

A comparative study of high resolution microscopy imaging modalities using a three-dimensional resolution measure

Jerry Chao^{1,2}, Sripad Ram², E. Sally Ward² and Raimund J. Ober^{1,2,*}

¹*Department of Electrical Engineering, University of Texas at Dallas,
Richardson, TX 75080, USA*

²*Department of Immunology, University of Texas Southwestern Medical Center,
Dallas, TX 75390, USA*

*[*ober@utdallas.edu](mailto:ober@utdallas.edu)*

Abstract: From an acquired image, single molecule microscopy makes possible the determination of the distance separating two closely spaced biomolecules in three-dimensional (3D) space. Such distance information can be an important indicator of the nature of the biomolecular interaction. Distance determination, however, is especially difficult when, for example, the imaged point sources are very close to each other or are located near the focal plane of the imaging setup. In the context of such challenges, we compare the limits of the distance estimation accuracy for several high resolution 3D imaging modalities. The comparisons are made using a Cramer-Rao lower bound-based 3D resolution measure which predicts the best possible accuracy with which a given distance can be estimated. Modalities which separate the detection of individual point sources (e.g., using photoactivatable fluorophores) are shown to provide the best accuracy limits when the two point sources are very close to each other and/or are oriented near parallel to the optical axis. Meanwhile, modalities which implement the simultaneous imaging of the point sources from multiple focal planes perform best when given a near-focus point source pair. We also demonstrate that the maximum likelihood estimator is capable of attaining the limit of the accuracy predicted for each modality.

© 2009 Optical Society of America

OCIS codes: (180.6900) Three-dimensional microscopy; (180.2520) Fluorescence microscopy; (100.6640) Superresolution; (110.3055) Information theoretical analysis; (030.5290) Photon statistics; (000.5490) Probability theory, stochastic processes, and statistics

References and links

1. N. G. Walter, C. Huang, A. J. Manzo, and M. A. Sobhy, "Do-it-yourself guide: how to use the modern single-molecule toolkit," *Nat. Methods* **5**, 475–489 (2008).
2. W. E. Moerner, "New directions in single-molecule imaging and analysis," *Proc. Natl. Acad. Sci. USA* **104**, 12596–12602 (2007).
3. X. Qu, D. Wu, L. Mets, and N. F. Scherer, "Nanometer-localized multiple single-molecule fluorescence microscopy," *Proc. Natl. Acad. Sci. USA* **101**, 11298–11303 (2004).
4. M. P. Gordon, T. Ha, and P. R. Selvin, "Single-molecule high-resolution imaging with photobleaching," *Proc. Natl. Acad. Sci. USA* **101**, 6462–6465 (2004).
5. K. A. Lidke, B. Rieger, T. M. Jovin, and R. Heintzmann, "Superresolution by localization of quantum dots using blinking statistics," *Opt. Express* **13**, 7052–7062 (2005).
6. S. Ram, E. S. Ward, and R. J. Ober, "Beyond Rayleigh's criterion: A resolution measure with application to single-molecule microscopy," *Proc. Natl. Acad. Sci. USA* **103**, 4457–4462 (2006).

7. E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess, "Imaging intracellular fluorescent proteins at nanometer resolution," *Science* **313**, 1642–1645 (2006).
 8. S. T. Hess, T. P. K. Girirajan, and M. D. Mason, "Ultra-high resolution imaging by fluorescence photoactivation localization microscopy," *Biophys. J.* **91**, 4258–4272 (2006).
 9. M. Rust, M. Bates, and X. Zhuang, "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)," *Nat. Methods* **3**, 793–795 (2006).
 10. A. Sharonov and R. M. Hochstrasser, "Wide-field subdiffraction imaging by accumulated binding of diffusing probes," *Proc. Natl. Acad. Sci. USA* **103**, 18911–18916 (2006).
 11. H. P. Kao and A. S. Verkman, "Tracking of single fluorescent particles in three dimensions: use of cylindrical optics to encode particle position," *Biophys. J.* **67**, 1291–1300 (1994).
 12. L. Holtzer, T. Meckel, and T. Schmidt, "Nanometric three-dimensional tracking of individual quantum dots in cells," *Appl. Phys. Lett.* **90**, 053902 (2007).
 13. T. Sun and S. B. Andersson, "Precise 3-D localization of fluorescent probes without numerical fitting," in *Proceedings of the International Conference of IEEE Engineering in Medicine and Biology Society* (IEEE, 2007) pp. 4181–4184.
 14. S. Ram, P. Prabhat, J. Chao, E. S. Ward, and R. J. Ober, "High accuracy 3D quantum dot tracking with multifocal plane microscopy for the study of fast intracellular dynamics in live cells," *Biophys. J.* **95**, 6025–6043 (2008).
 15. S. R. P. Pavani and R. Piestun, "Three dimensional tracking of fluorescent microparticles using a photon-limited double-helix response system," *Opt. Express* **16**, 22048–22057 (2008).
 16. S. Weiss, "Fluorescence spectroscopy of single biomolecules," *Science* **283**, 1676–1683 (1999).
 17. M. Born and E. Wolf, *Principles of Optics* (Cambridge University Press, Cambridge, UK, 1999).
 18. M. Pluta, *Advanced light microscopy, vol. 1: principles and basic properties* (Elsevier, Amsterdam, 1988).
 19. S. Inoué, "Foundations of confocal scanned imaging in light microscopy," in "Handbook of Biological Confocal Microscopy," J. B. Pawley, ed. (Springer Science+Business Media, LLC, New York, 2006), 3rd ed.
 20. C. R. Rao, *Linear statistical inference and its applications* (Wiley, New York, USA, 1965).
 21. J. Chao, S. Ram, A. V. Abraham, E. S. Ward, and R. J. Ober, "A resolution measure for three-dimensional microscopy," *Opt. Commun.* **282**, 1751–1761 (2009).
 22. T. D. Lacoste, X. Michalet, F. Pinaud, D. S. Chemla, A. P. Alivisatos, and S. Weiss, "Ultrahigh-resolution multi-color colocalization of single fluorescent probes," *Proc. Natl. Acad. Sci. USA* **97**, 9461–9466 (2000).
 23. A. Agrawal, R. Deo, G. D. Wang, M. D. Wang, and S. Nie, "Nanometer-scale mapping and single-molecule detection with color-coded nanoparticle probes," *Proc. Natl. Acad. Sci. USA* **105**, 3298–3303 (2008).
 24. S. Ram, E. S. Ward, and R. J. Ober, "How accurately can a single molecule be localized in three dimensions using a fluorescence microscope?" *Proc. SPIE* **5699**, 426–435 (2005).
 25. P. Prabhat, S. Ram, E. S. Ward, and R. J. Ober, "Simultaneous imaging of different focal planes in fluorescence microscopy for the study of cellular dynamics in three dimensions," *IEEE Trans. Nanobiosci.* **3**, 237–242 (2004).
 26. P. Prabhat, Z. Gan, J. Chao, S. Ram, C. Vaccaro, S. Gibbons, R. J. Ober, and E. S. Ward, "Elucidation of intracellular recycling pathways leading to exocytosis of the Fc receptor, FcRn, by using multifocal plane microscopy," *Proc. Natl. Acad. Sci. USA* **104**, 5889–5894 (2007).
 27. S. Ram, J. Chao, P. Prabhat, E. S. Ward, and R. J. Ober, "A novel approach to determining the three-dimensional location of microscopic objects with applications to 3D particle tracking," *Proc. SPIE* **6443**, 64430D1–D7 (2007).
 28. M. J. Młodzianowski, M. F. Juetten, G. L. Beane, and J. Bewersdorf, "Experimental characterization of 3D localization techniques for particle-tracking and super-resolution microscopy," *Opt. Express* **17**, 8264–8277 (2009).
 29. M. F. Juetten, T. J. Gould, M. D. Lessard, M. J. Młodzianowski, B. Nagpure, B. T. Bennett, S. T. Hess, and J. Bewersdorf, "Three-dimensional sub-100 nm resolution fluorescence microscopy of thick samples," *Nat. Methods* **5**, 527–529 (2008).
 30. S. Ram, E. S. Ward, and R. J. Ober, "A stochastic analysis of performance limits for optical microscopes," *Multidimens. Syst. Sig. Process.* **17**, 27–57 (2006).
 31. D. L. Snyder and M. I. Miller, *Random point processes in time and space* (Springer Verlag, New York, USA, 1991), 2nd ed.
 32. "EstimationTool," <http://www4.utsouthwestern.edu/wardlab/estimationtool>.
 33. "FandPLimitTool," <http://www4.utsouthwestern.edu/wardlab/fandplimittool>.
 34. S. Zacks, *The theory of statistical inference* (John Wiley and Sons, New York, USA, 1971).
 35. S. F. Gibson and F. Lanni, "Experimental test of an analytical model of aberration in an oil-immersion objective lens used in three-dimensional light microscopy," *J. Opt. Soc. Am. A* **9**, 154–166 (1992).
 36. P. Török and P. Varga, "Electromagnetic diffraction of light focused through a stratified medium," *Appl. Opt.* **36**, 2305–2312 (1997).
 37. O. Haeberlé, "Focusing of light through a stratified medium: a practical approach for computing microscope point spread functions. Part I: Conventional microscopy," *Opt. Commun.* **216**, 55–63 (2003).
 38. L. Tao and C. Nicholson, "The three-dimensional point spread functions of a microscope objective in image and object space," *J. Microsc.* **178**, 267–271 (1995).
-

1. Introduction

Single molecule microscopy (e.g., [1, 2]) has become an important tool for studying biological processes at the level of individual biomolecules. Besides enabling the direct visualization of biological events at the biomolecular level, the acquired images contain extractable information that can provide a more detailed understanding of the recorded events. In localization-based superresolution (e.g., [3, 4, 5, 6, 7, 8, 9, 10]) and high accuracy three-dimensional (3D) tracking (e.g., [11, 12, 13, 14, 15]), for example, the position of a single biomolecule is accurately determined from its image. Besides the localization of an individual molecule, however, of particular importance is the study of the interaction between two biomolecules.

In studying an interaction such as that between two proteins, an important problem is to resolve the two closely spaced molecules in the sense of determining the distance that separates them. An accurately determined distance of separation can provide biologically significant insight as it helps to elucidate the nature of the biomolecular interaction. Techniques based on fluorescence resonance energy transfer (e.g., [16]) have been limited to the probing of interactions involving distances of less than 10 nm. Meanwhile, based on Rayleigh's criterion, it has been widely believed that the conventional optical microscope can resolve two in-focus molecules only if they are separated by more than 200 nm. In the more general 3D scenario where the molecules are out of focus, the classical 3D resolution limit [17, 18, 19] predicts an even higher threshold distance. This leaves a distance range of 10 nm to 200 nm that is unaccounted for, and represents a significant obstacle to the study of many biomolecular interactions.

In [6], we proposed a new resolution measure that is based on the Cramer-Rao lower bound [20] from estimation theory, and it specifies the limit of the accuracy (i.e., the smallest possible standard deviation) with which a given distance can be determined. Using this result, we showed that, contrary to common belief, distances below Rayleigh's criterion can in fact be accurately determined with conventional optical microscopy. More precisely, this resolution measure predicts that arbitrarily small distances of separation can be estimated with prespecified accuracy, provided that a sufficient number of photons are detected from the pair of molecules. Since in general biomolecular interactions are not confined to a plane of focus, the resolution measure has been extended to the 3D context [21].

