airy disk' PSF, shown for comparison in Fig. 9(j). The Airy disk model produces the smallest radius PSF; the additional features of the 'complete' model do not produce a smaller PSF.

Since the supercritical zone of the BFP 'sees' only fluorophores close to the surface (see Fig. 7), surface selectivity can be achieved by imaging only that zone and blocking all of the subcritical zone; this was confirmed experimentally by Axelrod (2001). A computer calculation of the PSF with subcritical light completely blocked is shown in Fig. 9(l). This blockage introduces enhanced rings in the PSF and accounts for why the wide field images seen in Axelrod (2001) with subcritical blockage lose some sharpness.

(b) *Film present ($h > 0$).*Fig. 10 shows the PSF profiles for three of the cases corresponding to the BFP pattern shown in Fig. 8 ($\theta' = 0$ and $h = 0$, 1.6λ, 3.2λ). The presence of a film creates a broad background base to the PSF, most noticeable at higher film thicknesses. Since the BFP intensity develops sharp maxima containing high spatial frequencies in this h range, it is not surprising that it Fourier transforms into a broad PSF feature.

BFP phase effects

Since the smallest PSF occurs for the Airy disk model, one might conjecture that modification of the amplitude and phase at the BFP by some specially designed phase plate might narrow the complete model PSF prediction and thereby suggest a practical method to increase sharpness of the image (and perhaps better resolution of two nearby dipoles). In the Airy disk model, the BFP phase is constant everywhere (by definition). But in the complete model, the phase in the supercritical zone of the BFP varies with both azimuthal angle and radius, and differs from the constant phase of the subcritical zone. Furthermore, the phases of the x-polarized and y-polarized components of the electric field in the supercritical BFP are different from each other, meaning that the polarization there is generally elliptical, with an ellipticity that depends on position in the supercritical BFP zone. However, a practical modification of the supercritical BFP phase would have to be uniform to work equally with any fluorophore orientation. Such a uniform retardation or advance of the phase in the supercritical BFP annulus produces only a minor change (typically less than 10%) in the PSF width.

Summary and possible applications

The presence of a film on a substrate in bioresearch is common, both intentionally (e.g. a collagen or agarose layer deposited before cell plating, or a supported lipid monolayer, bilayer, or multilayer, or an acrylamide gel to immobilize beads or single molecules) and non-intentionally (e.g. a cell wall interposed between the substrate and cellular organelle, the plasma membrane interposed under a cytoplasmic organelle, or a layer of proteins deposited by cultured cells). This theoretical study shows that the presence of an intermediate refractive index dielectric film between glass and water greatly affects both the excitation and observed emission of a single molecule adsorbed at the film's aqueous interface. In both cases, resonance-like peaks are produced by interference among multiple reflections in the film. For TIR excitation, these peaks are manifested in the evanescent field intensity versus incidence angle curve. For observation of emission through a high aperture microscope objective, the peaks are manifested in the supercritical zone of the BFP.

The theory here presents general expressions for the emission fields at both the BFP and the image plane in which fluorophore orientation, distance to the substrate, near field capture, reflection and refraction at the substrate and apodization are taken into account. These expressions should make it possible for experimentalists to write custom programs to predict effects with different parameters (NA, refractive

