Journal of Microscopy, Vol. 228, Pt 3 2007, pp. 390–405
Received 10 June 2007; accepted 19 July 2007

Evaluating performance in three-dimensional fluorescence microscopy

JOHN M. MURRAY*, PAUL L. APPLETON†,
JASON R. SWEDLOW‡ & JENNIFER C. WATERS**
* *Department of Cell & Developmental Biology, University of Pennsylvania School of Medicine, Philadelphia, PA 19104-6058, U.S.A.*
†*Division of Cell and Developmental Biology and*
‡*Division of Gene Regulation and Expression, College of Life Sciences, University of Dundee, Dundee DD1 5EH, Scotland, U.K.*
** *Department of Cell Biology, Harvard Medical School, Boston MA 02115, U.S.A.*

Key words. Confocal microscopy, deconvolution, light throughput, photobleaching, signal to noise ration.

Summary

In biological fluorescence microscopy, image contrast is often degraded by a high background arising from out of focus regions of the specimen. This background can be greatly reduced or eliminated by several modes of thick specimen microscopy, including techniques such as 3-D deconvolution and confocal. There has been a great deal of interest and some confusion about which of these methods is 'better', in principle or in practice. The motivation for the experiments reported here is to establish some rough guidelines for choosing the most appropriate method of microscopy for a given biological specimen. The approach is to compare the efficiency of photon collection, the image contrast and the signal-to-noise ratio achieved by the different methods at equivalent illumination, using a specimen in which the amount of out of focus background is adjustable over the range encountered with biological samples. We compared spot scanning confocal, spinning disk confocal and wide-field/deconvolution (WFD) microscopes and find that the ratio of out of focus background to in-focus signal can be used to predict which method of microscopy will provide the most useful image. We also find that the precision of measurements of net fluorescence yield is very much lower than expected for all modes of microscopy. Our analysis enabled a clear, quantitative delineation of the appropriate use of different imaging modes relative to the ratio of out-of-focus background to in-focus signal, and defines an upper limit to the useful range of the three most common modes of imaging.

Correspondence to: John M. Murray. Tel: 215 898 3045; fax: 215 898 9871; e-mail: murray@cellbio.med.upenn.edu

Introduction

In biological fluorescence microscopy, image contrast is often degraded by high background arising from out of focus regions of the specimen. This background can be greatly reduced or eliminated by several methods of 'thick specimen microscopy', such as wide-field/deconvolution (WFD), spot scanning confocal (also called point scanning or laser scanning confocal), spinning disk confocal or multiphoton microscopy. There has been a great deal of interest and some confusion about which of these methods is 'better', in principle or in practice. It is for instance common to hear opinions such as: 'less photobleaching occurs when using a disk scanning confocal than other types of microscope and therefore disk scanning confocal is 'best for live-cell imaging'; or, 'the laser illumination used for spot scanning confocals photobleaches much more than the mercury arc illumination used for wide-field (WF) microscopes'. For biological microscopists these are critically important issues and it would be of great benefit to the community to have solid data on which to base these opinions. Perhaps even more valuable is to have a straightforward, generally accepted, test protocol that anyone could use to address these issues in the context of their own local facilities. In this study, we have developed a method to compare the performance of three of the most commonly used techniques: WFD, spot scanning confocal and spinning disk confocal. For our purposes, better performance will be defined as achieving a higher image quality at equivalent 3-D resolution under the constraints imposed by working with typical biological specimens.

Some useful measures of image quality include the overall efficiency of photon acquisition, image contrast and the

precision of quantitative measurements based on image pixel intensities. In fluorescence microscopy, the precision of image intensities, typically characterized as signal-to-noise ratio (SNR), should be limited only by the Poisson noise associated with stochastic absorption, excitation and re-emission of photons. The highest achievable SNR in a single image will therefore never be greater than the square root of the number of photons/pixel in the image. In practice, an upper limit to the SNR is set by the square root of the number of photons/pixel when the detector element reaches full capacity, or saturation. Out of focus background light acts to reduce the effective capacity of the detector; that is, photons from out of focus fluorescence saturate the detector, limiting accumulation of photons from the in-focus plane. Each method of thick specimen microscopy attempts to circumvent this problem in a different way. WFD microscopy aims to use the out of focus light to reconstruct the in-focus object. Spinning disk confocal, spot scanning confocal and multiphoton microscopy spatially restrict the region of excitation and/or detection of emitted photons to reduce the contribution of out of focus background, thus preserving more of the detector capacity for the in-focus light.

Illumination-dependent photobleaching of fluorophores in all specimens and phototoxicity that occurs during live cell imaging restricts the amount of fluorescence excitation light that can be practically used. This limits the available emission photons to a number far below the capacity of a detector element, significantly lowering the achievable image SNR. Each method of thick specimen microscopy has its own characteristic mixture of strengths and weaknesses, ensuring that rank order of performance among different methods will depend on the relative weight assigned to different aspects of performance. Biological specimens differ widely in the level of background compared to in-focus signal, in rate of photobleaching and in sensitivity to photodamage. The relative importance assigned to different measures of microscope performance is therefore strongly sample dependent.

The motivation for the experiments reported here is to establish methodology and guidelines for choosing the appropriate mode of thick specimen microscopy for a given biological specimen. The approach is to compare images acquired by the different methods at equal illumination, employing a specimen whose properties are well defined but adjustable over the range encountered with biological samples, using three different measures of performance. We are primarily concerned with the level of signal and noise in the context of a 3-D sample, but wish to have a test protocol that also rewards higher 3-D resolution. This combination is conveniently brought about using fluorescent beads that are comparable in size to the resolution limit of the high numerical aperture systems used for fluorescence microscopy. When these beads are imaged at appropriately high magnification, decreased resolution has the effect of smearing the available signal photons over an unnecessarily large number of pixels, thus decreasing the number of photons/pixel, and lowering the SNR and contrast. In this situation, measurements of image contrast and SNR, nominally a function of light throughput and added noise, also become sensitive to lowered resolution.

The remaining major difficulty in testing the performance of different microscope systems is in arranging for the comparison to be carried out with equal dosage of illumination. Clearly if there were no control of illumination dose, then, other things being equal, longer exposure and brighter illumination would give higher image SNR, up to the limit imposed by maximum detector capacity. As previously described (Murray, 1998), an essential feature of the comparisons is therefore to calibrate the illumination system on each microscope by measuring the photobleaching of a standardized specimen, whose photobleaching is a measure of the illumination time-irradiance integral. When this time-irradiance integral is adjusted to be the same for two different microscope systems, the quality of the images of small subresolution fluorescent objects becomes a system-independent measure of performance that is sensitive to light throughput, 3-D resolution and intrinsic noise.

A disadvantage of using the rate of photobleaching to calibrate illumination irradiance is that multiphoton microscopy cannot be part of the comparison. The unexpectedly severe photobleaching observed with 2-photon illumination, unpredictably dependent on the third to fifth power of illumination irradiance (Patterson & Piston, 2000), violates the fixed relation between photon emission and photobleaching that is required for our method. This study was therefore performed on three of the most commonly used 3-D imaging techniques for which photobleaching can be used as an accurate measure of illumination irradiance: WFD, spinning disk confocal and spot scanning confocal.

Methods

Background solution

AlexaFluor488 hydroxylamine (Molecular Probes #A30269) was dissolved in dry dimethylformamide to give a 10 mM stock solution which was stored at −20°C. Five μL of this stock was mixed with 0.5 mL of 10 mM glucose in 10 mM bicine pH 8.4 (final 100 μM fluorescent glucose, in ∼10 mM nonfluorescent glucose) and incubated at room temp for 1 h. NaN_3 was added to 10 mM and EDTA to 0.1 mM as preservatives, and the solution was frozen in aliquots at −20°C. Dilutions of this stock were made with 10 mM glucose, 10 mM bicine pH 8.4, 10 mM NaN_3, 0.1 mM EDTA to prepare background solutions with concentrations of AlexaFluor488 glucose lower than 100 μM.

Beads

A total of 0.4 mL red fluorescent (580 ex/605 em) amine modified latex microspheres, 0.21 μm diameter (Molecular

Probes #F8763; 2% solids; $\sim 3.9 \times 10^{12}$/mL; 5.3 μeq amine residues/mL) was washed twice with water, then suspended in 0.8 mL of 0.02% Tween 20, 12.5 mM $NaHCO_3$ pH 8.3. Fluorescein succinimidyl ester (Molecular Probes #F6130, 1.6 mg in 0.2 mL of DMSO) was added in 20 μL aliquots to the bead suspension with vigorous stirring. After rotation for 1 h at room temperature, 0.8 mL 1 M glycine in 0.1 M $NaHCO_3$, was added, followed by another hour of rotation, addition of bovine serum albumin (BSA) to a final concentration of 2%, rotation for 15 min, centrifugation at $\sim 10\,000 \times \mathbf{g}$ for 3 min, resuspension and washing three times with 0.02% Tween 20, 1% BSA, 10 mM TrisCl pH8.5. The bead suspension was allowed to stand overnight, then centrifuged and resuspended in 1 mL 10 mM TrisCl pH8.5 (no BSA, no Tween20) and stored at 4 °C until use.

The amount of fluorescein in a 300-fold diluted bead suspension ($=2\times10^{-8}$ M) was estimated by comparison of the green fluorescence of a washed bead suspension in 10 mM bicine pH 8.4 with dilutions in the same buffer of a 50.7 μM solution of fluorescein (Molecular Probes #F36915 NIST traceable fluorescein standard), fluorescence measured with 490 nm excitation, 520 nm emission, 10 nm bandpass. The number of beads in the same diluted bead suspension ($=5.8 \times 10^9$/mL) was estimated by comparison of its red fluorescence with dilutions of the original stock bead suspension (3.9×10^{12} beads/mL according to the manufacturer), fluorescence measured with 580 nm excitation, 605 nm emission, 10 nm bandpass. The results of those two comparisons were then used to calculate the apparent number of fluorescein molecules per bead (=2070). Fluorescence measurements were carried out in 3 mL quartz cuvettes using a Hitachi Model F2000 fluorescence spectrophotometer.