In [6] and [21], it was shown that even though distances under 200 nm in two-dimensional (2D) or 3D space can be estimated, the number of photons from the pair of molecules that is required to be able to obtain an acceptable accuracy increases substantially in nonlinear fashion with decreasing distance. Since attaining the required photon count is not always feasible (e.g., due to the photobleaching of the fluorophore), it would be useful to explore other means of achieving the desired limit of the accuracy. One such strategy is simply to separate the detection of the two closely spaced molecules, as the difficulty in determining small distances is attributable to the fact that the acquired image consists of two significantly overlapping spots (i.e., point spread functions). Separate detection can be achieved spectrally if the two fluorophores emit photons of different wavelengths (e.g., [22, 23]), or it can be achieved temporally. The latter can be realized through the natural photobleaching [3, 4, 6] or blinking [5] of the fluorophores, or the stochastic activation of photoactivatable [7, 8], photoswitchable [9], or other types [10] of fluorophores.

In dealing with the more general 3D scenario of resolving two out-of-focus molecules, the poor depth discrimination capability of the optical microscope introduces an additional challenge in the distance estimation problem. In particular, depth discrimination is especially poor near the microscope's focal plane, and it has been shown that the closer a single molecule is to the focal plane, the more difficult it is to accurately determine its axial (z) position [24]. This z -localization problem has direct implications on the distance problem. In [21], it was shown that even for two molecules that are separated by a relatively large distance, the limit of the dis-

tance estimation accuracy (i.e., the 3D resolution measure) typically deteriorates severely when either of the molecules is axially located near the microscope’s plane of focus. The resolution measure can again be improved by collecting more photons from the molecules, but very large numbers will be required, as in the case of small distances, to obtain an acceptable limit of the accuracy. This near-focus problem therefore represents a significant obstacle in the study of biomolecular interactions that occur in proximity to the focal plane.

A technique that can be expected to overcome poor depth discrimination near focus for the 3D distance estimation problem, however, is multifocal plane microscopy (MUM) [25]. MUM is an imaging configuration that uses multiple cameras to simultaneously capture images from distinct focal planes within the specimen. By positioning multiple focal planes at appropriate spacings along the microscope’s optical (z -)axis, MUM can be used to visualize and study cellular processes over a large depth range (e.g., [14, 26]). Importantly, by imaging a near-focus single molecule simultaneously from additional focal planes that are relatively far from it, MUM can significantly improve the limit of the accuracy with which the molecule’s z position can be determined [14, 27]. By the same principle, one can expect MUM to have an analogous effect on the limit of the distance estimation accuracy.

The main focus of this paper is to systematically compare different microscopy imaging modalities in terms of their ability to estimate the 3D distance of separation. These modalities make use of separate detection and/or the MUM technique, and represent imaging setups that have all been implemented and used in practice. Therefore, the method as well as the results of our comparisons can be used to generate or provide important guidelines in the design of an imaging experiment where distance estimation is an integral part of the data analysis. Whereas different imaging modalities have been compared in terms of their ability to localize a single point source (e.g., [28]), the work presented here provides a comparison of modalities in terms of their ability to resolve a pair of closely spaced point sources (i.e., to determine the distance of separation).

The comparisons are made using customizations of the 3D resolution measure [21] which we derive for each modality. Moreover, they are performed in the context of the important challenges posed by small distances of separation and near-focus depth discrimination. A third challenge that will be considered is that associated with the orientation of a point source pair with respect to the microscope’s z -axis. As we showed in [21], even a relatively large distance of separation can be difficult to determine with acceptable accuracy if, for example, one point source is positioned in front of, and hence obscures the other, in the z direction.

In addition to the comparisons, we show that the limits of the distance estimation accuracy predicted for the various imaging modalities by their respective resolution measures are attainable by the maximum likelihood estimator. This is an important result as it is useful to identify estimators that can achieve the limit of the accuracy in practice. We demonstrate the maximum likelihood estimator to be one such estimator via distance estimations that were carried out on simulated images of point source pairs.

The remainder of this paper is organized as follows. In Section 2, we provide a description of the four imaging modalities that are compared and the image data that they produce. In Section 3, we derive the 3D resolution measures corresponding to the different imaging modalities. In doing so, we demonstrate how the theory behind the resolution measure can be applied to different modalities. In Section 4, we use the derived resolution measures to compare the modalities in terms of their limits of the distance estimation accuracy. Specifically, we illustrate the dependence of the different resolution measures on the distance of separation, the axial position, and the orientation of a point source pair, and show that the best-performing modality depends on the specific scenario. In Section 5, we present the results of our maximum likelihood distance estimations with simulated images. Finally, we conclude our presentation in Section 6.

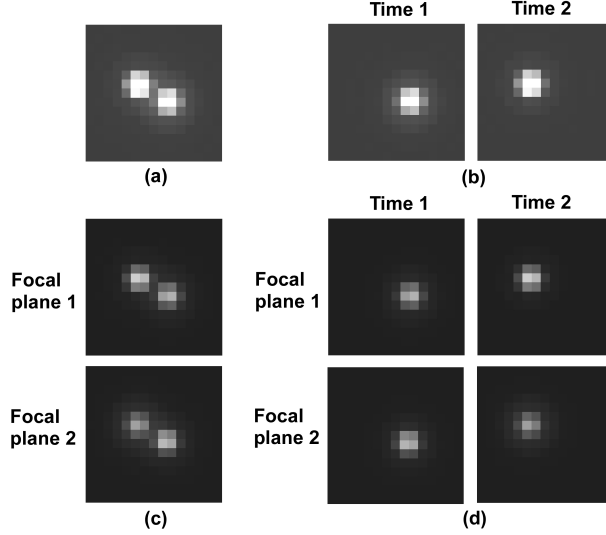


Fig. 1. Simulated images of a pair of like point sources in 3D space as acquired by (a) the SIM-SNG, (b) the SEP-SNG, (c) the SIM-MUM, and (d) the SEP-MUM imaging modalities. The point sources are deliberately simulated to be separated by a relatively large distance of 500 nm to clearly show the presence of a pair. The exposure time of each image is the same regardless of the modality. (a) The SIM-SNG modality abstracts the conventional fluorescence imaging setup which produces a single image of both point sources. (b) The SEP-SNG modality detects the two point sources separately in time and hence produces two images each capturing only one of the point sources. (c) The SIM-MUM modality uses two cameras to simultaneously detect the point source pair from two distinct focal planes, and hence produces two images each capturing both point sources. (d) The SEP-MUM modality combines separate detection with two-plane imaging and therefore produces a total of four images - two of one point source from the two distinct focal planes, and two of the other point source at a different time, but from the same two focal planes. In (c) and (d), the images are dimmer because the collected photons are split between the two focal planes.

2. The four imaging modalities

In this section, we give a description of each of the four imaging modalities for which we will compare the limits of the distance estimation accuracy. These modalities differ in terms of whether they employ the simultaneous or the separate detection of the two closely spaced point sources, and whether they image those point sources from a single focal plane or from multiple focal planes at the same time. The four modalities comprise all possible combinations of simultaneous (SIM) or separate (SEP) detection coupled with single focal plane (SNG) or multifocal plane (MUM) imaging. For brevity, each modality will be denoted by the appropriate combination of abbreviations here and throughout the paper. In particular, we describe the nature of the images that are produced by each modality, and from which distances of separation are to be estimated. The presentation is aided by Fig. 1, which shows simulated images of a pair of like point sources as acquired by each modality.

1. The **SIM-SNG (simultaneous detection, single focal plane)** modality represents the conventional fluorescence imaging setup where a single image is acquired of a pair of point sources and is used to estimate the distance of separation. A simulated image for this modality is shown in Fig. 1(a), where a relatively large 500 nm distance of separation

was specified to clearly illustrate the presence of two point sources. For much smaller distances of separation, the two spots would overlap significantly and be difficult to visually distinguish as two.

2. The **SEP-SNG (separate detection, single focal plane) modality** separates the detection of two closely spaced point sources either spectrally or temporally, and images each point source from one and the same focal plane. For ease of presentation, however, and without loss of generality, we will assume for this modality the model of temporal separation that relies on the use of photoactivatable or photoswitchable fluorophores (e.g., [7, 8, 9]). In this model, stochastically different subsets of the entire fluorophore population are converted to the fluorescing state in successive rounds of photoactivation. By keeping the subsets of photoactivated fluorophores small, it is likely that two closely spaced fluorophores are individually detected at different times. The output of the SEP-SNG modality therefore consists of two images acquired during disjoint time intervals as shown in Fig. 1(b). Each image captures only one of the point sources, but the two images are used together to estimate the distance that separates the two point sources.
3. The **SIM-MUM (simultaneous detection, multifocal plane) modality** (e.g., [14, 26]) detects two closely spaced point sources simultaneously, and images them from multiple focal planes at the same time. Though it in principle includes MUM setups that employ any number of focal planes, here we will assume for this modality a two-plane MUM configuration. In a two-plane MUM setup, a beam splitter is used to divide the fluorescence collected by the objective lens between two cameras that are positioned at distinct distances from the tube lens of the microscope. In this way, the two cameras can image distinct focal planes within the specimen at the same time. The output of this modality therefore comprises two images that are acquired during the same time interval, one by each camera. Each image captures both point sources as shown in Fig. 1(c), but from a distinct focal plane. The two images are used together to estimate the distance of separation. Note that the images of Fig. 1(c) are dimmer than those of Figs. 1(a) and 1(b) due to the splitting of the collected fluorescence between the two cameras.
4. The **SEP-MUM (separate detection, multifocal plane) modality** (e.g., [29]) combines separate detection with the MUM technique, and generally encompasses all methods of separation coupled with a MUM configuration employing any number of focal planes. In keeping with the assumptions made with the SEP-SNG and the SIM-MUM modalities, however, we will assume for this modality the use of photoactivatable or photoswitchable fluorophores with a two-plane MUM configuration. The output of the SEP-MUM modality therefore consists of four images as shown in Fig. 1(d). Two of the four images capture one and the same point source during the same time interval, but each is acquired by a different camera from a distinct focal plane. The remaining two images both capture the other point source during a time interval that is disjoint from the other, but are again acquired by the two different cameras from their respective focal planes. The four images are used together to perform the estimation of the distance of separation. Note that as in the case of the SIM-MUM modality, the images of Fig. 1(d) are dimmer because of the splitting of the collected fluorescence between the two cameras.

3. The 3D resolution measures for the four imaging modalities

The 3D resolution measure is a quantity that specifies the limit of the accuracy (i.e., the smallest possible standard deviation) that can be expected of the estimates produced by any unbiased estimator of the distance that separates two objects (e.g., point sources). More precisely, it is

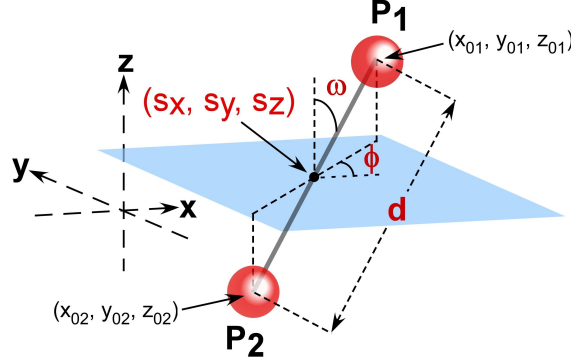


Fig. 2. Two point sources P_1 and P_2 , located respectively at (x_{01}, y_{01}, z_{01}) and (x_{02}, y_{02}, z_{02}) , in 3D space. In formulating the distance estimation problem, the 3D geometry of the pair is equivalently described by the six parameters $d, \phi, \omega, s_x, s_y, s_z$. The parameter d is the distance that separates P_1 and P_2 . The coordinates s_x, s_y , and s_z specify the location of the midpoint between P_1 and P_2 . The parameter ω is the angle between the line segment P_1P_2 and the positive z -axis. The parameter ϕ is the angle between the xy -plane projection of P_1P_2 and the positive x -axis.

defined as the square root of the Cramer-Rao lower bound [20] for estimating a distance of separation. The mathematical foundation on which the 3D resolution measure is based was presented in [30], and a detailed description of its derivation was given in [21].