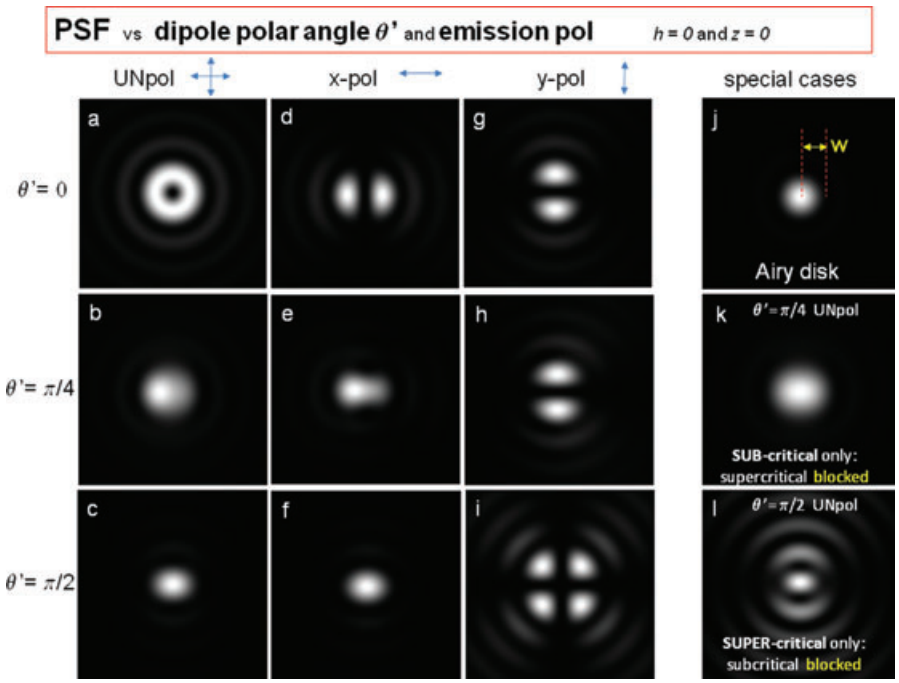

Fig. 9. PSF intensity pattern at the image plane for the same cases as the BFP patterns in the left column of Fig. (6): fluorophore orientations $(\theta', \phi') = (0,0)$, $(\pi/4,0)$ and $(\pi/2,0)$, all with $h = 0$ and $z = 0$. (**a–c**) NO polarizer; (**d–f**) linear polarizer oriented to transmit x-polarization only (left-right on the plane of this diagram); (**g–i**) linear polarizer oriented to transmit y-polarization only (up-down in the plane of this diagram). Two special cases are also shown. (**j**) An Airy disk generated by completely filling the aperture with unpolarized light. Characteristic radius $w = 0.61\ \lambda$ /NA (= 0.409 λ for NA = 1.49) sets the scale for all the panels. (**k**) Supercritical BFP light completely blocked. Note that the obvious x-direction asymmetry in panel d (and e) identifies the azimuthal angle as $\phi' = 0$. If the azimuthal angle were $\phi' = \pi$, the x-direction asymmetry would be reversed. However, if the supercritical light in the BFP is completely blocked, then the asymmetry disappears, making a reading of the azimuthal angle ambiguous (compare 9b with 9k). (**l**) Subcritical BFP light completely blocked. This introduces enhanced rings in the PSF (compare 9c with 9l). The greyscale intensities are adjusted within each panel for best viewing, but intensities cannot be compared among different panels. Parameters assumed in all these calculations are: NA = 1.49, $n_1 = 1.33$; $n_3 = 1.52$.

indices, etc) than those chosen here. The special case of no film is an important subset of the results.

Several specific aspects of the theory can be used to infer properties of the film and/or the fluorophore dipole in an experiment. These possible applications (all discussed in detail earlier) are summarized here.

(a) A suitable measure of the excitation intensity versus incidence angle of excitation light can report the presence, the refractive index and the thickness of an intermediate film, and possibly become the basis for a new form of image contrast. Fluorescence intensity can be considered as a proxy for excitation light intensity, but with some caution. If a single molecule fluorophore or a sheet of fluorophores is located at any particular unchanging single distance z from the film surface (and unchanging orientation), the fluorescence gathered by any objective will be exactly proportional to the excitation light intensity at that z position. But if a thicker layer of fluorophores (or a fluorophore solution in the water) inhabiting a range of z positions and z-dependent orientations is used, the proportionality is only approximate, because light power gathered by an objective from an excited fluorophore depends slightly upon its z position and on film thickness h, and strongly upon its orientation.

(b) A film can increase the amplitude of the evanescent field in TIR excitation for particular combinations of film refractive indices and thicknesses.

(c) The pattern of fluorescence emission in the supercritical annulus at the BFP can also report refractive index and thickness of an intermediate film.

(d) For the purpose of measuring film thickness, viewing the effect of excitation incidence angle (Fig. 2) or viewing the emission pattern in the BFP plane (Fig. 8) are probably equally sensitive methods, because they are based on the same phenomenon of multiple reflections. But the excitation incidence angle method will likely produce more photons because the sample can consist of large numbers of fluorophores spread over the field of view, whereas the BFP method assumes that the fluorescent source (such as a single molecule) is concentrated on the optical axis. By contrast, the BFP approach requires