The beads were also characterized by flow cytometry using a Becton–Dickinson FACSCalibur instrument, with 488 nm excitation, and fluorescence detection in three channels: 515–545 nm, 564–606 nm and >670 nm. This analysis yielded an estimate of fluorescein per bead equivalent to 2100 molecules of fluorescein in solution, in excellent agreement with the value measured by spectrophotometer. The coefficient of variation (CV) of the fluorescein signal in the flow cytometer was 16%, which is entirely accounted for by the Poisson statistical variation of the estimated number of photons (Shapiro, 2003) detected per bead in the fluorescein channel, ~40. Thus the fluctuation in fluorescein content per bead must be substantially less than 16%, but it is not possible to estimate the heterogeneity any more precisely than this.

Test specimens

To prepare bead specimens for imaging, 10 μL of bead suspension diluted 1:100 with dry EtOH was spread on a 22 mm #1.5 cover slip and allowed to air-dry. Twenty-five microlitres background solution was pipetted into a well formed using a Secure-Seal spacer (Molecular Probes #S24735). The cover slip with attached beads was inverted on this well of solution, pressed to make a firm seal and the edges of the cover slip were further sealed to the slide using nail polish.

A slide with a thin layer of fluorophore was prepared using fluorescein coupled to BSA. Fifty microlitres 10% BSA in water was mixed with 50 μL 0.1 M $NaHCO_3$ pH 8.8 and ~ 1 mg fluorescein succinimidyl ester (Molecular Probes #F6130) in 10 μL DMSO. After 1 h at room temperature the labelled protein was separated from free dye with a G25-Sephadex spin column. Five microlitres of labelled protein was spread evenly on a clean 22 × 22 mm cover slip, then rinsed with cold acetone. The cover slip was air-dried, the protein layer was fixed with 10% formaldehyde in PBS for 30 min, rinsed twice with 10 mM bicine pH 8.4, and mounted over a thin layer of the same buffer.

Microscopy

Images were collected on a variety of microscopes. In each case, the objective lens giving the brightest image was used and other optical components were chosen to give a pixel size of ~0.1 μm. In all cases, the refractive index of the immersion oil was carefully selected empirically to minimize spherical aberration. In the WF microscopes, the field stop was set to a diameter just slightly greater than the field of view of the camera. In the confocal microscopes, preliminary measurements established the range over which emitted fluorescence increased linearly with illumination irradiance (i.e. no significant depletion of the ground-state population). Linearity of AOTF response was established independently of fluorescence yield by measuring excitation light reflected from a mirror at various AOTF settings. The confocality of confocal microscopes is dramatically affected by lateral and axial chromatic aberration. To compensate for this, all spot scanning systems include an adjustable 'collimator' lens that attempts to make the focal position for excitation wavelengths identical to the focal position for emission wavelengths. When the system is truly 'confocal', the light collected by a pinhole with diameter equivalent to one Airy disk (image plane) should be ~70% of the total light that could be collected from a diffraction limited spot (Sandison *et al.*, 1995a). For spot scanning confocals, measurements were made of image intensity versus pinhole size to check that the expected plateau occurs at just beyond 1 Airy disk diameter, and re-alignment was carried out as necessary.

For each microscope, two types of data were collected: (1) multiple sequential images of the same field of view for measuring the rate of photobleaching and (2) 3-D stacks of optical sections with two excitation/emission wavelengths for intensity measurements on single beads.

(1) Photobleaching measurements utilized specimens with no AlexaFluor488 glucose in the background solution.

Fading of fluorescence in the green channel was recorded over time, at several levels of illumination irradiance.

(2) Small stacks of optical sections were acquired from fields having an appropriate density of well-separated single beads, using at least four different illumination irradiance/exposure time combinations for each specimen. Suitable regions were chosen by observation of the red fluorescence channel. In initial experiments, images were collected at a single visually identified focal plane, but it was found that variations in focus between different beads across the field of view was a large source of noise. Accordingly, all of the data presented here was acquired as 3-D stacks of three or five focal planes separated by 0.1 or 0.2 μm in *z*. For the fluorescein/AlexaFluor488 image ('green channel', 488 nm excitation, 500–550 nm emission), illumination irradiance and exposure times were different for specimens with different background fluorophore concentration, whereas the illumination condition for the 'red channel' images was the same for all specimens on a given microscope.

Zeiss LSM510 (spot scanning confocal). Plan-Neofluar 1.3 numerical aperture (NA) 40× or 1.45 NA 100× objective lens; green channel: 488 nm argon laser excitation, 500–550 nm emission; red channel: 543 nm HeNe laser excitation, >580 nm longpass emission; pixel size 0.1 μm; *z*-increment 0.2 μm; pinhole size 1.0 or 0.5 Airy disk units; pixel dwell time 1.28 μs; illumination irradiance set by laser tube current, which was held at 30% of maximum, and an acousto-optical tunable filter (AOTF; 0.5–3% used for these experiments)

Leica TCS SP2 AOBS (spot scanning confocal). A 63× PlanApo CS 1.4 NA objective lens; green channel: 488 nm laser excitation, 500–550 nm emission; red channel: 561 nm laser excitation, >580 nm longpass emission; pixel dwell time 2.5 μs; pixel size 0.1 μm; *z*-increment 0.2 μm; pinhole size 1.0 or 0.5 Airy disk unit. Illumination irradiance set by analogue laser tube current control (held at 'Low') and an AOTF.

Olympus FV1000 (spot scanning confocal). A 60× PlanApo 1.42 NA objective lens; green channel: 488 nm argon laser excitation, 500–550 nm emission; red channel: 543 nm HeNe laser excitation, >580 nm longpass emission; pixel size 0.1 μm; pinhole size 1 Airy disk unit; pixel dwell time 2 μs. Illumination irradiance set by laser tube current control (set to its minimum value) and an AOTF.

DeltaVision Spectris (WFD). Two different similarly equipped versions of this system were employed. The results were not significantly different; data from only one is reported here. Olympus IX70 base; 100× PlanApo 1.4 NA objective lens; Chroma Sedat Quad ET filter set #86000; green channel: 480–500 nm excitation, 507–543 nm emission; red channel: 543–568 nm excitation, 579–631 nm emission; CoolSnap HQ CCD camera (12-bit digitization, 6.4 μm pixels, full well capacity 18000 e^-, read noise 6 e^- rms; quantum efficiency (QE) ∼ 60% at 525 nm); 2 × 2 binning, image pixel size 0.13 μm in *xy*, 0.2 μm in *z*. Illumination irradiance controlled by software selectable ND filters (1, 3, 10, 32, 50, 100%T). Exposure times ranged from 80 to 500 msec. The illumination field stop was set to a diameter ∼30% greater than the field of view of the CCD.

Yokogawa CSU-10 (spinning disk confocal). Three versions of the spinning disk confocal using the Yokogawa CSU-10 scanhead were evaluated:

(1) Nikon TE2000U base, 100× 1.4 NA PlanApo objective lens. A Coherent 2.5 W ArKr laser and an AOTF were used for illumination and excitation wavelength selection (Prairie Technologies, Middleton, WI). A triple bandpass polychromatic mirror (Chroma # 53055) was used with single bandpass excitation and emission filters; green channel: 488 nm excitation, 500–550 nm emission; red channel: 568 nm excitation, 590–650 nm emission. Images were collected with a Hamamatsu ORCA-ER (12-bit digitization, 6.4 μm pixels, full well capacity 18000 e^-, read noise 8 e^- rms; QE ∼70% at 525 nm) and MetaMorph software, using 2 × 2 binning, image pixel size 0.13 μm. Illumination irradiance control set by AOTF. Exposure times 300 to 2000 msec.

(2) Nikon TE2000U base, 100× 1.4 NA PlanApo objective lens. A Melles-Griot 100 mW ArKr laser and Sutter filter wheels were used for illumination and excitation wavelength selection. A triple bandpass polychromatic mirror (Chroma # 53055) was used with single bandpass excitation and emission filters; green channel: 488 nm excitation, 500–550 nm emission; red channel: 568 nm excitation, 590–650 nm emission. Images were collected with a Hamamatsu ORCA-ER (12-bit digitization, 6.4 μm pixels, full well capacity 18 000 e-, read noise 8 e- rms; QE ∼ 70% at 525 nm) and MetaMorph software, using 2 × 2 binning, image pixel size 0.13 μm. Illumination irradiance set by analogue tube current dial (held at 'high' or 'low'). Exposure times from 75 to 8000 msec

(3) Nikon TE2000E base, 100× 1.45 PlanApo TIRF objective lens; Melles-Griot 50 mW ArKr gas laser; dual band-pass dichromatic mirror used with single wavelength excitation and single bandpass emission filter; green channel: 488 nm excitation, 500–550 nm emission; red channel: 568 nm excitation, 585–629 nm emission; Hamamatsu ORCA II ER camera, 14-bit digitization, 6.4 μm pixels, full well capacity 40000 e^- at 2 × 2 binning, read noise 4e^- rms; QE ∼ 70% at 525 nm; image pixel size 0.13 μm. Illumination irradiance set by analogue tube current dial (held at '8'). Exposure times 200 to 3000 msec.