The resolution measure as described in [21] assumes distance estimation based on a single image of a pair of point sources. As such, it is directly applicable to a conventional fluorescence imaging setup (i.e., the SIM-SNG imaging modality). However, to obtain the limits of the distance estimation accuracy for the SEP-SNG, the SIM-MUM, and the SEP-MUM modalities, the resolution measure needs to be adapted to consider multiple images of one or both point sources from different acquisition time intervals and/or focal planes (Figs. 1(b), 1(c), 1(d)). Since the derivations of these resolution measures are most easily presented as modifications to the derivation of the resolution measure for the SIM-SNG modality, we will begin by providing a succinct derivation of the SIM-SNG resolution measure that includes the necessary ingredients for arriving at the other resolution measures.

Note that only relatively high level descriptions will be given in this section, and that detailed descriptions with accompanying mathematical formulae can be found in the appendix.

3.1. The SIM-SNG modality

To obtain the resolution measure for the SIM-SNG modality, we consider the problem of estimating the vector of unknown parameters $\theta = (d, \phi, \omega, s_x, s_y, s_z)$, $\theta \in \Theta$, from a single image of two point sources (Fig. 1(a)). The symbol Θ denotes the parameter space which is an open subset of \mathbb{R}^6 . As shown in Fig. 2, the six parameters of θ collectively describe the geometry of a pair of point sources P_1 and P_2 in 3D space. Parameter d denotes the distance that separates P_1 and P_2 , parameter ϕ denotes the angle which the xy -plane projection of the line segment P_1P_2 forms with the positive x -axis, parameter ω denotes the angle which P_1P_2 forms with the positive z -axis, and parameters s_x, s_y , and s_z denote the coordinates of the midpoint between P_1 and P_2 .

In order to estimate θ (which includes the distance of separation d) from an image of the point sources P_1 and P_2 , a mathematical model that accurately describes the image is required. To account for the intrinsically stochastic nature of the photon emission (and hence the photon

detection) process, this image, which we assume to be acquired during the time interval $[t_0, t]$, is modeled as a spatio-temporal random process [30, 31]. The temporal portion models the time points at which photons are detected from P_1 and P_2 as an inhomogeneous Poisson process with intensity function $\Lambda_\theta(\tau)$, $\tau \geq t_0$. The spatial portion models the image space coordinates at which the photons are detected as a sequence of independent random variables with probability density functions $\{f_{\theta,\tau}\}_{\tau \geq t_0}$. As detailed in appendix section A.1, the intensity function $\Lambda_\theta(\tau)$ is simply the sum of the rates at which photons are detected from P_1 and P_2 (Eq. (3)), and each density function $f_{\theta,\tau}$ is simply a weighted sum of the images (i.e., the point spread functions) of the two point sources (Eq. (4)). Note that in general, both $\Lambda_\theta(\tau)$ and $\{f_{\theta,\tau}\}_{\tau \geq t_0}$ are functions of θ (i.e., the parameters to be estimated).

Having modeled our image through the functions $\Lambda_\theta(\tau)$ and $\{f_{\theta,\tau}\}_{\tau \geq t_0}$, a matrix quantity known as the Fisher information matrix [20] can be calculated. Denoted by $\mathbf{I}(\theta)$, this matrix provides a measure of the amount of information the image carries about the parameter vector θ to be estimated. Since θ is a six-parameter vector, $\mathbf{I}(\theta)$ is a 6-by-6 matrix.

The Fisher information matrix in turn allows us to apply the Cramer-Rao inequality [20], a well-known result from estimation theory which states that the inverse of $\mathbf{I}(\theta)$ is a lower bound on the covariance matrix of any unbiased estimator $\hat{\theta}$ of the parameter vector θ (Eq. (8)). It then follows that element $(1, 1)$ of $\mathbf{I}^{-1}(\theta)$, which corresponds to the distance parameter d (i.e., the first parameter in θ), is a lower bound on the variance of the distance estimates of any unbiased estimator. The 3D resolution measure for the SIM-SNG modality is accordingly defined as the quantity $\sqrt{[\mathbf{I}^{-1}(\theta)]_{11}}$, where the square root makes it a lower bound on the standard deviation, rather than the variance, of the distance estimates. Since it is a lower bound on the standard deviation, small values of the resolution measure correspond to good distance estimation accuracy, whereas large values correspond to poor accuracy.

We note that in the modeling of our image here, we have omitted details regarding image pixelation and extraneous noise sources (e.g., sample autofluorescence, detector readout). In the derivation of appendix section A.1, these important details which pertain to a practical image are taken into account.

3.2. The SEP-SNG, the SIM-MUM, and the SEP-MUM modalities

The SEP-SNG, the SIM-MUM, and the SEP-MUM modalities each produces multiple images as shown in Figs. 1(b), 1(c), and 1(d). To obtain the 3D resolution measure for any of these modalities, a single Fisher information matrix needs to be derived that combines the information carried by each of its images about the unknown parameter vector θ . Then, by applying the definition given in Section 3.1, the resolution measure follows easily as the square root of the first main diagonal element of the inverse matrix.

To derive the Fisher information matrix for any of these modalities, we make use of the assumption that temporally separated images, and also images from different focal planes, are formed independently of one another. By virtue of this independence, the Fisher information matrix for a modality can be obtained by simply adding the Fisher information matrices corresponding to the images it produces. For a given modality, the key to our derivation therefore lies in the determination of the matrix for each of its images. These matrices will be different from that for the SIM-SNG modality because the images to which they correspond contain one point source instead of two, and/or are acquired from a different focal plane. These image differences will necessitate a respecification of the spatio-temporal random process that describes the image formed from the photons detected from the point sources. This respecification can be achieved by simple redefinitions of the intensity function $\Lambda_\theta(\tau)$ and the probability density functions $\{f_{\theta,\tau}\}_{\tau \geq t_0}$ of, respectively, the temporal and spatial portions of the random process.

More specifically, for an image produced by the SEP-SNG or the SEP-MUM modality where

only a single point source is detected, the intensity function $\Lambda_\theta(\tau)$ will simply be the photon detection rate of only point source P_1 or P_2 , and not a sum of their detection rates. Likewise, each spatial density function $f_{\theta,\tau}$ will entail only the point spread function of P_1 or P_2 , and not a sum of their point spread functions. For images acquired from different focal planes by the SIM-MUM or the SEP-MUM modality, the spatial density functions $\{f_{\theta,\tau}\}_{\tau \geq t_0}$ need to be defined such that the axial position (i.e., defocus) of the point source pair is specified with respect to the focal plane from which the image was acquired. In addition, since the different focal planes are associated with different lateral magnifications, the density functions need to be defined using the appropriate magnification value.

Using these redefined intensity and density functions, the Fisher information matrices corresponding to the images produced by a modality can be calculated and subsequently added to yield a single matrix. Given this sum matrix $\mathbf{I}(\theta)$, the 3D resolution measure for the modality follows easily as the quantity $\sqrt{[\mathbf{I}^{-1}(\theta)]_{11}}$.

Detailed derivations of the 3D resolution measures for the SEP-SNG, the SIM-MUM, and the SEP-MUM modalities are given in appendix sections A.2, A.3, and A.4, respectively.

4. Comparison of modalities using the 3D resolution measure

In this section, we use the 3D resolution measures derived in Section 3 to compare the limits of the distance estimation accuracy for the four imaging modalities. To understand how the different modalities perform in the challenging scenarios of small distances of separation, near-focus depth discrimination, and near-parallel orientation with respect to the z -axis, we look at how their resolution measures depend on the parameters d , s_z , and ω (Fig. 2), respectively. More specifically, we plot the resolution measures as functions of two of the three parameters at a time, and we do so for all three possible pairings of the parameters.

To generate the plots for comparison, we assume the imaging of a pair of like point sources P_1 and P_2 that emit photons of wavelength $\lambda = 655$ nm. The image of each point source is assumed to be described by the classical 3D point spread function of Born and Wolf [17]. That is, the point spread functions of P_1 and P_2 are each of the form

$$q_{z_0}(x, y) = \frac{4\pi n_a^2}{\lambda^2} \left| \int_0^1 J_0 \left(\frac{2\pi n_a}{\lambda} \sqrt{x^2 + y^2} \rho \right) e^{j \frac{\pi n_a^2 z_0}{n \lambda} \rho^2} \rho d\rho \right|^2, \quad (x, y) \in \mathbb{R}^2, \quad z_0 \in \mathbb{R}, \quad (1)$$

where z_0 is the axial position of the point source in the object space, n_a is the numerical aperture of the objective lens, and n is the refractive index of the object space medium. For our comparisons, the objective lens is assumed to have a numerical aperture of $n_a = 1.4$ and a magnification of $M = 100$. The refractive index of the object space medium is assumed to be $n = 1.515$.

Each point source is further assumed to have a constant photon detection rate of 5000 photons per second, and the same acquisition time interval of 1 second is assumed for any given image regardless of the modality. In this way, an expected total of 10000 photons are detected from the two point sources in each modality, regardless of whether they are distributed across one, two, or four images (Fig. 1).

For details on settings and parameters not mentioned here, including those related to the two-plane MUM configuration of the SIM-MUM and the SEP-MUM modalities, see Fig. 3.

4.1. Small distances of separation

The smaller the distance of separation between two point sources, the greater the amount of overlap between their point spread functions in the acquired image. As a result, distance estimation can be expected to become increasingly difficult with decreasing distance. Since neither the SIM-SNG modality nor the SIM-MUM modality separates the detection of the point

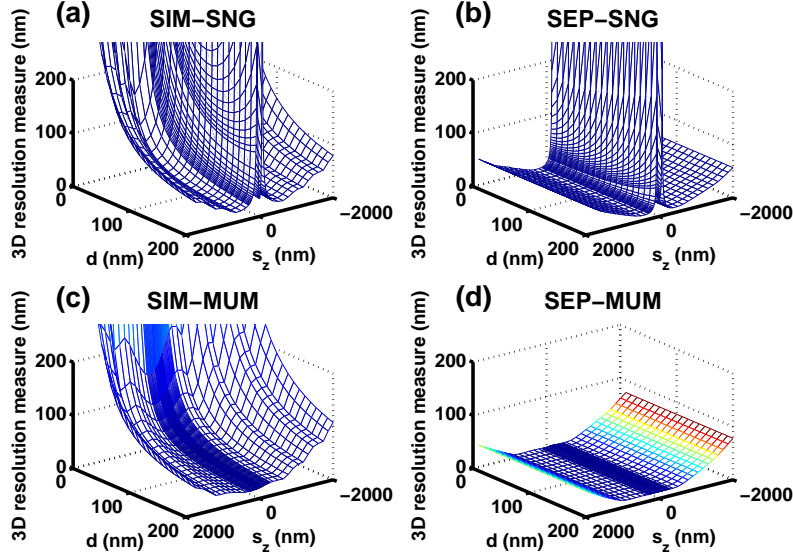


Fig. 3. Dependence of the 3D resolution measure (i.e., the best possible standard deviation for distance estimation) for (a) the SIM-SNG, (b) the SEP-SNG, (c) the SIM-MUM, and (d) the SEP-MUM modality on the distance of separation d and the axial position s_z of the imaged point source pair. In (a), (b), (c), and (d), the image of each point source in the pair is assumed to be described by the 3D point spread function of Born and Wolf. Each point source emits photons of wavelength $\lambda = 655$ nm, which are detected at a rate of $\Lambda_1 = \Lambda_2 = 5000$ photons per second. The image acquisition time is set to 1 second per image. The refractive index of the object space medium is set to $n = 1.515$, and the numerical aperture and the magnification of the objective lens are set to $n_a = 1.4$ and $M = 100$, respectively. A single image consists of a 21×21 array of $13 \mu\text{m}$ by $13 \mu\text{m}$ pixels, and the position of the point source pair in the xy -plane is set to $s_x = s_y = 1365$ nm, such that its image is centered on the pixel array given the 100-fold magnification. The point source pair is oriented such that it forms a 45° angle with the positive z -axis ($\omega = 45^\circ$) and projects at a 60° angle from the positive x -axis in the xy -plane ($\phi = 60^\circ$). In (c) and (d), the spacing between the two focal planes is set to $\Delta z_f = 500$ nm. The first focal plane corresponds to the focal plane of the modalities in (a) and (b), and is located at $s_z = 0$ nm with an associated magnification of 100. The second focal plane is located above the first at $s_z = 500$ nm, and its associated magnification is set to $M' = 97.98$ (computed using Eq. (12) by assuming a standard tube length of $L = 160$ mm). The collected photons are assumed to be split 50:50 between the two focal planes. In (a) and (b), the mean of the background noise in each image is assumed to be a constant and is set to $\beta(k) = 80$ photons per pixel. In (c) and (d), however, it is set to $\beta(k) = 40$ photons per pixel due to the equal splitting of the collected fluorescence. In (a), (b), (c), and (d), the mean and the standard deviation of the readout noise in each image are respectively set to $\eta_k = 0 e^-$ and $\sigma_k = 8 e^-$ per pixel.