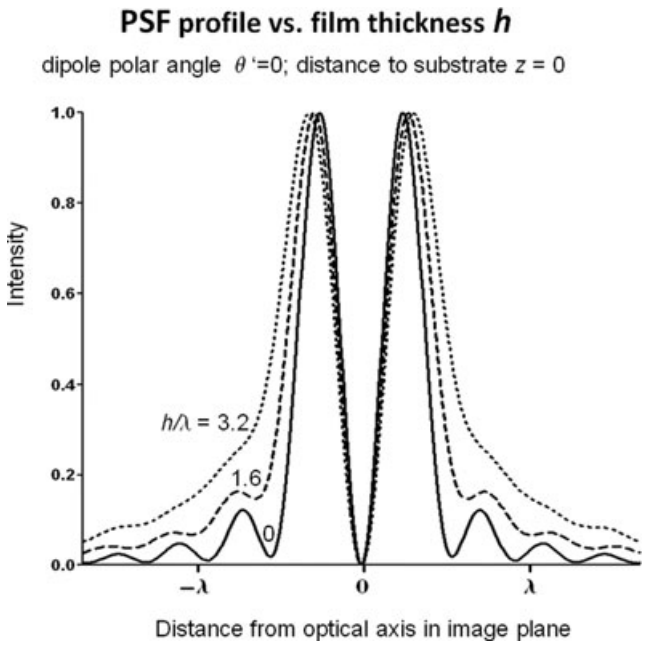

Fig. 10. Effect of film thickness on the PSF, shown in intensity profile across a diameter for the $(\theta', \phi') = (0,0)$, z = 0 case, for three different h/λ (= 0, 1.6, and 3.2). The maximum amplitude of each profile is normalized to 1.0 here to show most vividly the wider spread with greater h. The actual total power of each PSF (equal to its total power at the BFP) varies only slightly with h, as shown in Fig. 7c. Parameters assumed in these calculations are: NA = 1.49, $n_1 = 1.33$; $n_2 = 1.42$; $n_3 = 1.52$.

no special experimental manipulations (like varying incidence angle): the information is already available at the BFP. Other thickness-sensitive effects such as those depicted in Fig 7(b) (the rather subtle effect of h on the supercritical vs. subcritical power in the BFP) or Fig 10 (the broadening of the PSF, which is hard to measure accurately because the PSF typically spreads over very few camera pixels) are not good choices for measuring film thickness.

(e) The theory presented here predicts that the presence of a film may complicate attempts to measure z-dependent fluorophore concentration by varying TIR incidence angle.

(f) The ratio of supercritical to subcritical power in the back focal plane is a sensitive measure of distance of a fluorophore from the surface.

(g) Unambiguous reading of the orientation (both polar and azimuthal) of a fluorophore can be obtained easily from the BFP pattern (as previously shown by Lieb *et al.*, 2004 and by Burghardt and Ajtai, 2009), and also obtained from the image plane PSF pattern produced by an objective with NA sufficiently high to gather near-field light.

(h) The full theoretical treatment produces PSF patterns that are very different from, and generally larger than, the classical Airy disk pattern.

(i) When viewed through a polarizer, the PSF images of a single oriented dipole predicted by the full theoretical treatment are rather complicated. But these patterns are needed to accurately predict the emission polarization from a sample consisting of many fluorophores with a distribution of orientations, such as a labelled biological membrane containing submicroscopic indentations (e.g. exo- and endocytotic sites). Any precise analysis of BFP or PSF single molecule emission intensity patterns emitted is somewhat limited by the number of emitted photons that are captured and detected through a microscope. That number is capped by photobleaching and possibly by a short time window imposed by fast dynamics under study.

(j) In techniques such as PALM and STORM, the PSF of a single molecule is fit to a symmetrical Gaussian in order to locate the centre to superresolution accuracy. However, the actual asymmetry of the PSF from an obliquely oriented single molecule, which is due entirely to the near-field contribution as it appears in the supercritical zone of the BFP, will cause a purely orientational feature to be misinterpreted as a positional shift on the order of 1/25 of the wavelength.

(k) The theory for excitation and emission at a dielectric-coated surface as presented here can be generalized directly to metal film-coated surfaces, simply by using the appropriate metal refractive index for n_2. A metal typically has a complex dielectric constant ε_2 with a negative real part and a positive imaginary part (at the frequency of the light), and n_2 is just the square root of ε_2 in the complex plane. Surface plasmon effects, beyond the scope of discussion in this paper, will then emerge from the equations presented herein.

Acknowledgements

The author thanks Drs. Geneva M. Omann and Ronald V. Holz for reviewing a draft of this manuscript and making suggestions for changes, and Drs. Edward H. Hellen and Nancy L. Thompson for useful discussions. This project was supported by NIH grants 1R21NS073686-01 and 2R56NS038129-11 to Dr. R.V.H and D.A.