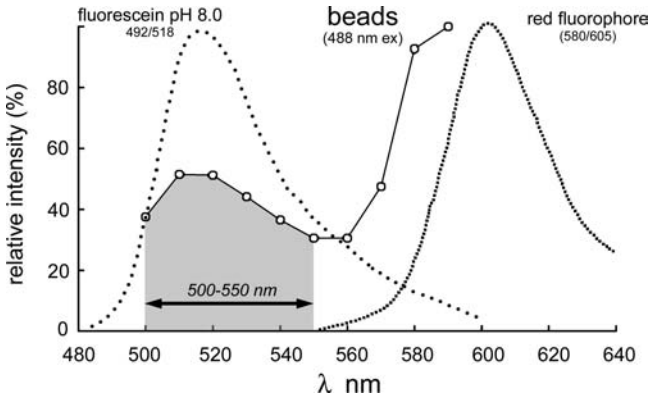

Fig. 1 Emission spectra of fluorescein with 492 nm excitation, the bead red fluorophore with 580 nm excitation, and beads containing both the red fluorophore and covalently attached fluorescein, with 488 nm excitation. The 500–550 nm emission band used for the bead intensity measurements is shown in grey.

Detector calibration

For each microscope system, the proportionality factor (g) relating image intensity values (grey levels) to the apparent number of photons detected was measured. The value was obtained from a graph of variance versus mean intensity for a set of images. To determine image variance, a pair of images of a uniform field was collected in rapid succession with no change in conditions. Subtracting one member of this pair from the other gives a difference image in which all systematic differences are removed, leaving predominantly Poisson noise with variance equal to twice the variance of a single image. Repeating this paired acquisition with several different illumination conditions and plotting the variances so obtained against the mean intensity value in each pair gives a straight line whose slope is $2/g$. This value changes with the gain setting for CCD cameras, and with PMT voltage (inversely proportional to V^n, $n = 6$–7) and pixel dwell time (increases linearly with t) for PMT based systems. For CCD cameras, gain was kept at the lowest setting. For the spot scanning confocals, g was determined for a range of PMT voltages and dwell times spanning the values used for the bead analysis. The signal from PMT detectors contains a small amount of 'extra' noise generated by the amplification process, and thus the variance computed as described above slightly overestimates the Poisson noise, resulting in a slight underestimate for g. For the range of PMT voltages used here (800–1100 v), the error in g is ~10% (Shapiro, 2003).

Data collection

Figure 1 shows the emission spectrum of the beads when excited at 488 nm. To eliminate contributions from the red (photostable) fluorophore, images were collected using 500–550 nm bandpass filters. Figure 2(B) shows the relationship between the background fluorescence from out of focus fluorophores and the depth of the solution under the cover slip imaged in a WF microscope with high NA objective lens. Background fluorescence increases until the underlying solution is >100 μm deep, even though the integrated intensity from a single point falls to zero when it is more than ~4 μm from the focal plane of the objective lens (Fig. 2B inset). To maintain consistent levels of background, slides were mounted over a solution with a depth of ~130 μm for all experiments.

Photobleaching measurements provided the calibration of the illumination system for each microscope that was prerequisite for comparison between different systems at equal illumination irradiance. We determined the rate constant for photobleaching, measured in units of fractional amount of photobleaching per acquisition. This allowed calculation of an *exposure index*; i.e. an exposure index of 0.02 corresponds to an illumination dose that would photobleach 2% of the fluorescein molecules on the bead during one image acquisition.

One 'acquisition' is defined as the imaging required to acquire all the information needed to determine total fluorescence emitted from the bead as well as the contribution from in-focus and out of focus background fluorescence, and thus to calculate net integrated bead intensity. Removal of out of focus background is accomplished automatically in a confocal when the light passes through the pinhole, so a single optical section of an in-focus bead contains all the information necessary for calculating net integrated bead intensity. In the WF case, deconvolution is required to remove the out of focus background, which requires information from the out of focus as well as in-focus planes. Thus the raw data for the performance comparison (one 'acquisition') comprises single images for the confocals, but small 3-D stacks for the WF microscopes. Stacks of three images separated by 0.2 μm in z were found to be optimal. Increasing the number of images in the stack did not increase the final image quality sufficiently to compensate for the increased illumination.

Image analysis

All measurements on images were performed using ImageJ (http://rsb.info.nih.gov/nih-image/). Results were exported as text files and imported into Excel for secondary calculations.

(1) Photobleaching curves:

Using the first image of the time series, upper and lower intensity threshold values were chosen to form a mask that excluded background regions and large clumps of beads, but included single beads and small clusters. The average intensity within the masked region was measured for images in the time series, corrected for background, divided by the value computed for the first time point and plotted versus cumulative exposure time on a semilog plot. The exponential decay constant was determined by a least-squares fit to the initial portion of this curve (<20% photobleaching). The values of the

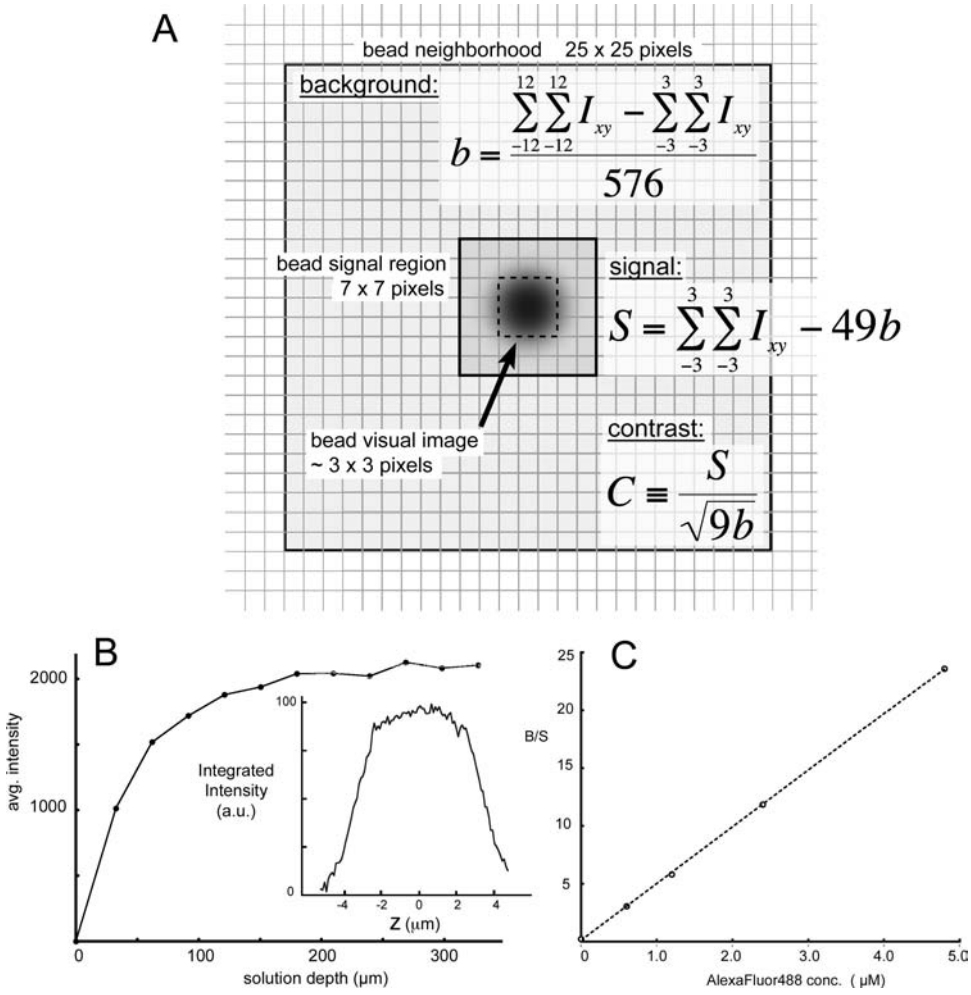

Fig. 2 (A) Diagram of the measurement regions described in the text. The diagram represents a portion of a digital image containing a single bead. Grey lines denote pixel boundaries. A grey-scale representation of the bead is at the centre of the diagram. *Raw bead intensity* is defined as the sum of the pixel values in the central 7 × 7 box. The local per-pixel *background* (*b*) is measured in a 25 × 25 pixel neighbourhood that excludes the central 7 × 7 box. *Net integrated bead intensity* (*S*) is calculated by subtracting the background in the 7 × 7 box (i.e. 49*b*) from the *raw bead intensity*. *Contrast* (*C*) is defined as the magnitude of *S* relative to the magnitude of the expected Poisson fluctuations in background over the central 3 × 3 pixels, which is approximately the area occupied by the visual image of a bead. (B) The relationship between average background intensity from out of focus fluorescence and fluorophore solution depth for an NA1.4 lens focused at the cover slip. The inset shows the integrated intensity from a single subresolution fluorescent bead as a function of defocus for the same objective lens and light source. (C) The ratio of background intensity from out of focus fluorescence to the net integrated intensity from a single bead (*B*/*S*), at various concentrations of Alexa Fluor 488 in the background solution. The beads are invisible by eye above 1 μM, but can be seen on a good monitor with contrast enhancement up to ~4 μM (*B*/*S* ~20)

decay constants measured at different illumination irradiance were plotted against the appropriate illumination system parameter (e.g. AOTF% setting, ND filter% transmittance, etc.), and the regression coefficients for the resulting straight line were obtained by least-squares fitting. In all cases these plots showed that the rate of photobleaching varied in strict linear proportion to the illumination irradiance parameter. The regression coefficients were used to interpolate values of the decay constant for each set of conditions used in collecting bead images. We used the interpolated decay constant in the form of an exposure index (*E*) as a measure of the illumination irradiance-time integral for each experimental condition. If the decay constant for a particular acquisition is denoted by D, then the exposure index is given by $E = 1.0 - e^{-D}$.