sources, they can be expected to perform particularly poorly at very small distances of separation. This is seen in Figs. 3(a) and 3(c), where the 3D resolution measures for the SIM-SNG and the SIM-MUM modalities are respectively shown as functions of a point source pair's distance of separation d and axial position s_z . For a given axial position, a deterioration of the limit of the distance estimation accuracy (i.e., an increase in the resolution measure) is observed for both modalities as the distance of separation is decreased from 200 nm to 0 nm. This deterioration is nonlinear, progressing relatively slowly at larger values of d , but becoming significantly sharper when the value of d is in the low tens of nanometers. (Note that at $d = 0$ nm, the resolution measure is infinity due to the Fisher information matrix becoming singular. This special scenario corresponds to the degenerate case where the images of the point sources completely coincide. Also, note that for some s_z values close to 0 nm in Fig. 3(a) where the resolution measure is very large, the described pattern of deterioration does not hold. Specifically, the pattern is interrupted by a sharp deterioration of the resolution measure over values of d that put one of the point sources near the focal plane at $s_z = 0$ nm. This problem of especially poor depth discrimination near focus is discussed in Section 4.2.)

In contrast, the deteriorative effect of small distances of separation is effectively neutralized in the SEP-SNG and the SEP-MUM modalities. By separating the detection of the two point sources, these two modalities completely remove the overlap of their point spread functions. Therefore, good limits of the distance estimation accuracy can be expected even when the two point sources are very close to each other. This is corroborated by Figs. 3(b) and 3(d), where the resolution measures corresponding to the SEP-SNG and the SEP-MUM modalities are respectively plotted as functions of d and s_z . For a given value of s_z , the resolution measures for both modalities stay relatively small and constant regardless of the distance of separation. (Note the exception at $d = 0$ nm, where the resolution measure is infinity. In addition, the exception to the rule is again present for some near-focus s_z values in Fig. 3(b) where the resolution measure is very large.)

To give an example of how the four modalities compare in terms of precise numbers, Fig. 4(a) shows in one plot the 2D slice $s_z = 367.2$ nm from each of the four 3D plots of Fig. 3. At this axial position, a distance of $d = 200$ nm can be estimated with best possible accuracies of ± 22.95 nm and ± 18.51 nm when the SIM-SNG and the SIM-MUM modalities, respectively, are used. These numbers correspond to approximately $\pm 10\%$ of the 200 nm distance. When the SEP-SNG and the SEP-MUM modalities are used, however, the best possible accuracies that can be expected are ± 9.92 nm and ± 12.70 nm, respectively, or approximately just $\pm 5\%$ of the 200 nm distance. Though a nontrivial, factor of two improvement in the limit of the accuracy can already be seen at $d = 200$ nm when the SEP-SNG and the SEP-MUM modalities are used, the advantage of using separate detection is even more significant at smaller distances.

At $d = 100$ nm, for example, best possible accuracies of ± 45.34 nm and ± 33.26 nm can be expected from the SIM-SNG and the SIM-MUM modalities, respectively. These numbers are no better than $\pm 30\%$ of the 100 nm distance. When the SEP-SNG and the SEP-MUM modalities are used, however, the resolution measures are ± 9.58 nm and ± 12.60 nm, respectively, corresponding to just approximately $\pm 10\%$ of the 100 nm distance. If the distance is halved again to $d = 50$ nm, then the limits of the accuracy for the SIM-SNG and the SIM-MUM modalities deteriorate further to ± 88.48 nm and ± 64.60 nm, respectively. These numbers are greater than the 50 nm distance, and are clearly unacceptable. In sharp contrast, by using the SEP-SNG and the SEP-MUM modalities instead, resolution measures of ± 9.47 nm and ± 12.54 nm, respectively, can still be expected. These limits of the accuracy correspond to a perhaps still acceptable $\pm 20\%$ and $\pm 25\%$ of the 50 nm distance, respectively.

We note that the numbers given here pertain to the scenario where an expected photon count of 5000 is detected from each of the two point sources. For the SIM-SNG modality, we showed

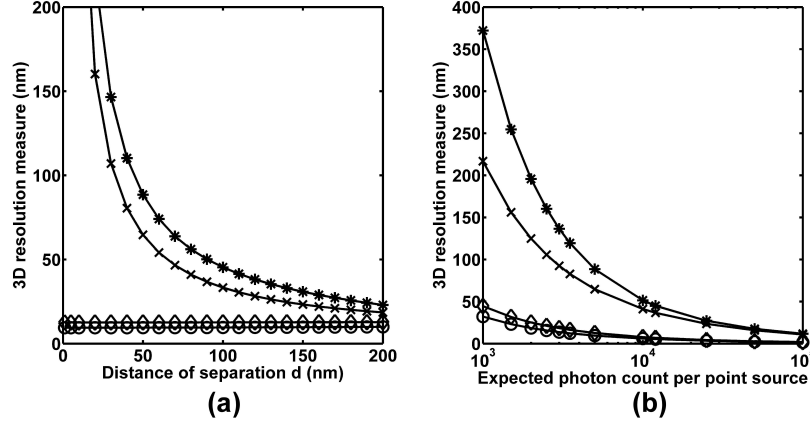


Fig. 4. Dependence of the 3D resolution measures for the SIM-SNG (*), the SEP-SNG (\circ), the SIM-MUM (\times), and the SEP-MUM (\diamond) modalities on (a) the distance of separation d of, and (b) the expected photon count detected from, the imaged point source pair. In (a), the curves shown correspond to the 2D slice $s_z = 367.2$ nm from each of the four 3D plots of Fig. 3. In (b), the resolution measures shown pertain to the point source pair with a distance of separation of $d = 50$ nm from (a). The expected photon count per point source is varied from 1000 to 100000 photons. In both (a) and (b), all experimental, noise, and other point source pair parameters are as given in Fig. 3.

in [21] that the resolution measure can be improved by detecting more photons from the point source pair and hence increasing the amount of data that is collected. To demonstrate this photon count dependence for all modalities considered here, Fig. 4(b) shows, for each modality, the resolution measure corresponding to the point source pair with separation distance $d = 50$ nm from Fig. 4(a) as a function of the expected number of detected photons. For each modality, the improvement (i.e., decrease) in the resolution measure can roughly be described by an inverse square root dependence on the increase in photon count. By doubling the expected photon count to 10000 per point source, for example, the best possible accuracies with which the 50 nm distance can be estimated are improved to ± 51.37 nm and ± 41.01 nm for the SIM-SNG and the SIM-MUM modalities, respectively, and ± 5.96 nm and ± 7.72 nm for the SEP-SNG and the SEP-MUM modalities, respectively. A further tenfold increase to 100000 photons per point source would improve the resolution measures to less than $\pm 24\%$ of the 50 nm distance for the SIM-SNG (± 11.68 nm) and the SIM-MUM (± 10.73 nm) modalities, and to less than $\pm 4\%$ for the SEP-SNG (± 1.51 nm) and the SEP-MUM (± 1.81 nm) modalities.

In closing this subsection, we point out that the same general observations concerning the effect of small distances of separation can be made with the plots of Fig. 5, where the resolution measures corresponding to the four modalities are shown as functions of d and the orientation angle ω . On the one hand, for a given value of ω , the nonlinear deterioration of the resolution measure with decreasing distance of separation is observed for the SIM-SNG (Fig. 5(a)) and the SIM-MUM (Fig. 5(c)) modalities. On the other hand, for a given angle, the resolution measure remains low and essentially constant over all distances (except $d = 0$ nm) for the SEP-SNG (Fig. 5(b)) and the SEP-MUM (Fig. 5(d)) modalities. (Note that there is an exception at $\omega = 0^\circ$, where the resolution measure is infinity for all modalities regardless of the distance d . This is due to the Fisher information matrix being singular when one point source is situated exactly in front of the other.)

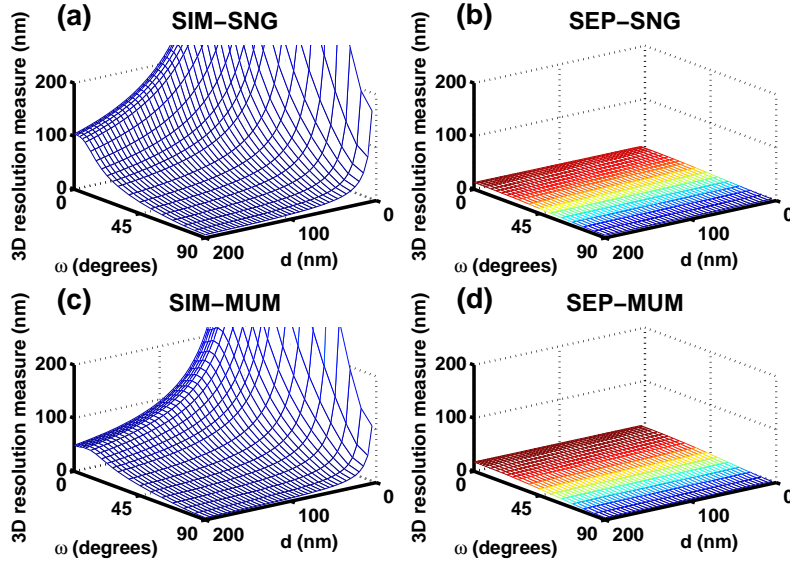


Fig. 5. Dependence of the 3D resolution measure for (a) the SIM-SNG, (b) the SEP-SNG, (c) the SIM-MUM, and (d) the SEP-MUM modality on the distance of separation d and the orientation angle ω of the imaged point source pair. In (a), (b), (c), and (d), the point source pair is axially centered at $s_z = 400$ nm. Except for d and ω which are varied in these plots, all experimental, noise, and other point source pair parameters are as given in Fig. 3.

4.2. Near-focus depth discrimination

Due to the poor depth discrimination capability of the optical microscope, the accuracy with which the z position of a point source can be determined is especially poor when it is located near the focal plane [24]. Since the distance estimation problem can be formulated equivalently as the simultaneous estimation of the locations (x_{01}, y_{01}, z_{01}) and (x_{02}, y_{02}, z_{02}) (see Fig. 2) of the two point sources, it follows that especially poor accuracy for determining the distance of separation can also be expected when either of the point sources is near-focus. In more technical terms, the unknown parameter vector θ can alternatively be defined as $(x_{01}, y_{01}, z_{01}, x_{02}, y_{02}, z_{02})$, and estimates of the distance of separation d can be obtained indirectly by simply computing the Euclidean distance using the estimated coordinates. Therefore, a poor accuracy in the estimation of z_{01} or z_{02} can be expected to translate to a poor accuracy in the determination of d .