References

Acher, O., Adenot, A.L. & Duverger, F. (2000) Fresnel coefficients at an interface with a lamellar composite material. *Phys. Rev. B* **62**, 13748–13756.

Aguet, F., Geissbuhler, S., Marki, I., Lasser, T. & Unser, M. (2009) Super-resolution orientation estimation and localization of fluorescent dipoles using 3-D steerable filters. *Opt. Express* **17**, 6829–6848.

Axelrod, D. (2001) Selective imaging of surface fluorescence with very high aperture microscope objectives. *J. Biomed. Opt.* **6**, 6–13.

Benešová, M. & Tománek, P. (1999) Thickness measurement of thin dielectric films by evanescent total reflection fluorescence. *J. Microsc.* **194**, 434–438.

Bohmer, M. & Enderlein, J. (2003) Orientation imaging of single molecules by wide-field epifluorescence microscopy. *J. Opt. Soc. Am. B* **20**, 554–559.

Born, M. &Wolf, E. (1975) *Principle of Optics*, 5th edn. Pergamon Press, Oxford.

Burghardt, T.P. (2011) Single molecule fluorescence image patterns linked to dipole orientation and axial position: application to myosin cross-bridges in muscle fibers. *Plos One* **6**, e16772/1–13.

Burghardt, T.P. & Ajtai, K. (2009) Mapping microscope object polarized emission to the back focal plane pattern. *J. Biomed. Opt.* **14**, 034036/1–8.

Burghardt, T.P. & Thompson, N.L. (1984) Effect of planar dielectric interfaces on fluorescence emission and detection: evanescent excitation with high-aperture collection. *Biophys. J.* **46**, 729–737.

Challener, W.A., Edwards, J.D., McGowan, R.W., Skorjanec, J. & Yang Z. (2000) A multilayer grating-based evanescent wave sensing technique. *Sensors and Actuators B* **71**, 42–46.

Chance, R.R., Prock, A. & Silbey, R. (1978) Molecular fluorescence and energy transfer near interfaces. *Adv. Chem Phys.* **37**, 1–65.

Chiu, K.P., Kao, T.S. & Tsai, D.P. (2008) Evanescent field enhancement due to plasmonic resonances of a metamaterial slab. *J. Microsc.* **229** pt 2, 313–319.

Courtois, J.-Y., Courty, J.-M. & Mertz, J.C. (1996) Internal dynamics of multilevel atoms near a vacuum-dielectric interface. *Phys. Rev. A* **53**, 1862–1878.

Cyphersmith, A., Maksov, A., Hassey-Paradise, R., McCarthy, K.D. & Barnes, M.D. (2011) Defocused emission patterns from chiral fluorophores: application to chiral axis orientation determination. *J. Phys. Chem. Lett.* **2**, 661–665.

Dickson, R.M., Norris, D.J. & Moerner, W.E. (1998) Simultaneous imaging of individual molecules aligned both parallel and perpendicular to the optic axis. *Phys. Rev. Lett.* **81**, 5322–5325.

Ekgasit, S., Thammacharoen, C. & Knoll, W. (2004) Surface plasmon resonance spectroscopy based on evanescent field treatment. *Anal. Chem.* **76**, 561–568.

Ekgasit, S., Yub, F. & Knoll, W. (2005a) Fluorescence intensity in surface-plasmon field-enhanced fluorescence spectroscopy. *Sensors and Actuators B* **104**, 294–301.

Ekgasit, S., Thammacharoen, C., Yu, F. & Knoll W. (2005b) Influence of the metal film thickness on the sensitivity of surface plasmon resonance biosensors. *Appl. Spect.* **59**, 661–667.

Enderlein, J., Ruckstuhl, T. & Seeger, S. (1999) Highly efficient optical detection of surface-generated fluorescence. *Appl. Opt.* **38**, 724–732.

Fattinger, C. & Lukosz, W. (1984) Optical-environment-dependent lifetimes and radiation patterns of luminescent centers in very thin films. *J. Luminesc* 31–32, 933–935.

Feng, S., Elson, J.M., & Overfelt, P.L. (2005) Optical properties of multilayer metal-dielectric nanofilms with all-evanescent modes. *Opt. Express* **13**, 4113–4124.

Ford, G.W. & Wber, W.H. (1984) Electromagnetic interactions of molecules with metal surfaces. *Phys. Rep* **113**, 195–287.

Foreman, M.R., Romero, C.M. & Török, P. (2008) Determination of the three-dimensional orientation of single molecules. *Opt. Lett.* **33**, 1020–1022.