(2) Net integrated intensity of single beads:

A plugin for ImageJ that implements a feature point detection and tracking algorithm (Sbalzarini & Koumoutsakos, 2005) was used to locate candidate single beads in the middle optical section of the red channel of each stack. The plane of best focus and refined centre of mass of the bead was then determined by a search in the neighbourhood of this initial location. In specimens with low background

fluorophore concentration, this search could be carried out accurately in both red and green channels, which served to establish the lateral and axial displacements between the red and green channel images characterizing each microscope system (typically < 1 pixel). Knowing these displacements, the location of the green bead image could then be accurately determined from the position of the red bead image in specimens where the high background fluorophore concentration lowered the visibility of the green bead so much that direct searching in the green channel was not feasible.

Pixel intensities within a 7×7 pixel box centred on the refined centre of mass of each bead were summed to give the raw integrated bead intensity in the red and green channels (Fig. 2A). From this sum, the local background intensity (b), measured in a 25×25 box surrounding each bead, was subtracted to give the net integrated single bead intensity (S). Specifically, if the sum of the 49 pixels in the 7×7 box is denoted T7, the sum of the 625 pixels in the entire concentric 25×25 box is T25, then $b = (\mathrm{T25} - \mathrm{T7})/576$ and $S = \mathrm{T7} - 49b$.

Displaying the histogram of the net integrated red bead intensities for all the images collected on a microscope system typically revealed a bimodal distribution with two clear peaks, corresponding to single beads and to pairs of beads too close together to be resolved. This histogram was used to set a threshold that excluded double beads. The pared list was then further filtered to remove beads near the edge of the image, or too close to a neighbour, situations that interfered with measuring the local background intensity.

Results

Design of the test specimen and performance assay

As described in Methods, 0.2 μm beads, labelled with a small but constant amount of a green rapidly bleachable fluorophore (fluorescein) and much larger quantity of a red photostable fluorophore (Fig. 1), were adsorbed to a cover slip and mounted over a ~130-μm-thick layer of solution of a green photostable fluorophore (Alexa Fluor 488) with excitation/emission spectrum essentially identical to fluorescein. The red fluorophore on the bead is not used in the quantitative measurements. It serves merely to provide a means of locating the beads and setting the focus without photobleaching the green fluorophore, even when the green fluorescence of the beads is 'invisible' due to high background.

It is useful to have a parameter for quantitatively characterizing specimens with respect to the severity of problems caused by out-of-focus background fluorescence. Such a parameter could be a suitable guide in deciding which type of microscope would be most effective. For this purpose, an informative parameter is the ratio of background intensity to the intensity of the smallest objects of interest in the focal plane. In terms of the test specimens used here, the parametric ratio is taken as the background intensity summed over a bead-sized area, denoted B, divided by the total net intensity from a single in-focus bead, denoted S (Fig. 2A). The visible extent of a single bead is $\sim 3 \times 3$ pixels, so B is taken as nine times the average background per pixel; i.e. $B = 9b$ (see Fig 2A). Figure 2(C) shows the change in the ratio B/S as a function of concentration of AlexaFluor 488 in the solution under the cover slip. The dashed line is a least-squares fit to the data points.

Note that by definition B/S characterizes a specimen in terms of its image in a WF microscope; that is, in a microscope in which all out of focus fluorescence appears in the image (as opposed, for instance, to a confocal in which much of the out of focus fluorescence has been blocked from the image). Anticipating results to be described below, as the concentration of AlexaFluor 488 in the background solution increases beyond 5 μM it will be found impossible to measure S, and hence impossible to evaluate B/S directly, using a WF microscope with a CCD detector of typical well capacity and dynamic range. However, it is necessary to have a parameter that quantitatively compares the severity of out of focus fluorescence over the entire range of specimens used here, not limited merely to specimens suitable for imaging by WF microscopy. For the purposes of this paper, we could simply use the background AlexaFluor 488 concentration as that parameter. It is however preferable to use a parameter like the B/S ratio that will be useful in principle for any specimen (i.e. based on image intensities), not just the artificial bead test specimens we consider here. The regression line in Fig. 2(C) provides an indirect method of determining the expected value of B/S at any concentration of AlexaFluor 488, regardless of whether the denominator could be directly measured by WF microscopy. The equation for this regression line was used to calculate the expected ratio B/S for all concentrations of AlexaFluor 488, including those higher than 5 μM. To avoid confusion, it is best to not use the symbols B and S in this connection, since those symbols will be in constant use throughout the remainder of the paper for directly measured quantities in images from both WF and confocal microscopes. For want of a standard name for this ratio, we will call it the *Haziness index* and denote it H.

Figure 3 shows a montage of typical images obtained from the different types of microscopes at various concentrations of background fluorophore, i.e. from specimens with different values of H. As expected, image quality decreases with increasing H. For this overview figure, the images have been digitally contrast enhanced, but even so the decreasing visibility of single beads with increasing H is obvious. To characterize different modes of microscopy, we used beads in a defined background to measure the integrated intensity after background subtraction (i.e. the net intensity) and the contrast generated between the beads and the background. In addition, we measured the fluctuation of the net integrated intensity within populations of beads under identical conditions to assess the precision of different imaging modalities.

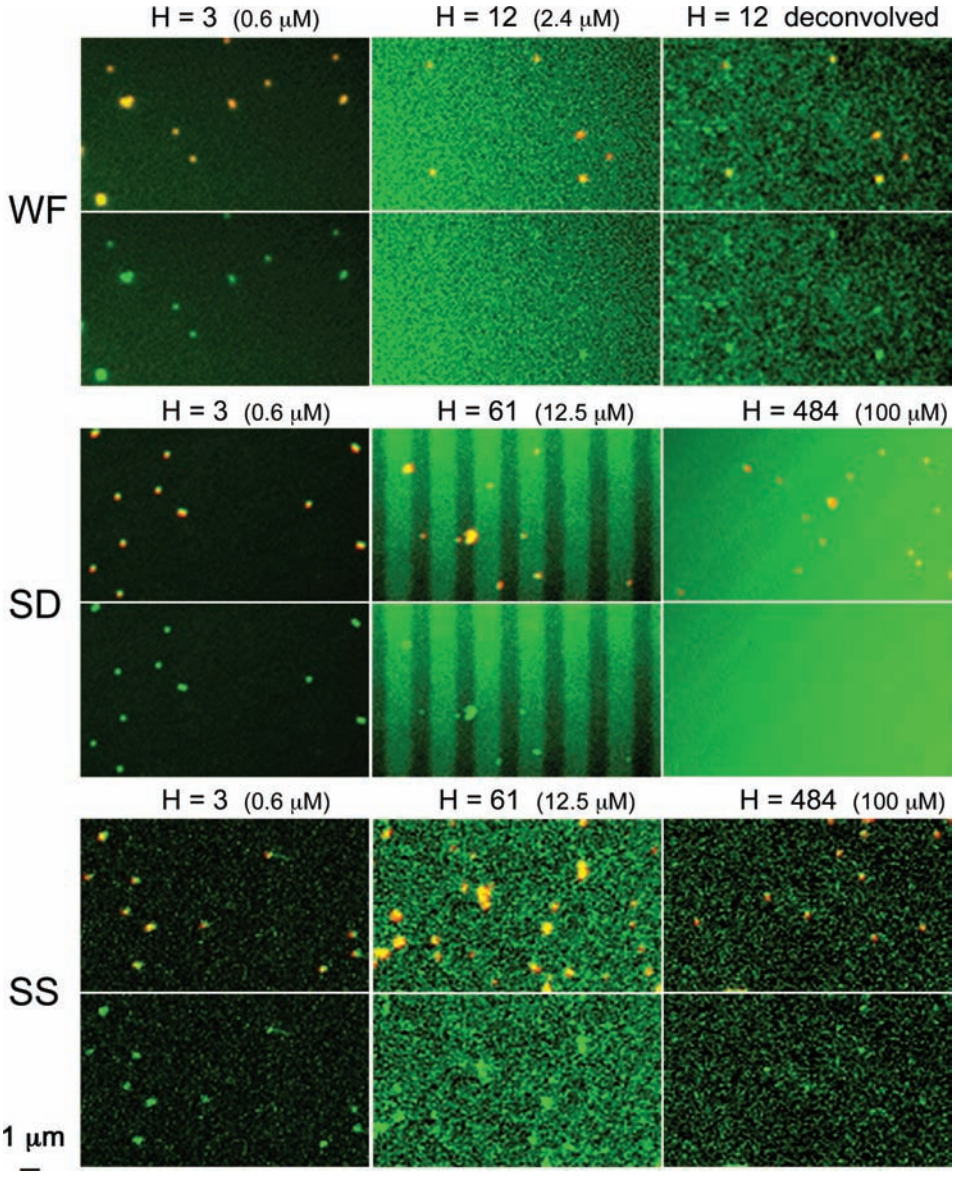

Fig. 3 Montage of images of beads with different amounts of out of focus fluorescence (*haziness index, H)* from wide-field (WF), spinning disk (SD) and spot scanning (SS) confocal microscopes. The concentration of AlexaFluor 488 in the background solution is given in parentheses for each image. For this figure, nonlinear contrast stretching has been employed to optimize single bead visibility in the green channel in the images with high background. The smallest objects visible in the red channel are single beads. Larger objects are clusters of multiple beads.

Image collection and data analysis

Two types of data were collected on each microscope: (1) a series of sequential images of the same field of view of beads in the absence of background fluorophore, used for measurements of photobleaching rate to calibrate the illumination system and (2) optical sections of beads in the presence of different amounts of background to determine the average bead intensity and its standard deviation at several levels of illumination dose.