The severe deterioration of the limit of the distance estimation accuracy when a near-focus point source is involved has previously been reported in [21] for the SIM-SNG modality. Here, this phenomenon can be seen in Fig. 3(a), for any distance of separation d , as the two sharp increases of the resolution measure near the focal plane $s_z = 0$ nm. (Note that the presence of two sharp increases can be seen more clearly in the 2D plot of Fig. 6(b). Additionally, we note the exception at $d = 0$ nm where the resolution measure is infinity regardless of the axial position.) The sharp deterioration below the focal plane corresponds to point source P_1 coming very close to being in focus (i.e., value of z_{01} approaching 0 nm), and the sharp deterioration above the focal plane corresponds to point source P_2 coming very close to being in focus (i.e., value of z_{02} approaching 0 nm).

Similar sharp deteriorations of the resolution measure about the focal plane can be seen for

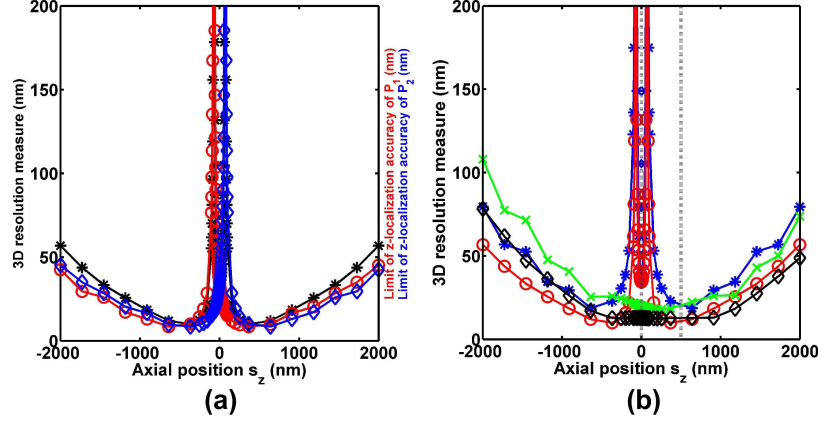


Fig. 6. (a) Correlation of the 3D resolution measure (i.e., limit of the distance estimation accuracy) (*) for a point source pair with the limit of the z -localization accuracy of each of the point sources P_1 (◊) and P_2 (◊) in the pair. The resolution measure curve corresponds to the 2D slice $d = 200$ nm from Fig. 3(b) (i.e., the SEP-SNG modality). (b) Dependence of the 3D resolution measures for the SIM-SNG (*), the SEP-SNG (◊), the SIM-MUM (×), and the SEP-MUM (◊) modalities on the axial position s_z of the imaged point source pair. The curves shown correspond to the 2D slice $d = 200$ nm from each of the four 3D plots of Fig. 3. The dashed vertical line at $s_z = 0$ nm marks the focal plane of the SIM-SNG and the SEP-SNG modalities, and focal plane 1 of the SIM-MUM and the SEP-MUM modalities. The dashed vertical line at $s_z = 500$ nm marks focal plane 2 of the SIM-MUM and the SEP-MUM modalities. In both (a) and (b), all experimental, noise, and other point source pair parameters are as given in Fig. 3.

the SEP-SNG modality in Fig. 3(b). This is due to the fact that while this modality removes the overlap of the point spread functions of the two point sources, it does not solve the problem of having to localize an individual near-focus point source. To demonstrate that poor (good) z -localization accuracy translates to poor (good) distance estimation accuracy, we plot in Fig. 6(a) the 2D slice $d = 200$ nm from Fig. 3(b), together with the corresponding limits of the localization accuracy of the axial coordinates z_{01} and z_{02} of the two point sources at each axial (s_z) position of the point source pair. This limit of the z -localization accuracy is based on the same theoretical framework [30] as the resolution measure, and its derivation can be found in [14]. As it is a lower bound on the standard deviation of the z location estimates of a point source, high (low) values indicate poor (good) accuracy, just as is the case with the resolution measure.

Fig. 6(a) shows that the behavior of the resolution measure for the SEP-SNG modality correlates with the behaviors of the limits of the z -localization accuracy of the two point sources. Above a defocus of roughly $|s_z| = 500$ nm, both the resolution measure and the limits of the z -localization accuracy exhibit a deteriorating trend as the point source pair is positioned farther away, in either direction, from the focal plane at $s_z = 0$ nm. Below a defocus of roughly $|s_z| = 500$ nm, the resolution measure exhibits, as explained above, the two sharp increases about the focal plane. The sharp deterioration below the focal plane, which corresponds to P_1 coming close to being in focus, coincides accordingly with the sharp deterioration of the z -localization accuracy of P_1 , wherein z_{01} approaches 0 nm. Similarly, the sharp deterioration above the focal plane corresponds to P_2 coming close to being in focus, and coincides with the sharp deterioration of the z -localization accuracy of P_2 , wherein z_{02} approaches 0 nm.

Unlike the SIM-SNG and the SEP-SNG modalities, the SIM-MUM modality is able to overcome the near-focus depth discrimination problem. For this modality, Fig. 3(c) shows that the two sharp deteriorations about the focal plane at $s_z = 0$ nm are absent. Instead, for a given value of d (except $d = 0$ nm), the resolution measure remains flat in the near-focus region, indicating that good distance estimation accuracy can be expected from the SIM-MUM modality even when one of the point sources is near-focus. This is due to the fact that while a point source may be near-focus from the perspective of the focal plane at $s_z = 0$ nm, it is a good distance away from the second focal plane at $s_z = 500$ nm. The image acquired from the second focal plane therefore contains enough information to compensate for the lack thereof in the image acquired from the first focal plane. Thus, by using both images in the estimation of the distance of separation, the combined information is sufficient to yield a good accuracy.

The SEP-MUM modality incorporates the two-plane MUM configuration, and can therefore also be expected to overcome the near-focus depth discrimination problem. Fig. 3(d) shows this to be the case, with the resolution measure staying flat in the near-focus region for a given value of d not equal to 0 nm. Importantly, the SEP-MUM modality is the only modality out of the four that is able to overcome both the small distance of separation problem and the near-focus depth discrimination problem. By combining the principles of separate detection and MUM, Fig. 3(d) shows that it is able to achieve a comparatively low and flat resolution measure over all distances of separation 200 nm and below (except $d = 0$ nm), and over the four-micron axial range centered at the focal plane.

We note that for all four modalities, the plots of Fig. 3 show that the resolution measure generally increases as a non-near-focus point source pair is moved away from the focal plane(s) in either direction along the z -axis. This outward deterioration can be explained intuitively by the fact that the farther away two closely spaced point sources are from focus, the more they will appear to be a single point source in the acquired image, and hence the more difficult the distance estimation.

For a comparison of the four modalities with precise numbers, Fig. 6(b) plots in one set of axes the 2D slice $d = 200$ nm from each of the four 3D plots of Fig. 3. To give an example of the important advantage gained with MUM when a point source is near-focus, consider the point source pair centered at $s_z = 64.60$ nm, which corresponds to $z_{02} = -6.11$ nm and hence places point source P_2 near the focal plane. With the SIM-SNG and the SEP-SNG modalities, extremely poor resolution measures of ± 525.60 nm and ± 277.07 nm (out of the plot's range), respectively, can be expected in estimating the 200 nm distance of separation. In significant contrast, resolution measures of no worse than $\pm 10\%$ of the 200 nm distance can be expected from the SIM-MUM (± 19.23 nm) and the SEP-MUM (± 12.74 nm) modalities.

Note that the same general effects of near-focus depth discrimination can be observed from the plots of Fig. 7, where the resolution measures for the modalities are shown as functions of s_z and the orientation angle ω . For a given value of ω , all four modalities lose accuracy as a non-near-focus point source pair is moved away from the focal plane(s). In the near-focus region, however, the SIM-SNG (Fig. 7(a)) and the SEP-SNG (Fig. 7(b)) modalities exhibit the pair of sharp deteriorations about the focal plane, whereas the SIM-MUM (Fig. 7(c)) and the SEP-MUM (Fig. 7(d)) modalities do not. (Note that when $\omega = 90^\circ$ such that the two point sources have the same z position (i.e., $z_{01} = z_{02}$), only a single point deterioration is observed exactly at the focal plane $s_z = 0$ nm in Figs. 7(a) and 7(b). In this special scenario, the resolution measure actually remains small for axial positions immediately about $s_z = 0$ nm. Also, as noted in Section 4.1, the resolution measure is infinity when $\omega = 0^\circ$. This is the case for all modalities for any s_z .)

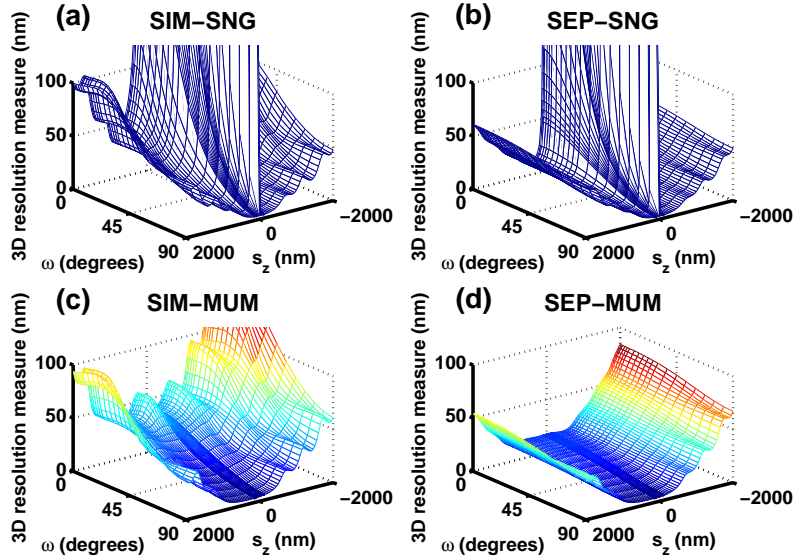


Fig. 7. Dependence of the 3D resolution measure for (a) the SIM-SNG, (b) the SEP-SNG, (c) the SIM-MUM, and (d) the SEP-MUM modality on the axial position s_z and the orientation angle ω of the imaged point source pair. In (a), (b), (c), and (d), the two point sources are separated by a distance of $d = 200$ nm. Except for s_z and ω which are varied in these plots, all experimental, noise, and other point source pair parameters are as given in Fig. 3.

4.3. Near-parallel orientations with respect to the z -axis

In addition to its distance of separation and axial position, the orientation of a point source pair with respect to the z -axis (i.e., the angle ω ; see Fig. 2) is an important determinant of the limit of the distance estimation accuracy [21]. Given a point source pair, the task of estimating its distance of separation can be expected to be relatively easy when the two point sources are positioned side by side ($\omega = 90^\circ$; perpendicular to the z -axis). This is because this side-by-side orientation can be expected to produce the least overlap between the point spread functions of the two point sources in the acquired image. However, as this point source pair is rotated towards parallel alignment with the z -axis ($\omega = 0^\circ$), such that one point source starts to go in front of, and hence obscure the other, the distance estimation problem becomes tougher as more point spread function overlap can be expected.

This effect of orientation is demonstrated by Figs. 5(a) and 5(c), which correspond respectively to the SIM-SNG and the SIM-MUM modalities. For a given distance of separation d (except $d = 0$ nm) in these plots, it can be seen that the 3D resolution measure increases as the point source pair is rotated from the side-by-side orientation of $\omega = 90^\circ$ towards the front-and-back orientation of $\omega = 0^\circ$. This deterioration of the resolution measure progresses relatively slowly when going from $\omega = 90^\circ$ to roughly $\omega = 45^\circ$, but continues at a significantly faster rate as ω decreases further before leveling off when approaching 0° (at which point the resolution measure is infinity as mentioned previously).