Fourkas, J.T. (2001) Rapid determination of the three-dimensional orientation of single molecules. *Opt. Lett.* **26**, 211–213.

Goodman, J.W. (1968) *Introduction to Fourier Optics*. McGraw-Hill, New York.

Gu, M. (2000) *Advanced Optical Imaging Theory*. Springer Verlag, Berlin.

Hellen, E.H. & Axelrod, D. (1987) Fluorescence emission at dielectric and metal-film interfaces. *J. Opt. Soc. Am. B* **4**, 337–350.

Hohlbein, J. & Hübner, C.G. (2008) Three-dimensional orientation determination of the emission dipoles of single molecules: the shot-noise limit. *J. Chem. Phys.* **129**, 094703/1–8.

Ishitobi, H., Nakamura, I., Hayazawa, N., Sekkat, Z. & Kawata, S. (2010) Orientational imaging of single molecules by using azimuthal and radial polarizations. *J. Phys. Chem. B* **114**, 2565–2571.

Kaiser, R., Lévy, Y., Vansteenkiste, N., Aspect, A., Seifert, W., Leipold, D. & Mlynek, J. (1994) Resonant enhancement of evanescent waves with a thin dielectric waveguide. *Opt. Comm.* **104**, 234–240.

Ke, P.C., Gan, X.S., Szajman J., Schilders, S. & Guy, M. (1997) Optimizing the strength of an evanescent wave generated from a prism coated with a double-layer thin-film stack. *Bioimaging* **5**, 1–8.

Le Moala, E., Forta, E., Lévêque-For, S., Cordelièresc, F.P., Fontaine-Aupart, M.-P. & Ricolleaua, C. (2007) Enhanced fluorescence cell imaging with metal-coated slides. *Biophys. J.* **92**, 2150–2161.

Lieb, M.A., Zavislan, J.M. & Novotny, L. (2004) Single-molecule orientations determined by direct emission pattern imaging. *J. Opt. Soc. Am. B* **21**, 1210–1215.

Mattheyses, A.L. & Axelrod, D. (2005) Fluorescence emission patterns near glass and metal-coated surfaces investigated with back focal plane imaging. *J. Biomed. Opt.* **10**, 054007/1–6.

Mertz, J. (2000) Radiative absorption, fluorescence, and scattering of a classical dipole near a lossless interface: a unified description. *J. Opt. Soc. Am. B* **17**, 1906–1913.

Nesnidal R.C. & Walker, T.G. (1996) Multilayer dielectric structure for enhancement of evanescent waves. *Appl. Opt.* **35**, 2226–2229.

Oheim, M., Loerke, D., Preitz, B. & Stuhmer, W. (1998) A simple optical configuration for depth-resolved imaging using variable angle evanescent-wave microscopy. *SPIE* **3568**, 131–140.

Ölveczky, B.P. Periasamy, N. & Verkman, A.S. (1997) Mapping fluorophore distributions in three dimensions by quantitative multiple angle-total internal reflection fluorescence microscopy. *Biophys. J.* **73**, 2836–2847.

Patra, D., Gregor, I. & Enderlein, J. (2004) Image analysis of defocused single-molecule images for three-dimensional molecule orientation studies. *J. Phys. Chem. A* **108**, 6836–6841.

Pavani, S.R.P., Thompson, M.A., Biteen, J.S., Lord, S.J., Liu, N., Twieg, R.J., Piestun, R. & Moerner, W.E. (2009) Three-dimensional, single-molecule fluorescence imaging beyond the diffraction limit by using a double-helix point spread function. *Proc. Natl. Acad. Sci. USA* **106**, 2995–2999.

Polerecky', L., Hamrle, J. & MacCraith, B.D. (2000) Theory of the radiation of dipoles placed within a multilayer system. *Appl. Opt.* **39**, 3968–3977.

Ruckstuhl, T. & Verdes, D. (2004) Supercritical angle fluorescence (SAF) microscopy. *Opt. Express* **12**, 4246–4254.

Rust, M.J., Bates, M. & Zhuang, X. (2006) Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat. Methods* **3**, 793–795.

Sikorski, Z. & Davis, L.M. (2008) Engineering the collected field for single molecule orientation determination. *Opt. Express* **16**, 3660–3673.

Wood, B., Pendry, J.B. & Tsai, D.P. (2006) Directed subwavelength imaging using a layered metal-dielectric system. *Phys. Rev. B* **76**, 115116/1–8.