Figure 4 shows typical photobleaching curves for a range of illumination dose on different types of microscopes. The initial rate of bleaching is accurately modelled by a single exponential decay process, but beyond ~30% loss of intensity the curves sometimes become more complex. The decay constant for the initial rate of photobleaching increases linearly with illumination irradiance over the entire range of irradiance achievable in the WF and spinning disk systems (Fig. 4D and E), but spot scanning confocals can easily achieve irradiance high enough to deplete the population of ground-state fluorophores, giving rise to the curvature seen in Fig. 4(F). For measurements on the spot scanning confocals, optical sections were collected with illumination irradiance restricted to the linear range. Pixel dwell times were also kept short (<3 μs), since it was

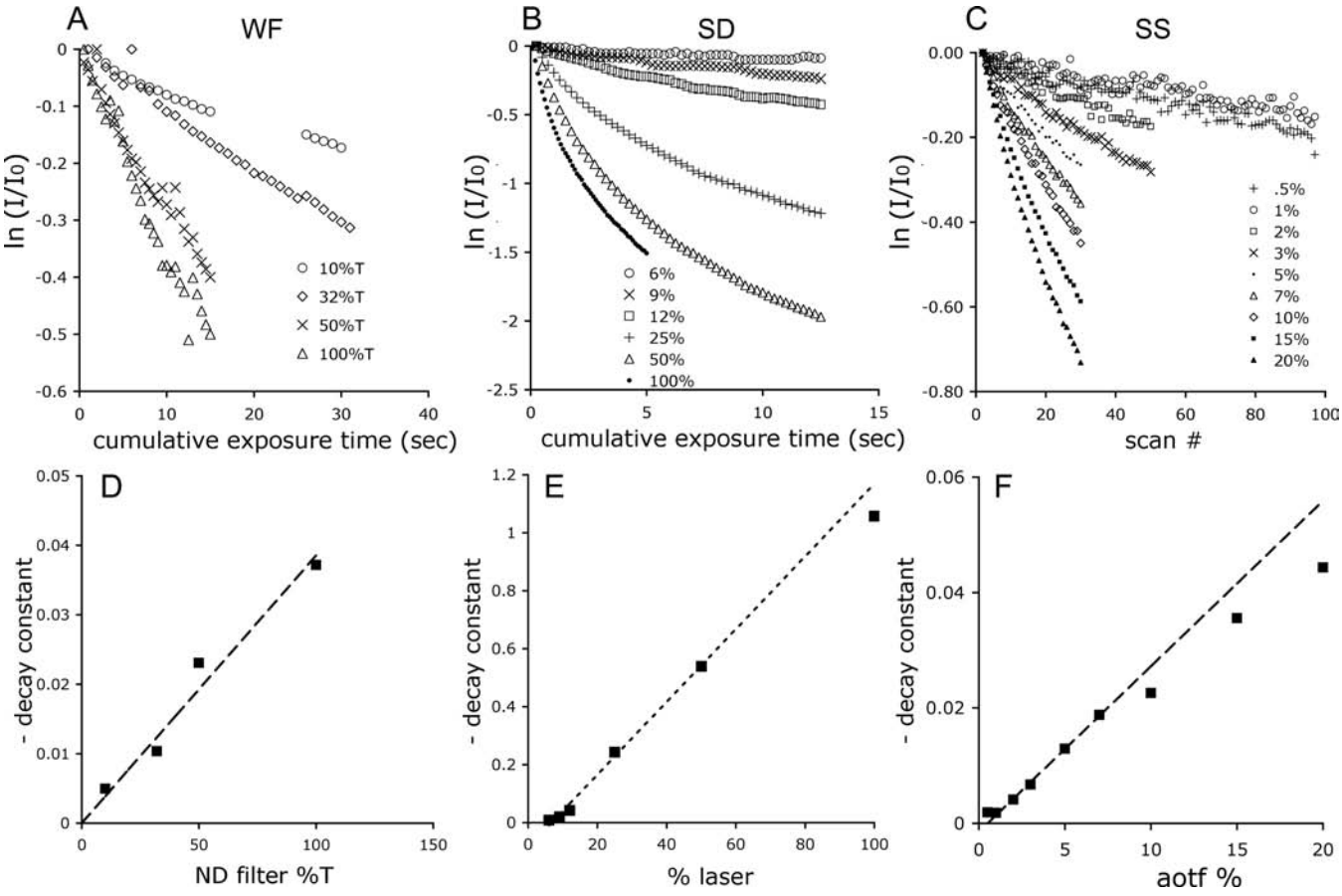

Fig. 4 Photobleaching curves used for calibration of the illumination systems. (A, B, C) Semilogarithmic plot of the normalized fluorescence intensity ($\ln(I/I_0)$) of the green fluorophore in a series of images of the same field of view of beads, taken in WF, spinning disk confocal (SD) and spot scanning confocal (SS) microscopes. In each system, decay curves for several different levels of illumination irradiance are shown. (D, E, F) Least-squares estimated exponential decay constants from (A–C) (solid squares) plotted against the systems' illumination irradiance parameters, with the least-squares fits to each dataset shown as a dashed line. Note the deviation from linearity for the spot scanning confocal system (F) at AOTF settings >10%.

found that the illumination irradiance required for significant ground state depletion decreases as the dwell time per pixel is increased. This observation is consistent with ground state depletion occurring due to accumulation of fluorophores in a state with a relatively long lifetime; i.e. in a triplet state with lifetime of many microseconds, rather than in the excited singlet state, which has a lifetime of ~5 ns.

The measurements of Fig. 4 show that for each microscope photobleaching is proportional to illumination irradiance. In order to use photobleaching as a standard measure of illumination irradiance, it is necessary that the proportionality constant relating photobleaching rate to illumination irradiance be the same for all microscopes. As there is an approximately 10^6-fold difference between illumination irradiance in WF and spot-scanning confocals, the possible existence of different mechanisms of photobleaching and hence different proportionality constants needs to be considered seriously. To rule out this possibility, we made one set of direct measurements of both illumination dose and photobleaching at these two extremes of illumination irradiance. Imaging conditions in a WF and a spot-scanning confocal were adjusted by trial and error to give equal amounts of photobleaching during acquisition of a single image of a very thin layer of fluorescein (see Methods), using the same specimen slide for both. With these same conditions, measurements using a light meter at the specimen plane showed that the illumination dose to the specimen was identical for the two microscopes within the error of the measurement (<10%). Taken together, our measurements show that over the range of illumination irradiance spanning typical WF and spot-scanning confocal microscopy with CW lasers, the amount of photobleaching bears a fixed linear relationship to the illumination dose.

How many photons are acquired from a single bead?

Figure 5(A) shows a typical comparison of photon acquisition efficiency among several WF, deconvolution, spinning disk

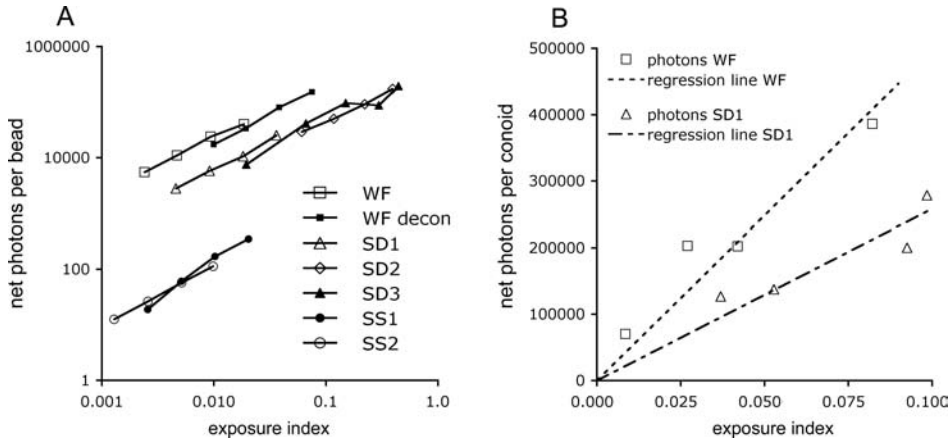

Fig. 5 (A) The average number of photons acquired from a single bead in images from WF, deconvolved WF (WFD), spinning disk confocal (SD) and spot scanning confocal microscopes (SS), plotted against standardized illumination dosage. Note that both axes are logarithmic. (B) Net photons acquired from a small biological test object, the conoid of YFP-tubulin transgenic *Toxoplasma gondii*, when imaged with two of the same microscopes as used for (A), a WF and a spinning disk confocal microscope, plotted against standardized illumination dosage. The dashed lines are least-squares fits to the data points. Note that both axes are linear in this comparison.

and spot scanning confocal microscopes expressed as the total number of photons above the background collected from single beads as a function of 'exposure index' (E). The exposure index is a system independent measure of the amount of excitation light used in forming the images of single beads. It is determined for each system by measuring photobleaching rates, and can be interpreted as the fractional bleaching caused by acquisition of a single image (see Methods). Note that Fig. 5(A) is on a logarithmic scale to allow the plotting of all three types of microscope on one graph. On this scale, the differences between individual microscopes (i.e. microscopes from different manufacturers) are very small compared to the differences between optical designs. The rank order of efficiency is as expected: for the same amount of illumination, WF microscopes acquire more photons than a spinning disk confocal, and both of these collect far more than a spot scanning confocal. What was unexpected is the scale of the differences in efficiency, ~4-fold between WF and spinning disk, and ~50-fold between spinning disk and spot scanning confocals. Some of the difference between spinning disk and spot scanning confocal is accounted for by the large difference in quantum efficiency between the CCD camera of a spinning disk system (~60%) and the PMT of a spot scanning confocal (~10%), but the reasons for the remainder of the efficiency gap are unclear. Equally surprising from a theoretical standpoint is the large difference between WF and spinning disk confocals, approximately fourfold in this set of comparisons. We made no effort to ascertain which of the many differences in optical or electronic components is responsible for the large gaps, since the different microscopes had hardly anything in common (different objective lenses, light sources, filters sets, cameras). However the data all fall on the same lines, even though collected from systems supplied by several different vendors of each optical design, on microscopes located in different cities, over an interval of >1 year. It therefore seems that the differences in performance shown in Fig. 5(A) are a reproducible feature distinguishing the different classes. It is likely that the specimen used for the performance assay accentuates the differences between optical designs, because comparison using a somewhat larger, much brighter, and more photostable biological test object (Fig. 5B; YFP-tubulin in the ~0.4 μm diameter conoids of *Toxoplasma gondii*, Swedlow *et al.*, 2002; Hu *et al.*, 2006), we observed a ~2-fold difference between a WF and a spinning disk system, using two of the same microscopes used for Fig. 5(A). The conoid is ~2× larger in diameter, ~10-fold brighter and ~10-fold less prone to photobleaching than our test beads; in other words, it is much easier to image. A general observation throughout the course of our experiments has been that differences between microscopes become much more pronounced as the specimens become more challenging. Therefore, tiny photolabile beads of only modest brightness are a good test object for performance comparisons.