Just as the loss of accuracy due to small distances of separation can be avoided by separating the detection of the two closely spaced point sources, the deteriorative effect of near-parallel orientations (i.e., small ω) can be significantly lessened by the same principle. This can be seen in Figs. 5(b) and 5(d), where for a given value of d (except $d = 0$ nm), the resolution measures

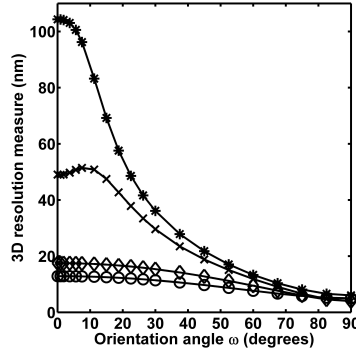


Fig. 8. Dependence of the 3D resolution measures for the SIM-SNG (*), the SEP-SNG (○), the SIM-MUM (×), and the SEP-MUM (◊) modalities on the orientation angle ω of the imaged point source pair. The curves shown correspond to the 2D slice $d = 200$ nm from each of the four 3D plots of Fig. 5. All experimental, noise, and other point source pair parameters are as described in Fig. 5.

for the SEP-SNG and the SEP-MUM modalities, respectively, stay low and almost flat over the entire range of values for ω (except at $\omega = 0^\circ$ where they are infinity). The advantage conferred by these two modalities here can again be attributed to the removal of the overlap of the point spread functions of the two point sources, which in this case is caused by small values of ω . It is important to note at this point that, by also being able to overcome the deteriorative effect associated with near-parallel orientations, the SEP-MUM modality is the only modality considered here that is capable of addressing all three challenges associated with distance, depth discrimination, and orientation.

To compare the four modalities with precise numbers, Fig. 8 shows in one set of axes the 2D slice $d = 200$ nm from each of the four 3D plots of Fig. 5. At values of ω larger than approximately 45° , resolution measures of around, or better than, ± 20 nm (i.e., $\pm 10\%$ of the 200 nm distance) can be expected for all four modalities. At $\omega = 75^\circ$, for example, the resolution measures for the SIM-SNG, the SEP-SNG, the SIM-MUM, and the SEP-MUM modalities are ± 8.10 nm, ± 5.82 nm, ± 6.62 nm, and ± 5.96 nm, respectively. At values of ω smaller than approximately 45° , however, substantial differences can be observed between the modalities. Whereas the SIM-SNG and the SIM-MUM modalities both lose a significant amount of accuracy, the SEP-SNG and the SEP-MUM modalities experience only a small deterioration. At $\omega = 7.5^\circ$, for example, poor limits of the accuracy of ± 96.23 nm and ± 51.42 nm can be expected, respectively, from the SIM-SNG and the SIM-MUM modalities. These numbers correspond, respectively, to about $\pm 50\%$ and $\pm 25\%$ of the 200 nm distance. In contrast, for the same small angle, much better resolution measures of ± 12.77 nm and ± 17.37 nm can be expected, respectively, from the SEP-SNG and the SEP-MUM modalities. In fact, across the entire range of values for ω (except $\omega = 0^\circ$), these two modalities are able to maintain accuracies of better than $\pm 10\%$ of the 200 nm distance.

Note that the same general effect that the orientation has on the resolution measure for each modality can be seen in the plots of Fig. 7. For a given value of s_z , the nonlinear deterioration of the resolution measure is observed for the SIM-SNG (Fig. 7(a)) and the SIM-MUM (Fig. 7(c)) modalities as the value of ω is decreased from 90° to 0° . For the SEP-SNG (Fig. 7(b)) and the SEP-MUM (Fig. 7(d)) modalities, the resolution measure remains, for a given s_z , relatively low and flat throughout the entire range of angles (except $\omega = 0^\circ$). (Note that due to the near-focus depth discrimination problem (Section 4.2), the patterns described are not observed in Figs.

7(a) and 7(b) for some s_z values in the region occupied by the pair of sharp deteriorations about the focal plane. In these cases, the patterns are interrupted by a sharp deterioration over values of ω which put one of the point sources near the focal plane.)

5. Maximum likelihood estimation with simulated images

As defined in Section 3, the 3D resolution measure represents the smallest possible standard deviation that can be expected of the distance estimates produced by an unbiased estimator. In practice, it is useful to know that estimators exist which are capable of attaining this limit of the distance estimation accuracy. In [21], we presented results of estimations, performed on simulated images of point source pairs, which demonstrate the maximum likelihood estimator to be capable of achieving the resolution measure corresponding to the SIM-SNG imaging modality. Here, we extend those results with that from estimations which were carried out on images simulated for the SEP-SNG, the SIM-MUM, and the SEP-MUM modalities. These results show that the maximum likelihood estimator is also capable of attaining the resolution measures corresponding to these modalities. At the same time, they provide a verification of the correctness of the theory presented in Section 3 and the appendix.

5.1. Methods

Since the pixels comprising an image can be assumed to be independent measurements of photon count, the log-likelihood function that is maximized in our estimations is simply the sum of the log-likelihood functions for the individual pixels. For an image of N_p pixels produced by the SIM-SNG modality, the log-likelihood function is then given by

$$\ln(\mathcal{L}(\theta|z_1, \dots, z_{N_p})) = \sum_{k=1}^{N_p} \ln(p_{\theta,k}(z_k)), \quad (2)$$

where $p_{\theta,k}(z_k)$, $k = 1, \dots, N_p$, is the probability that z_k photons are detected at the k^{th} pixel, and is given by Eq. (7). The parameter vector θ that is estimated is as given in Section 3.1, consisting of the six parameters (including the distance of separation d) that together describe the 3D geometry of a point source pair (Fig. 2).

For an image of N_p pixels produced by the SEP-SNG, the SIM-MUM, or the SEP-MUM modality, the log-likelihood function is also as given by Eq. (2), but with one important caveat. Depending on whether the image is of point source P_1 or P_2 or both, and of focal plane 1 or 2, $p_{\theta,k}(z_k)$ of Eq. (7) is evaluated using the appropriate redefinitions of the intensity function Λ_θ and the spatial density functions $\{f_{\theta,\tau}\}_{\tau \geq t_0}$ detailed in appendix sections A.2, A.3, and A.4. In addition, since these modalities entail the use of either two or four independently-formed images (Figs. 1(b), 1(c), 1(d)) to arrive at a single estimate of the distance d , the overall log-likelihood function to be maximized is just the sum of the log-likelihood functions for the individual images.

Maximum likelihood estimations were performed on nine different data sets of simulated images of point source pairs. Depending on the imaging modality to which it corresponds, a data set consists of 500 images, 500 pairs of images, or 500 sets of four images (Fig. 1). In this way, 500 distance estimates are obtained for each data set regardless of the modality. Each image is a 15x15 pixel array of 13 μm by 13 μm pixels. In addition to the imaging modality, the various data sets differ in terms of the point source pair's distance of separation d , axial position s_z , and orientation angle ω .

Each individual image in a data set corresponding to the SIM-SNG modality is simulated according to the description of a practically acquired image in appendix section A.1. That is, the total photon count at each pixel is simulated as the sum of the photon counts due to the

point source pair, the background noise, and the detector measurement noise. More precisely, the photon count due to the point source pair is a realization of a Poisson-distributed random variable with mean given by Eq. (5). This mean is evaluated by assuming an expected photon count of 5000 photons per point source, and that the image of each point source is described by the 3D point spread function of Born and Wolf (Eq. (1)). The photon count due to the background noise is a realization of a Poisson-distributed random variable with a mean of 80 photons. The photon count due to the measurement noise is a realization of a Gaussian-distributed random variable with a mean of $0 e^-$ and a standard deviation of $8 e^-$.

An individual image in a data set corresponding to the SEP-SNG, the SIM-MUM, or the SEP-MUM modality is simulated similarly as described for an image of the SIM-SNG modality. As with the computation of the log-likelihood function, however, the mean given by Eq. (5) is evaluated using the appropriate redefinitions of the functions Λ_θ and $\{f_{\theta,\tau}\}_{\tau \geq t_0}$ presented in appendix sections A.2, A.3, and A.4, depending on whether the image is of point source P_1 or P_2 or both, and of focal plane 1 or 2. Also, since a 50:50 splitting of the collected fluorescence between the two focal planes is assumed for the SIM-MUM and the SEP-MUM modalities, the expected photon count per point source is halved to 2500, and the mean of the background noise at each pixel is halved to 40 photons for these modalities.

The software used to simulate the data sets, perform the estimations, and compute the resolution measures is implemented using the technical programming language of MATLAB (The MathWorks, Inc., Natick, MA) and its optimization toolbox. It comprises the core functionalities of the software packages EstimationTool [32] and FandPLimitTool [33].

5.2. Results

Table 1 shows the estimation results for the nine simulated data sets. In every case, the true distance of separation is very closely recovered by the mean of the distance estimates. Moreover, the standard deviation of the distance estimates of each data set is in close agreement with the 3D resolution measure corresponding to the data set. These results therefore demonstrate the ability of the maximum likelihood estimator to attain the limit of the distance estimation accuracy predicted for each modality.

Data sets 1 through 4 simulate the imaging, using each of the four modalities, of a point source pair with a relatively large distance of separation of $d = 200$ nm, an axial position of $s_z = 400$ nm that places both point sources well away from the focal plane ($z_{01} = 450$ nm, $z_{02} = 350$ nm), and an orientation of $\omega = 60^\circ$ that is far from parallel alignment with the z -axis. Given this relatively easy scenario, resolution measures of no worse than $\pm 7\%$ of the 200 nm distance are predicted and attained for the SIM-SNG (± 13.71 nm) and the SIM-MUM (± 12.09 nm) modalities. Even better accuracy limits of no worse than $\pm 5\%$ of 200 nm are predicted and attained for the SEP-SNG (± 7.82 nm) and the SEP-MUM (± 9.59 nm) modalities.

Data sets 5 and 6 simulate the imaging of a point source pair with a smaller distance of separation of $d = 100$ nm and an orientation of $\omega = 30^\circ$ that is closer to parallel alignment with the z -axis. These two data sets demonstrate that, given the more challenging separation distance and orientation, good accuracy limits of ± 11.12 nm and ± 15.47 nm are still predicted and attained, respectively, for the SEP-SNG and the SEP-MUM modalities. (Poor accuracy limits of ± 73.22 nm and ± 56.78 nm are predicted for the SIM-SNG and the SIM-MUM modalities, respectively.) Data sets 7 and 8 entail a point source pair with the same distance of separation and orientation as that of data sets 1 through 4, but with an axial position of $s_z = 75$ nm that places point source P_2 only 25 nm away from the focal plane. Given this point source that is close to the focal plane, these data sets show that excellent accuracy limits of ± 12.29 nm and ± 9.58 nm are still predicted and attained, respectively, for the SIM-MUM and the SEP-MUM modalities. (Poor accuracy limits of ± 85.27 nm and ± 49.23 nm are predicted for the SIM-

Table 1. Results of maximum likelihood estimations

Data set no.	Imaging modality	Axial position s_z (nm)	Orientation angle ω (degrees)	Distance of separation d (nm)	Mean of d estimates (nm)	Resolution measure (nm)	Standard deviation of d estimates (nm)
1	SIM-SNG	400	60	200	201.09	13.71	13.93
2	SEP-SNG	400	60	200	200.60	7.82	7.91
3	SIM-MUM	400	60	200	201.43	12.09	12.10
4	SEP-MUM	400	60	200	200.81	9.59	9.53
5	SEP-SNG	400	30	100	99.95	11.12	11.29
6	SEP-MUM	400	30	100	100.38	15.47	15.49
7	SIM-MUM	75	60	200	200.65	12.29	12.03
8	SEP-MUM	75	60	200	200.63	9.58	9.66
9	SEP-MUM	75	30	100	100.44	15.54	15.85

Results of maximum likelihood estimations on nine sets of simulated images of point source pairs in 3D space. The image of each point source is simulated according to the Born and Wolf 3D point spread function, and the expected photon count detected per point source is 5000. For each data set, the mean and the standard deviation of 500 estimates of the distance of separation are shown with the corresponding resolution measure as well as the axial position s_z , the orientation angle ω , and the distance of separation d which are different between the data sets. The data sets also differ in terms of the imaging modality. A data set corresponding to the SIM-SNG modality contains 500 simulated images. A data set corresponding to the SEP-SNG or the SIM-MUM modality contains 500 pairs of simulated images. A data set corresponding to the SEP-MUM modality contains 500 sets of four images. (See Fig. 1.) In all data sets, a single image is a 15×15 array of $13 \mu\text{m}$ by $13 \mu\text{m}$ pixels. With an assumed magnification of $M = 100$, the position of the point source pair in the xy -plane is set to $s_x = s_y = 975 \text{ nm}$ so that its image is centered on the pixel array. All other experimental, noise, and point source pair parameters are as given in Fig. 3.