It is quite instructive to compare the performance shown in Fig. 5(A) with the same measurements on specimens with increasing amounts of out of focus fluorescence. Figure 6 shows the same group of microscopes, but with specimens containing 4.8, 25 or 100 μM AlexaFluor488 in the background solution, corresponding to **H** values of approximately 25, 120 and 480. For CCD detectors, increasing background fluorescence mandates a decrease in exposure index to avoid detector saturation. As a consequence, the net photons per bead decreases by more than 10 fold (note the 10-fold decrease in vertical axis scale for each change in background fluorescence), since most of the detector capacity is being used for background photons. The pinholes of the

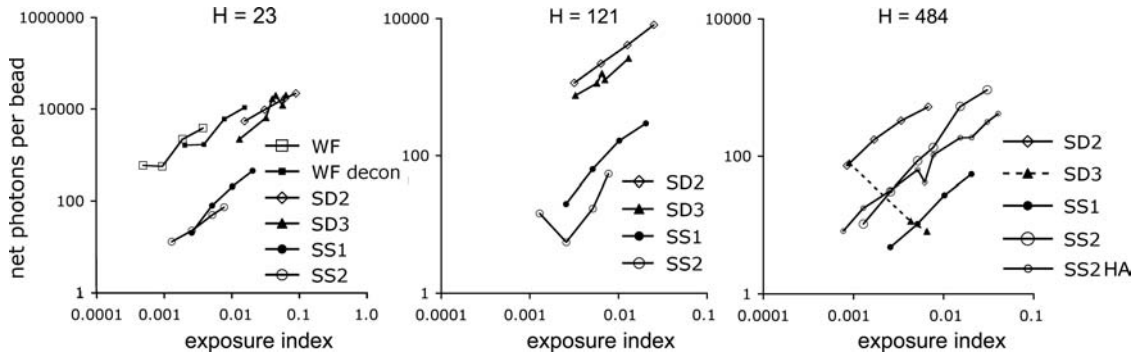

Fig. 6 Average net photons acquired from a single bead by microscopes of different types, plotted against standardized illumination dosage, for three different levels of out-of-focus background (*haziness index*, **H**), on a logarithmic scale. In the $H = 484$ (100 μM AlexaFluor 488 background solution) plot, 'SS2 HA' indicates spot scanning confocal #2 operated with the pinhole set to 1/2 Airy disk diameter.

spinning disk remove most of the background, permitting higher exposure before detector saturation occurs. For the PMT based spot scanning confocals, exposure does not need to be reduced at high background levels because the gain can be varied over several orders of magnitude, and the number of photons collected per bead can therefore be kept constant.

Note that in Fig. 6, even at the highest values of H, the number of photons per bead in images from one of the spinning disk confocals is almost 10-fold larger than for any of the spot scanning devices. However, the benefit gained from this considerable advantage can be judged only after considering two other performance measures described below, the image contrast and the precision of measurement of net bead intensity. For the other spinning disk, SD3, the background fluorescence leaks into the image so much that measurements of signal from the bead are dominated by noise, and completely unreliable (see 'Contrast', below). In trying to interpret the apparent precipitous drop of net photons collected by SD3 with increasing exposure index in the $H = 484$ graph, it is helpful to keep in mind that the ~50 net photons at highest E represents the average difference between values of background plus signal of ~400 050 photons and background alone of ~400 000 photons. Unsurprisingly, the result is unreliable.

To validate our measurements of photon efficiency, it is worth considering how many photons could in principle be collected from a single bead and comparing this value to actual measurements. By comparing the fluorescence of a dilute bead suspension with that of a fluorescein solution standard in a fluorescence spectrophotometer, and also by flow cytometry analysis (see Methods), the number of fluorescein molecules per bead was estimated to be ~2000. Using published values for the quantum yields for fluorescence (0.9) and for photobleaching (3×10^{-5}) of fluorescein in aqueous solution, a fluorescein molecule would emit ~30 000 photons on average before photobleaching (Tsien & Waggoner, 1995), or $\sim 6 \times 10^7$ photons per bead. Taking the overall efficiency of the WF microscope/CCD camera to be ~6% (calculation in Sandison *et al.*, 1995b, adapted to our conditions), and taking into account the fraction of the fluorescein emission spectrum admitted by our 500–550 nm filters (~75%), the expected maximum number of acquired photons per bead is $\sim 2.7 \times 10^6$. The highest efficiency measurement from our entire dataset (WF microscope, no background fluorophore) is in excellent agreement with this calculation; 23 842 photons per bead acquired on average at an exposure index of 0.0093, which extrapolates to 2.6×10^6 photons per bead at complete photobleaching. This calculation verifies the accuracy of our measurements for the WF microscopes, and therefore suggests that there is room for considerable improvement in the photon efficiency of spinning disk confocals, and even more room for improvement of spot scanners.

Contrast: How visible is the bead against the background fluorescence?

A quantitative measure of contrast that correlates well with the perceived visibility of single beads is given by the ratio of the average above-background intensity of the bead to the average size of fluctuations in a bead-sized patch of background. The visual impression of the extent of a single bead is limited to ~3 × 3 pixels and, empirically, the visual perception of contrast seems to depend on background fluctuation over a region of roughly that size. Motivated by this admittedly somewhat subjective empirical observation, we therefore defined contrast as $S/B^{1/2}$, where B, the sum of the background in a bead-sized patch, is again defined as 9*b*. (Note that in a superimposed average of many beads, intensities significantly above background are detectable over a larger area, and therefore the definition of S utilizes a 7×7 box in order to ensure complete collection of the bead signal. However, most of S is contained in the central 3×3 pixel region.)

When the out of focus fluorescent background is low (H less than 20), the contrast in a WF or WFD image surpasses

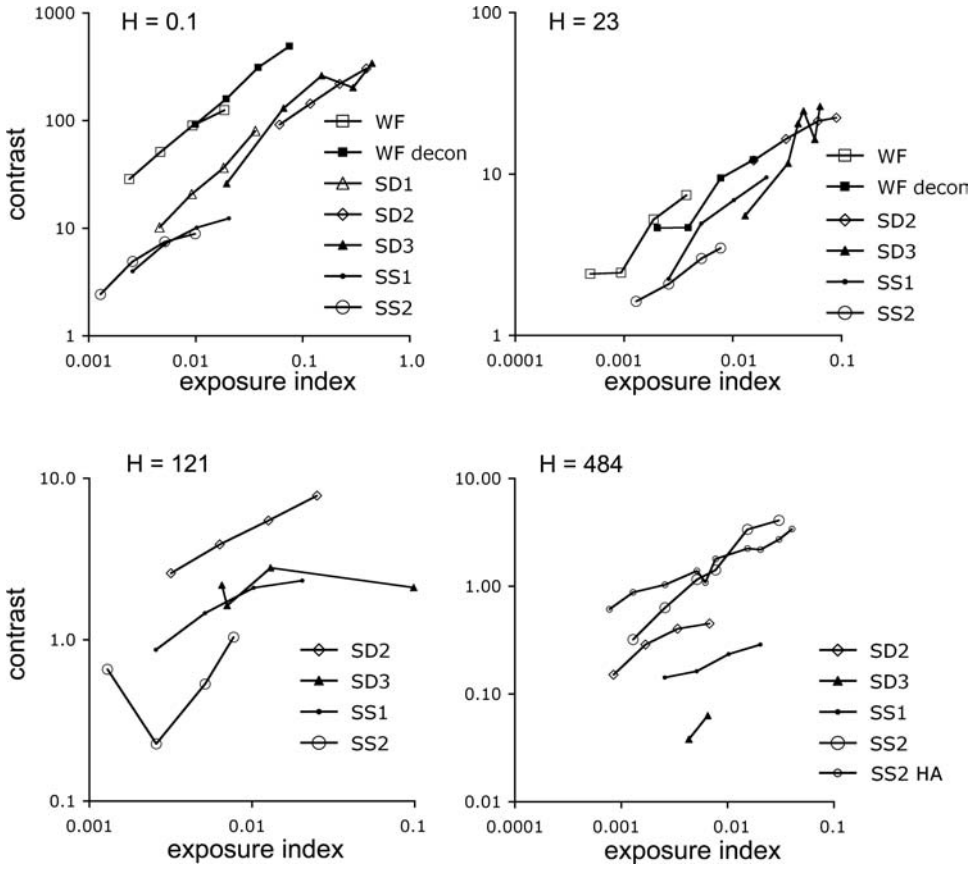

Fig. 7 Contrast in images of single beads acquired on different types of microscopes at 4 different levels of out-of-focus background (*haziness index*, *H*), plotted against standardized illumination dosage, on a logarithmic scale. Note the 100-fold difference in vertical scale among the graphs. In the $H = 484$ (100 μM AlexaFluor 488 background solution) plot, 'SS2 HA' indicates spot scanning confocal #2 operated with the pinhole set to 1/2 Airy disk diameter.

all of the confocal methods (Fig. 7). As with the photon efficiency measure, at low H differences between individual microscopes are minor compared to the differences between types. At higher H, and over a broad range ($20 < H < 200$) spinning disk confocals achieve superior contrast, with differences in performance between individual microscopes becoming increasingly significant as H increases. Notice that at the highest H, the difference in contrast between SD2 and SD3 is at least 4-fold, which is in accord with the dramatic difference in photon efficiency between these two spinning disk confocals shown in Fig. 6 at this level of H. For SD3, fluorescence from the background has so contaminated the image that the signal from the bead is lost in the noise. For $H > 200$, single beads are visible only in the images from a spot scanning confocal (cf. Fig 3), and then only with reduced pinhole size. At these very high levels of background, the contrast is extremely sensitive to excess illumination irradiance. If ground-state depletion becomes significant, increased exposure index reduces contrast, as the illumination then excites out of focus fluorophores more efficiently than fluorophores in the focal plane.

Precision: How much variation exists among individual images from a population of identical beads?

Images of 20 to 200 individual beads were analyzed under each set of conditions on each microscope, a total of more than 32 000 bead measurements. The average number of photons above background from single beads was divided by its standard deviation to give a measure of SNR. If the dominant source of noise in the measurements were stochastic fluctuations in the number of photons acquired per bead (Poisson noise), then the SNR should increase in proportion to $S^{1/2}$. However, this was not the case; the noise measured was far greater than the expected contribution of Poisson noise (see Discussion) and the SNR for individual microscopes showed no consistent trend relative to exposure (data not shown). Accordingly the measurements at all exposure indexes were combined to give an average SNR for each value of H (Fig. 8). Although some differences between individual microscopes of each type were apparent, they were relatively minor, and unimportant for the present discussion, so SNR values were lumped together

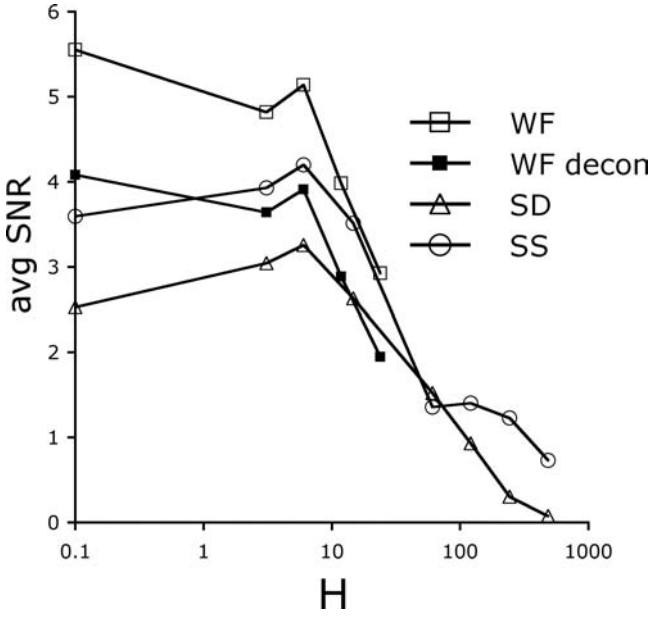

Fig. 8 Average SNR in populations of images of single beads acquired on different types of microscopes, plotted against the haziness index H (logarithmic scale for X-axis only). Data for all illumination dosages and all individual microscopes of each class have been lumped together.

for each optical design to simplify the presentation. Figure 8 shows a plot of SNR versus H on a logarithmic scale for WF, WFD, spinning disk and spot scanning confocals. At low H, WF gives the most precise measurements, but the SNR drops rapidly as H increases, and plummets to near zero at $H > 20$. Deconvolution, though it improved contrast significantly, amplifies the differences between individual bead images, lowering SNR. Among the confocals, the data from spot scanning devices is significantly more precise than from spinning disk confocals. As H exceeds 50, SNR begins to fall for both types of confocal, a practical limit for acquiring useful data being $H = 500$ to 1000 with a spot scanning confocal.

A striking and disappointing aspect of Fig. 8 is the low SNR achieved even under the best conditions. Several factors contribute to this (see Discussion) but one very obvious problem is variation in illumination and sensitivity across the field of view. Figure 9 shows representative examples of this, by no means the worst examples encountered, from all three types of microscope. Perhaps surprisingly, the spot scanning confocals are the least offensive in this regard. Spinning disk confocals are the worst offenders. The most obvious problem is the banding seen in Fig. 9, also prominent in Fig 3. This is believed to be partly the result of aliasing induced by the mismatch between two temporal sampling intervals: the interval between successive pinholes sweeping over the same pixel, and the integration time on the CCD (Chong *et al.*, 2004). Wobble of the spinning disk also leads to banding. In addition to banding, quite a large nonuniformity in response is indicated by the marked shading of the peripheral field of view in images from spinning disk confocals.

Figure 9 makes it clear that post-acquisition processing to compensate for nonuniform illumination and detection across the field of view ('flat-fielding') would significantly improve the SNR, but further analysis of the data reveals that other equally large sources of excess variance are present. For the most precise subset of the data (WF images with zero background fluorophore), restricting the measurements to beads located within the central 1/10th of the field of view reduced the variation in net photons by 2-fold to ~10%, but the residual variation is still more than 10-fold larger than expected on the basis of photon counts, even for this best-case subset.

Discussion

Our assay has, for the first time, quantified the differences between the major modes of 3-D fluorescence microscopy when collecting images in 3-D with different levels of background fluorescence. We find relatively small differences between different commercial versions of the same mode, but striking differences between the different modalities. We find a 2- to 4-fold difference in the light efficiency of WFD systems and spinning disk systems and a further 50-fold difference between spinning disk and spot scanning systems.

Our approach extends a previous report, where a simpler performance assay utilizing the SNR in single pixels of an

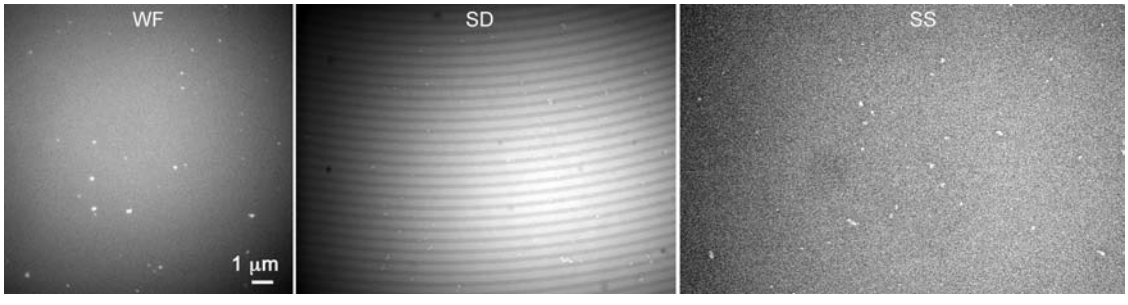

Fig. 9 Nonuniformity across the field of view in images from samples with high background acquired on (*left to right)* WF, spinning disk confocal (SD) and spot scanning confocal (SS) microscopes. The bright objects are large clumps of beads. Single beads are not visible in the presence of this much background at this low magnification.

image of small photobleachable beads was described. This simpler approach has been recently used to characterize the difference between spot scanning and spinning disk confocal microscopes (Wang *et al.*, 2005). However those comparisons used a test specimen with no out of focus fluorescence. Clearly, when comparing different types of confocal imaging modes or confocal and WFD microscopy, some account needs to be taken of their relative efficiency in removing the contribution from out of focus regions. For example, most spot scanning confocals have a continuously adjustable pinhole diameter, whereas in spinning disk confocals, the pinhole size is fixed or adjustable only in large steps, thus sometimes necessitating use of a larger than optimal pinhole. For a thin specimen that has little or no fluorescent material in out of focus planes, a larger pinhole diameter would be an advantage, allowing capture of more of the emitted fluorescence in a single optical section. However, this advantage turns into a serious liability as the amount of out of focus fluorescence in the specimen increases, because the larger pinhole passes more out-of focus fluorescence and degrades the image quality.

WF, spinning disk and spot scanning confocal microscopes differ greatly in many ways that potentially affect their performance in imaging thick specimens, so that *a priori* prediction of relative performance, even in theory, is difficult. However, the influence of some of those differences can be understood in at least a qualitative manner. For instance, WF microscopes collect an entire image plane ($\sim 10^6$ pixels) in parallel, while spinning disk confocals collect $\sim 10^3$ pixels in parallel, and spot scanners collect 1 pixel at a time. Collecting a large number of points (spinning disk) or an entire image plane (WF) in parallel of course gives a large improvement in speed compared to single spot collection, but there are in addition two important consequences of the parallel data collection that improve the image SNR as well. First, the rapid fluctuations in illumination irradiance that degrade images from spot scanning confocals (Swedlow *et al.*, 2002) are greatly attenuated. Second, the ability to use CCD detectors with quantum efficiency of $\sim$60% instead of PMT detectors with $\sim$10% quantum efficiency is an enormous advantage. However, in both WF and spinning disk systems, there are other factors that may outweigh these advantages for some specimens. In the WF case, it is necessary to collect a 3-D stack instead of a single image plane to address out of focus blurring, even if 3-D information is not needed, thus increasing the exposure and photobleaching. Additionally, the calculations necessary to remove the contributions from out of focus planes tend to amplify the noise. In specimens with modest background, the amplification of noise stemming from the deconvolution calculations is usually compensated by the increase in signal due to collection of additional images (Swedlow *et al.*, 2002). However, in samples with a very large out of focus component, collecting additional focal planes may give a much larger increase in total background than in total signal, reducing the SNR. In the multiple-spot confocals (spinning disk or pinhole array scanning), the loss-of confocality associated with 'cross-talk' between different pinholes (Conchello & Lichtman, 1994) may begin to degrade contrast and SNR in images of specimens that are thicker than $\sim d/(2 \tan \alpha)$, where d is the spacing between adjacent spots, and α is the objective aperture half-angle, both measured in object space. This cross-talk is likely to be at least partially responsible for the lower contrast achieved by spinning disk compared to spot scanning confocals at very high H in the current measurements.