SNG and the SEP-SNG modalities, respectively.) Lastly, data set 9 demonstrates that, given the combination of $d = 100 \text{ nm}$, $\omega = 30^\circ$, and $s_z = 75 \text{ nm}$, a good accuracy limit of $\pm 15.54 \text{ nm}$ is still predicted and attained for the SEP-MUM modality. (Poor accuracy limits of $\pm 488.43 \text{ nm}$, $\pm 66.84 \text{ nm}$, and $\pm 57.33 \text{ nm}$ are predicted for the SIM-SNG, the SEP-SNG, and the SIM-MUM modalities, respectively.)

It is well known that under relatively weak conditions the maximum likelihood estimator of a parameter vector θ is asymptotically Gaussian distributed with asymptotic mean θ and covariance $\mathbf{I}^{-1}(\theta)$ (e.g., [34]), where the latter is the Cramer-Rao lower bound (Eq. (8)) on which the resolution measure is based. This means that our estimator is asymptotically unbiased and asymptotically efficient. However, in the case of finite data sets such as those shown in Table 1, it is difficult to analytically determine the accuracy of our complex estimator and whether or not it is biased. We have therefore relied on simulation studies instead. Having examined many data sets such as those shown in Table 1, we can draw the conclusion that if there is bias, then it is a very small one provided that the resolution measure is small compared to the estimated distance. For example, for every data set in Table 1, the mean differs from the true distance by less than 1% of the true distance. These data sets entail relatively small resolution measures that are no worse than $\pm 16\%$ of the corresponding true distances. Importantly, this shows that under the practically desirable condition where the resolution measure predicts a good accuracy limit, the maximum likelihood estimator is able to recover the true distance with very little bias if it in fact exists. Moreover, under this condition, we can conclude based on the consistency with which it does so in our simulation studies, that the maximum likelihood estimator is able

to attain the resolution measure.

6. Conclusions

Using a Cramer-Rao lower bound-based 3D resolution measure which specifies the best possible accuracy (standard deviation) with which the distance separating two point sources can be estimated, we have compared the limits of the distance estimation accuracy that can be expected for images acquired using four different fluorescence microscopy imaging modalities. These modalities represent imaging setups that have been realized in practice, and comprise all possible combinations of simultaneous or separate fluorophore detection coupled with a single focal plane or a multifocal plane imaging configuration.

Specifically, we have compared the resolution measures for the four modalities in the context of the challenges posed by small distances of separation, near-focus depth discrimination, and orientations near parallel to the microscope's z -axis. We have shown that the conventional fluorescence imaging setup generally performs poorly under these conditions. However, modalities that implement the MUM technique are able to overcome the near-focus depth discrimination problem, and modalities that employ the separate detection of fluorophores are able to deal with point source pairs with small distances of separation and/or near-parallel orientations. Moreover, though it does not always yield the best limit of the distance estimation accuracy numerically, the modality that combines the principles of separate detection and multifocal plane imaging is the only modality that is able to provide generally good accuracy limits when all three problems are present.

The method of comparison using the resolution measure and the results we report are useful in practice as they can be used to produce important guidelines in the design of an imaging experiment where distance estimation needs to be performed on the acquired images. For the actual estimation of the distance of separation, we have also shown with simulated images that the maximum likelihood estimator is capable of attaining the limits of the accuracy predicted by the resolution measures corresponding to all four modalities. This estimator might therefore make a good choice for the analysis of real image data.

In comparing the resolution measures for the different modalities, we have used a specific point spread function and a specific set of parameters (e.g., numerical aperture, detector pixel size, magnification, plane spacing of the two-plane MUM configuration, etc.). Therefore, the precise values of the resolution measures that are shown in the plots and given in the text apply specifically to the conditions we have assumed. In practice, the resolution measure should be computed with a point spread function (e.g., [35, 36, 37]) and parameters that pertain to the specific imaging setup that is used. As can be seen from the generality of the theory presented in Section 3 and the appendix, the mathematical framework on which the resolution measure is based can easily support such necessary customization.

Importantly, the conclusions we have drawn from our comparisons are generally true regardless of the specific point spread function and the specific values of the parameters. While a different set of conditions would produce different numerical values for the resolution measure, it would not change the general patterns seen in the various plots on which our conclusions are based. To illustrate this point, we show in Fig. 9 the same 2D plots as in Figs. 4(a), 6(b), and 8, but for a numerical aperture of $n_a = 1.2$, an immersion medium refractive index of $n = 1.33$, and a magnification of $M = 63$. We see that while the plots of Fig. 9 do not show the same numerical values as in Figs. 4(a), 6(b), and 8, they demonstrate the same general behavior of the resolution measures as functions of a point source pair's distance of separation, axial position, and orientation.

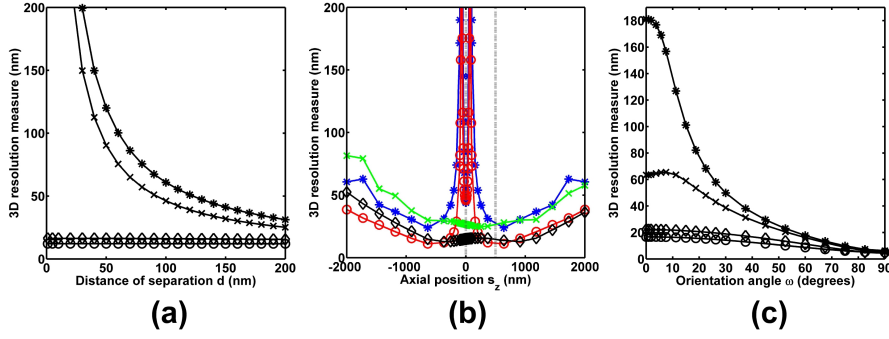


Fig. 9. Dependence of the 3D resolution measures for the SIM-SNG (*), the SEP-SNG (o), the SIM-MUM (x), and the SEP-MUM (◊) modalities on (a) the distance of separation d , (b) the axial position s_z , and (c) the orientation angle ω of the imaged point source pair. The plots in (a), (b), and (c) are analogous to the plots of Figs. 4(a), 6(b), and 8, respectively, but with the object space medium refractive index changed to $n = 1.33$, and the numerical aperture and the magnification of the objective lens changed to $n_a = 1.2$ and $M = 63$, respectively. (In order that the image of the point source pair is still centered on the 21×21 array of $13 \mu\text{m}$ by $13 \mu\text{m}$ pixels given the new magnification, the position of the point source pair in the xy -plane has also been changed to $s_x = s_y \approx 2167 \text{ nm}$. Also, the magnification associated with focal plane 2 of the SIM-MUM and the SEP-MUM modalities is accordingly changed to $M' = 62.42$.) All other experimental, noise, and point source pair parameters are as described in the corresponding Figs. 4(a), 6(b), and 8.

Acknowledgments

This work was supported in part by the National Institutes of Health (R01 GM071048 and R01 GM085575), and in part by the National Multiple Sclerosis Society (FG-1798-A-1).

Appendix

A.1 The 3D resolution measure for the SIM-SNG modality

Let us consider a single image of both point sources P_1 and P_2 (Fig. 1(a)) that is acquired during the time interval $[t_0, t]$, and from which the unknown parameter vector $\theta = (d, \phi, \omega, s_x, s_y, s_z)$ is to be estimated (see Fig. 2). Assuming at first the absence of any extraneous noise sources (e.g., sample autofluorescence, detector dark current, detector readout), such an image is formed strictly from the photons detected from the point source pair. Since photon emission (and hence photon detection) is stochastic by nature, this image is modeled as a spatio-temporal random process [30, 31]. The temporal aspect models the time points at which the photons are detected as an inhomogeneous Poisson process with intensity function

$$\Lambda_\theta(\tau) = \Lambda_1(\tau) + \Lambda_2(\tau), \quad \tau \geq t_0, \quad (3)$$

where Λ_1 and Λ_2 are the photon detection rates of P_1 and P_2 , respectively. The spatial aspect models the image space coordinates of the detected photons as a sequence of independent random variables with probability density functions $\{f_{\theta,\tau}\}_{\tau \geq t_0}$, where each $f_{\theta,\tau}$ is given by

$$f_{\theta,\tau}(x, y) = \frac{1}{M^2} \left[\varepsilon_1(\tau) q_{z_{01},1} \left(\frac{x}{M} - x_{01}, \frac{y}{M} - y_{01} \right) + \varepsilon_2(\tau) q_{z_{02},2} \left(\frac{x}{M} - x_{02}, \frac{y}{M} - y_{02} \right) \right], \quad (4)$$

where $(x, y) \in \mathbb{R}^2$, $\varepsilon_i(\tau) = \Lambda_i(\tau) / (\Lambda_1(\tau) + \Lambda_2(\tau))$, $i = 1, 2$, $\tau \geq t_0$, M denotes the lateral magnification of the microscope, (x_{01}, y_{01}, z_{01}) and (x_{02}, y_{02}, z_{02}) denote, respectively, the 3D co-

ordinates of the positions of P_1 and P_2 in the object space (see Fig. 2), and $q_{z_{01},1}$ and $q_{z_{02},2}$ denote, respectively, the point spread functions of P_1 and P_2 . Note that the position coordinates of P_1 and P_2 are functions of the unknown parameter vector θ , and can be written explicitly as $x_{01} = s_x + d \sin \omega \cos \phi/2$, $y_{01} = s_y + d \sin \omega \sin \phi/2$, $z_{01} = s_z + d \cos \omega/2$, $x_{02} = s_x - d \sin \omega \cos \phi/2$, $y_{02} = s_y - d \sin \omega \sin \phi/2$, and $z_{02} = s_z - d \cos \omega/2$. Also, the point spread functions $q_{z_{01},1}$ and $q_{z_{02},2}$ are each specified in the form of an image function [30], which gives the image, at unit lateral magnification, of a point source that is located at $(0, 0, z_0)$, $z_0 \in \mathbb{R}$, in the object space.

The spatio-temporal random process we have described readily models a non-pixelated image of infinite size. For a finite-sized image consisting of N_p pixels, however, the collected data is a sequence of independent random variables $\{S_{\theta,1}, \dots, S_{\theta,N_p}\}$, where $S_{\theta,k}$ is the number of photons detected at the k^{th} pixel. In [30], we showed that the photon count $S_{\theta,k}$ is Poisson-distributed with mean

$$\mu_{\theta}(k) = \int_{t_0}^t \int_{C_k} \Lambda_{\theta}(\tau) f_{\theta,\tau}(x,y) dx dy d\tau, \quad (5)$$

where $[t_0, t]$ is again the image acquisition time interval, C_k is the region in the detector plane that is occupied by the k^{th} pixel, and the functions $\Lambda_{\theta}(\tau)$ and $f_{\theta,\tau}(x,y)$ are as defined above.