While we have compared photon efficiency, contrast and SNR is this study, other measures, such as sensitivity to optical imperfections and speed of data acquisition, might be more important for particular applications. For example, a significant difference between WFD and confocal microscopes is manifested by the effect of optical aberrations on their performance (Sandison *et al.*, 1995b). In a WF microscope, imperfectly matched filter sets and lens chromatic aberration lead to wavelength-dependent axial and lateral image shifts and differences in magnification, with only minimal effect on image intensity. However, in spot scanning or spinning disk confocals, image shifts are minimal but image intensity is severely affected by chromatic aberration. The speed of imaging, both in terms of duration of a single exposure and the maximum repetition rate, is also frequently a major consideration. Even when only one 3-D stack of images at a single time point is needed, the speed of imaging can be limiting, since any significant movement of the specimen during collection of the stack makes 3-D deconvolution problematic, and complicates the interpretation of relationships between objects in different planes.

An extremely powerful methodology for cell biologists was opened with the advent of the fluorescent proteins (Giepmans *et al.*, 2006). An important aspect of this methodology is the capability it confers to count the number of molecules of a specific protein in a living cell. Accurate counting requires accurate measurement of fluorescence, and because the purpose of counting is inevitably to compare two counts, the precision of counting is of central concern. Ideally the standard deviation of a set of measurements, such as net photons in the image of individual beads, ought to be predictable based simply on the known average number of photons in the signal. In practice, the observed standard deviation of net photons collected from populations of identical beads is much larger, 10 to 100 fold, than the prediction. Every microscope performed far worse than the limit imposed by Poisson statistics, suggesting that substantial improvements can be made in commercial digital imaging systems.

What accounts for the very large discrepancy between the SNR expected on the basis of photon counts and the actual measured performance? Experimentally, we find that over a small neighbourhood, the background (b) standard deviation is typically only very slightly larger than the expected value, $b^{1/2}$. Therefore very little extra noise is added by

the detectors or associated electronics. Instead, the excess variability must be already present in the number of photons delivered to the detection system by the microscope optics. As described under Results, part of the problem is variations in illumination and sensitivity across the field of view. Another problem is small amounts of defocus, due either to incomplete flatness of field correction in the objective lenses, or random axial displacements of the beads. Although the data analysis incorporated a search for the local optimum focus (see Methods), for practical reasons the z-increment between planes in the 3-D stacks could not be made smaller than 0.2 μm. Thus a random focus error of up to 0.1 μm could be present. For a 1.4 NA objective lens in a WF microscope, a focus error of 0.1 μm would decrease the peak pixel intensity by ∼15%. The effect on integrated intensity would be much smaller, but the integration would have to be carried out over a large neighbourhood. In practice, the integration in our data analysis was limited to a neighbourhood of 4 × 4 Airy disk diameters, so a small effect of defocus error on integrated intensity is probable. In the confocal microscopes, both the peak and the integrated intensity would decrease by ∼15% with 0.1 μm defocus. To estimate an upper limit for the consequences of such a defocus error, random normally distributed decrements in the net photons per bead ranging from 0% to 15% were applied to a model population of beads with average photon counts the same as our WF microscope data. In this worst-case scenario, the calculated SNR fell by ∼4 fold from that predicted by Poisson statistics of the average photon counts. However, this calculated SNR is >10-fold higher than the actually observed SNR. We conclude that defocus errors make only a minor contribution to the low precision of the measurements. Adding in both the maximal defocus error and the most pessimistic estimate of the heterogeneity in fluorescein content per bead (see Methods) leads to a further 2- to 3-fold decrease in expected SNR, but the final result is still several-fold better than actually observed. It is clear from these considerations that the precision of all types of digital microscope imaging systems used for 3-D microscopy could be substantially improved. Improvements in precision would immediately increase the value of these systems for live-cell imaging, where the loss of precision typically cannot be compensated by collecting more photons.

So that our work may achieve its goal of establishing guidelines for choosing between different modes of 3-D microscopy, we summarize here the main result of our comparisons. First, the magnitude of a simple specimen-dependent parameter that we have called the *haziness index* (H) provides a numerical criterion for choosing the most appropriate mode of microscopy. H is computed as a ratio of background in a WF image over signal: the intensity of background from out-of-focus fluorescence divided by the intensity of the fluorescence from a small in-focus object. 'Small' here means comparable in size to the Airy disk. In practice, the intensity of such small objects will be unmeasurable in a WF microscope for specimens with very large background, so it will typically be necessary to estimate H by some indirect means. For instance, it will often be possible to find a thin edge of the specimen where the background is not overwhelming and measurement of the intensity of some small object of interest is possible in a WF microscope. This number could then be combined with a WF measurement of the background in the thickest region of the specimen to give a good estimate of H. Alternatively, if both signal and background could be measured in a thin specimen, then knowing its thickness relative to a thicker specimen would be sufficient to calculate the value of H in the regions too thick for WF microscopy. In this regard, the graph in Fig. 2(B) will be useful, showing the decreasing contribution to the background from planes that are very far from focus.

Naturally, this sort of extrapolation will be somewhat imprecise, but fortunately all that is required is an order-of-magnitude estimate of H. For specimens with H less than ∼20, WF microscopy with deconvolution can provide the best images. Over the range $20 < H < 200$, spinning disk confocals are the best choice. From $200 < H < 1000$, it will probably be necessary to use a spot scanning confocal, and beyond $H = 1000$, none of these methods is likely to be satisfactory.

Acknowledgements

The authors thank the academic and commercial faculty and students of the Analytical and Quantitative Light Microscopy Course at the Marine Biological Laboratory for helpful discussions, and Rainer Heintzmann, Ted Salmon and Greenfield Sluder for critical reading of the manuscript. For providing expertise and access to various microscope systems, we wish to acknowledge: Jennifer Atkins, Paul Goodwin and Kyla Teplitz, Applied Precision LLC; Brad Calloway and Geoff Daniels, Leica Microsystems Inc.; Louie Kerr, Marine Biological Laboratory; Bruce Gonzaga, Ed Rader and Penny Tavormina, Molecular Devices Corp.; Kunio Toshimitsu, Nikon Inc.; Melinda Frame and Kenji Matsubara, Olympus America Inc.; Yash Sabharwal, Optical Insights LLC; Greg Law, Perkin-Elmer Life and Analytical Sciences; Phong Tran, University of Pennsylvania; Paul McCormick, Jason Kirk, Rudi Rottenfusser, Julie Siefkert and Sebastian Tille, Carl Zeiss Inc. We thank Andrea Stout for initial analysis of some preliminary data. JRS is a Wellcome Trust Senior Research Fellow. This work was supported by NIH grant 2R01-AI-049301 to JMM.

References

Chong, F.K., Coates, C.G., Denvir, D.J., McHale, N., Thornbury, K. & Hollywood, M. (2004) Optimization of spinning disk confocal microscopy: synchronization with the ultra-sensitive EMCCD. *Three-Dimensional and Multidimensional Microscopy: Image Acquistion and Processing XI* (ed. by J.A. Conchello, C.J. Cogswell and T. Wilson), pp. 65–76. SPIE, Bellingham, WA.

Conchello, J.A. & Lichtman, J.W. (1994) Theoretical-analysis of a rotating-disk partially confocal scanning microscope. *Appl. Opt.* **33**, 585–596.

Giepmans, B.N., Adams, S.R., Ellisman, M.H. & Tsien, R.Y. (2006) The fluorescent toolbox for assessing protein location and function. *Science* **312**, 217–224.

Hu, K., Johnson, J., Florens, L., *et al.* (2006) Cytoskeletal components of an invasion machine-the apical complex of *Toxoplasma gondii*. *PLoS Pathog.* **2**, e13.

Murray, J.M. (1998) Evaluating the performance of fluorescence microscopes. *J. Microsc.* **191**, 128–134.

Patterson, G.H. & Piston, D.W. (2000) Photobleaching in two-photon excitation microscopy. *Biophys. J.* **78**, 2159–2162.

Sandison, D.R., Piston, D.W., Williams, R.M. & Webb, W.W. (1995a) Quantitative comparison of background rejection, signal-to-noise ratio, and resolution in confocal and full-field laser scanning microscopes. *Appl. Opt.* **34**, 3576–3588.

Sbalzarini, I.F. & Koumoutsakos, P. (2005) Feature point tracking and trajectory analysis for video imaging in cell biology. *J. Struct. Biol.* **151**, 182–195.

Shapiro, H.M. (2003) *Practical Flow Cytometry*. Wiley-Liss, New York.

Swedlow, J.R., Hu, K., Andrews, P.D., Roos, D.S. & Murray, J.M. (2002) Measuring tubulin content in *Toxoplasma gondii*: a comparison of laser-scanning confocal and wide-field fluorescence microscopy. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 2014–2019.

Tsien, R.Y. & Waggoner, A. (1995) Fluorophores for confocal microscopy: Photophysics and photochemistry. *Handbook of Biological Confocal Microscopy*, 2nd edn. (ed. by J.B. Pawley), pp. 267–279. Plenum Press, New York.

Wang, E., Babbey, C.M. & Dunn, K.W. (2005) Performance comparison between the high-speed Yokogawa spinning disc confocal system and single-point scanning confocal systems. *J. Microsc.* **218**, 148–159.

Sandison, D.R., Williams, R.M., Wells, K.S., Strickler, J. & Webb, W.W. (1995b). Quantitative fluorescence confocal laser scanning microscopy (CLSM). *Handbook of Biological Fluorescence Microscopy*, 2nd edn. (ed. by J.B. Pawley), pp. 39–53. Plenum Press, New York.