We have thus far given a description of an image that is formed from only the photons detected from the point sources. An image produced by a practical microscopy imaging experiment, however, will also contain photons contributed by extraneous noise sources such as autofluorescence from the imaged sample and the detector readout process. To account for these signal-deteriorating photons, a practical image of N_p pixels is modeled as a sequence of independent random variables $\{\mathcal{S}_{\theta,1}, \dots, \mathcal{S}_{\theta,N_p}\}$, where $\mathcal{S}_{\theta,k}$ is the total photon count at the k^{th} pixel. At each pixel k , the total photon count $\mathcal{S}_{\theta,k}$ is modeled as the sum of the three mutually independent random variables $S_{\theta,k}$, B_k , and W_k . The random variable $S_{\theta,k}$ represents again the number of photons from the point source pair which are detected at the k^{th} pixel. It is dependent on the unknown parameter vector θ , and is Poisson-distributed with mean $\mu_{\theta}(k)$ given by Eq. (5). The random variable B_k represents the number of spurious photons at the k^{th} pixel which arise from extraneous noise sources such as sample autofluorescence and the detector dark current. It is assumed to be Poisson-distributed with mean $\beta(k)$. The random variable W_k represents the number of photons at the k^{th} pixel which are due to measurement noise sources such as the detector readout process. It is assumed to be Gaussian-distributed with mean η_k and standard deviation σ_k . Note that neither B_k nor W_k is dependent on θ , which contains parameters pertaining only to the point source pair.

Given our stochastic model of a practical image, the Fisher information matrix $\mathbf{I}(\theta)$ corresponding to an image of N_p pixels is then given by [30]

$$\mathbf{I}(\theta) = \sum_{k=1}^{N_p} \left(\frac{\partial \mu_{\theta}(k)}{\partial \theta} \right)^T \frac{\partial \mu_{\theta}(k)}{\partial \theta} \left(\int_{\mathbb{R}} \frac{\left(\sum_{l=1}^{\infty} \frac{[v_{\theta}(k)]^{l-1} e^{-v_{\theta}(k)}}{(l-1)!} \cdot \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{1}{2} \left(\frac{z-l-\eta_k}{\sigma_k} \right)^2} \right)^2}{p_{\theta,k}(z)} dz - 1 \right), \quad (6)$$

where for $k = 1, \dots, N_p$, $v_{\theta}(k) = \mu_{\theta}(k) + \beta(k)$, and

$$p_{\theta,k}(z) = \frac{1}{\sqrt{2\pi}\sigma_k} \sum_{l=0}^{\infty} \frac{[v_{\theta}(k)]^l e^{-v_{\theta}(k)}}{l!} e^{-\frac{1}{2} \left(\frac{z-l-\eta_k}{\sigma_k} \right)^2}, \quad z \in \mathbb{R}. \quad (7)$$

Applying the Cramer-Rao inequality [20]

$$\text{Cov}(\hat{\theta}) \geq \mathbf{I}^{-1}(\theta), \quad (8)$$

where $\hat{\theta}$ is any unbiased estimator of θ , the 3D resolution measure is defined as the quantity $\sqrt{[\mathbf{I}^{-1}(\theta)]_{11}}$. It is therefore a lower bound on the standard deviation of the estimates of any unbiased estimator of the distance of separation d .

A.2 The 3D resolution measure for the SEP-SNG modality

The SEP-SNG modality produces two images that are separated in time, one of point source P_1 only, and the other of point source P_2 only (Fig. 1(b)). We first consider an image of point source P_1 that is acquired during the interval $[t_0, t_1]$. Since P_1 is the only point source that is detected, the intensity function of the Poisson process is simply given by the photon detection rate of P_1 , i.e., $\Lambda_\theta(\tau) = \Lambda_1(\tau)$, $t_0 \leq \tau \leq t_1$. Likewise, the spatial density function will involve only the point spread function of P_1 , and is given by $f_\theta(x, y) = \frac{1}{M^2} q_{z_{01}, 1} \left(\frac{x}{M} - x_{01}, \frac{y}{M} - y_{01} \right)$. Note that the subscript τ has been dropped from f_θ because unlike the density function of Eq. (4), the f_θ here does not depend on time. Substituting these redefined functions into Eq. (5), the number of photons $S_{\theta, k}$ detected at the k^{th} pixel of this image is then Poisson-distributed with mean

$$\mu_\theta^{(1)}(k) = \frac{1}{M^2} \int_{t_0}^{t_1} \int_{C_k} \Lambda_1(\tau) \cdot q_{z_{01}, 1} \left(\frac{x}{M} - x_{01}, \frac{y}{M} - y_{01} \right) dx dy d\tau, \quad (9)$$

where $\mu_\theta(k)$ is superscripted with (1) to denote image of P_1 . Finally, by evaluating Eq. (6) with $\mu_\theta^{(1)}$, we obtain the Fisher information matrix $\mathbf{I}^{(1)}(\theta)$ which corresponds to the image of P_1 .

By applying the same reasoning to an image of point source P_2 acquired during the time interval $[t_2, t_3]$ that is disjoint from $[t_0, t_1]$, we get the mean photon count

$$\mu_\theta^{(2)}(k) = \frac{1}{M^2} \int_{t_2}^{t_3} \int_{C_k} \Lambda_2(\tau) \cdot q_{z_{02}, 2} \left(\frac{x}{M} - x_{02}, \frac{y}{M} - y_{02} \right) dx dy d\tau, \quad (10)$$

where the superscript (2) of $\mu_\theta(k)$ denotes image of P_2 . Substituting it into Eq. (6), we obtain the Fisher information matrix $\mathbf{I}^{(2)}(\theta)$ which corresponds to the image of P_2 .

By the independence of the two spatio-temporal random processes, the Fisher information matrix for the SEP-SNG modality is just

$$\mathbf{I}(\theta) = \mathbf{I}^{(1)}(\theta) + \mathbf{I}^{(2)}(\theta), \quad (11)$$

and by definition the 3D resolution measure is the quantity $\sqrt{[\mathbf{I}^{-1}(\theta)]_{11}}$.

A.3 The 3D resolution measure for the SIM-MUM modality

The SIM-MUM modality produces two images, each of both point sources, that are acquired during the same time interval, but from different focal planes (Fig. 1(c)). Let us consider the acquisition time interval $[t_0, t]$. Since each image acquired during this time is that of both point sources P_1 and P_2 , the intensity function of the temporal portion of its respective spatio-temporal random process will remain the same as that for the SIM-SNG modality. That is, it is still the sum of the photon detection rates of P_1 and P_2 , given by Eq. (3).

For each image, the density functions $\{f_{\theta, \tau}\}_{\tau \geq t_0}$ of the spatial portion of its respective random process will look similar to that for the SIM-SNG modality. However, due to the different focal planes that are associated with the two images, two important modifications need to be made. First, for each image the axial position of the point source pair needs to be specified with respect to the focal plane from which the image was acquired. Second, the lateral magnification of the two images will be different because their respective focal planes are located at different distances from the objective lens.

For ease of reference, let us designate the two distinct focal planes as focal plane 1 and focal plane 2. Without loss of generality, let focal plane 1 be the plane that is positioned closer

to the objective lens, and let the z coordinates of point sources P_1 and P_2 (i.e., z_{01} and z_{02}) be given with respect to focal plane 1. Furthermore, let M denote the lateral magnification associated with focal plane 1. Then, for the image corresponding to focal plane 1, each density function $f_{\theta,\tau}$ is exactly as given by Eq. (4). Accordingly, the Fisher information matrix $\mathbf{I}^{(1)}(\theta)$ corresponding to the image from focal plane 1 is readily given by Eq. (6).

Let focal plane 2 be positioned at a plane spacing of Δz_f above focal plane 1. Then, with respect to focal plane 2, the z coordinates z_{01} and z_{02} of the same point sources become $z_{01} - \Delta z_f$ and $z_{02} - \Delta z_f$, respectively. Moreover, the lateral magnification M' that is associated with focal plane 2 can be determined using the geometrical optics-based relationship [38]

$$M' = M \frac{L - \frac{L \cdot M^2 \cdot \Delta z_f}{n \cdot L + M^2 \cdot \Delta z_f}}{L}, \quad (12)$$

where n is the refractive index of the object space medium, and L is the tube length of the microscope. By substituting the modified axial positions and magnification into Eq. (4), we obtain the density functions $\{f_{\theta,\tau}\}_{\tau \geq t_0}$ for the image corresponding to focal plane 2. That is, each density function $f_{\theta,\tau}$ will be given by

$$f_{\theta,\tau}(x, y) = \frac{1}{(M')^2} \left[\varepsilon_1(\tau) q_{z_{01} - \Delta z_f, 1} \left(\frac{x}{M'} - x_{01}, \frac{y}{M'} - y_{01} \right) + \varepsilon_2(\tau) q_{z_{02} - \Delta z_f, 2} \left(\frac{x}{M'} - x_{02}, \frac{y}{M'} - y_{02} \right) \right], \quad (13)$$

where the photon detection rate-dependent functions ε_1 and ε_2 are as defined for Eq. (4). Finally, by using Eq. (13) in Eq. (5) and substituting the result into Eq. (6), we obtain the Fisher information matrix $\mathbf{I}^{(2)}(\theta)$ corresponding to the image from focal plane 2.

By the independence of the two images, the Fisher information matrix $\mathbf{I}(\theta)$ for the SIM-MUM modality is then given by Eq. (11), and the 3D resolution measure is just the quantity $\sqrt{[\mathbf{I}^{-1}(\theta)]_{11}}$.

A.4 The 3D resolution measure for the SEP-MUM modality

The SEP-MUM modality combines separate detection with the two-plane MUM technique. It therefore produces two pairs of images (Fig. 1(d)). One pair, each of point source P_1 only, is acquired simultaneously by two cameras during time interval $[t_0, t_1]$, but from two distinct focal planes. The other pair, each of point source P_2 only, is acquired simultaneously by the same two cameras during time interval $[t_2, t_3]$ that is disjoint from $[t_0, t_1]$, but from the same two distinct focal planes. Each of the four images is hence characterized by a unique combination of time and focal plane, and can be assumed to be formed independently of the other three images.

By the independence of the four images, we can again add their respective Fisher information matrices. These four matrices can be obtained by applying the same arguments used in appendix sections A.2 and A.3. Using the same designation of the focal planes as in appendix section A.3, the matrices for the two images corresponding to focal plane 1 are readily obtained by evaluating Eq. (6) with Eqs. (9) and (10). Use of Eq. (9) (Eq. (10)) gives the matrix for the image of P_1 (P_2) acquired during $[t_0, t_1]$ ($[t_2, t_3]$).

The matrices for the two images corresponding to focal plane 2 are easily obtained by evaluating Eq. (6) with modified versions of Eqs. (9) and (10). These functions are given by

$$\mu_{\theta}^{(1)}(k) = \frac{1}{(M')^2} \int_{t_0}^{t_1} \int_{C_k} \Lambda_1(\tau) \cdot q_{z_{01} - \Delta z_f, 1} \left(\frac{x}{M'} - x_{01}, \frac{y}{M'} - y_{01} \right) dx dy d\tau \quad (14)$$

and

$$\mu_{\theta}^{(2)}(k) = \frac{1}{(M')^2} \int_{t_2}^{t_3} \int_{C_k} \Lambda_2(\tau) \cdot q_{z_{02} - \Delta z_f, 2} \left(\frac{x}{M'} - x_{02}, \frac{y}{M'} - y_{02} \right) dx dy d\tau, \quad (15)$$

where M' is again the lateral magnification of focal plane 2 given by Eq. (12), and Δz_f is again the spacing between the two focal planes. Use of Eq. (14) (Eq. (15)) gives the matrix for the image of P_1 (P_2) acquired during $[t_0, t_1]$ ($[t_2, t_3]$).

The 3D resolution measure for the SEP-MUM modality is then the quantity $\sqrt{[\mathbf{I}^{-1}(\boldsymbol{\theta})]_{11}}$, where $\mathbf{I}(\boldsymbol{\theta})$ is the sum of the four Fisher information matrices corresponding to the four images.