

Model of bleaching and acquisition for superresolution microscopy controlled by a single wavelength

Alex Small

*Department of Physics, California State Polytechnic University, 3801 West Temple Avenue,
Pomona, CA 91768, USA*

*arsmall@csupomona.edu

Abstract: We consider acquisition schemes that maximize the fraction of images that contain only a single activated molecule (as opposed to multiple activated molecules) in superresolution localization microscopy of fluorescent probes. During a superresolution localization microscopy experiment, irreversible photobleaching destroys fluorescent molecules, limiting the ability to monitor the dynamics of long-lived processes. Here we consider experiments controlled by a single wavelength, so that the bleaching and activation rates are coupled variables. We use variational techniques and kinetic models to demonstrate that this coupling of bleaching and activation leads to very different optimal control schemes, depending on the detailed kinetics of fluorophore activation and bleaching. Likewise, we show that the robustness of the acquisition scheme is strongly dependent on the detailed kinetics of activation and bleaching.

© 2011 Optical Society of America

OCIS codes: (100.6640) Superresolution; (180.2520) Fluorescence microscopy

References and links

1. K. Lidke, B. Rieger, T. Jovin, and R. Heintzmann, "Superresolution by localization of quantum dots using blinking statistics," *Opt. Express* **13**, 7052–7062 (2005).
2. 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).
3. M. J. Rust, M. Bates, and X. Zhuang, "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (storm)," *Nat. Methods* **3**, 793–795 (2006).
4. 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).
5. R. E. Thompson, D. R. Larson, and W. W. Webb, "Precise nanometer localization analysis for individual fluorescent probes," *Biophys. J.* **82**, 2775–2783 (2002).
6. R. J. Ober, S. Ram, and E. S. Ward, "Localization accuracy in single-molecule microscopy," *Biophysical Journal* **86**, 1185–1200 (2004).
7. H. Shroff, C. G. Galbraith, J. A. Galbraith, and E. Betzig, "Live-cell photoactivated localization microscopy of nanoscale adhesion dynamics," *Nat. Methods* **5**, 417–423 (2008).
8. S. A. Ribeiro, P. Vagnarelli, Y. Dong, T. Hori, B. F. McEwen, T. Fukagawa, C. Flors, and W. C. Earnshaw, "A super-resolution map of the vertebrate kinetochore," *Proc. Natl Acad. Sci. U.S.A.* **107**, 10484–10489 (2010).
9. D. Greenfield, A. L. McEvoy, H. Shroff, G. E. Crooks, N. S. Wingreen, E. Betzig, and J. Liphardt, "Self-organization of the *escherichia coli* chemotaxis network imaged with super-resolution light microscopy," *PLoS Biol.* **7**, e1000137 (2009).
10. N. A. Frost, H. Shroff, H. Kong, E. Betzig, and T. A. Blanpied, "Single-molecule discrimination of discrete perisynaptic and distributed sites of actin filament assembly within dendritic spines," *Neuron* **67**, 86–99 (2010).

11. S. T. Hess, T. J. Gould, M. V. Gudheti, S. A. Maas, K. D. Mills, and J. Zimmerberg, "Dynamic clustered distribution of hemagglutinin resolved at 40 nm in living cell membranes discriminates between raft theories," *Proc. Natl. Acad. Sci. U.S.A.* **104**, 17370 (2007).
12. T. A. Brown, R. D. Fetter, A. N. Tkachuk, and D. A. Clayton, "Approaches toward super-resolution fluorescence imaging of mitochondrial proteins using palm," *Methods* **51**, 458–463 (2010).
13. B. Huang, S. A. Jones, B. Brandenburg, and X. Zhuang, "Whole-cell 3d storm reveals interactions between cellular structures with nanometer-scale resolution," *Nat. Methods* **5**, 1047 (2008).
14. J. Fölling, M. Bossi, H. Bock, R. Medda, C. A. Wurm, B. Hein, S. Jakobs, C. Eggeling, and S. W. Hell, "Fluorescence nanoscopy by ground-state depletion and single-molecule return," *Nat. Methods* **5**, 943–945 (2008).
15. C. Steinhauer, C. Forthmann, J. Vogelsang, and P. Tinnefeld, "Superresolution microscopy on the basis of engineered dark states," *J. Am. Chem. Soc.* **130**, 16840–16841 (2008).
16. D. Baddeley, I. D. Jayasinghe, C. Cremer, M. B. Cannell, and C. Soeller, "Light-induced dark states of organic fluorochromes enable 30 nm resolution imaging in standard media," *Biophys. J.* **96**, L22–L24 (2009).
17. M. Heilemann, S. van de Linde, A. Mukherjee, and M. Sauer, "Super-resolution imaging with small organic fluorophores," *Angew. Chem. Int. Ed.* **48**, 6903–6908 (2009).
18. I. Testa, C. A. Wurm, R. Medda, E. Rothermel, C. Von Middendorf, J. Fölling, S. Jakobs, A. Schnle, S. W. Hell, and C. Eggeling, "Multicolor fluorescence nanoscopy in fixed and living cells by exciting conventional fluorophores with a single wavelength," *Biophys. J.* **99**, 2686–2694 (2010).
19. S. Lee, M. Thompson, M. A. Schwartz, L. Shapiro, and W. E. Moerner, "Super-resolution imaging of the nucleoid-associated protein hu in *caulobacter crescentus*," *Biophys. J.* **100**, L31–L33 (2011).
20. J. Vogelsang, T. Cordes, C. Forthmann, C. Steinhauer, and P. Tinnefeld, "Controlling the fluorescence of ordinary oxazine dyes for single-molecule switching and superresolution microscopy," *Proc. Natl. Acad. Sci. U.S.A.* **106**, 8107–8112 (2009).
21. E. Shore and A. Small, "Optimal acquisition scheme for subwavelength localization microscopy of bleachable fluorophores," *Opt. Lett.* **36**, 289–291 (2011).
22. A. Small, "Theoretical limits on errors and acquisition rates in localizing switchable fluorophores," *Biophys. J.* **96**, L16–L18 (2009).
23. F. Huang, S. L. Schwartz, J. M. Byars, and K. A. Lidke, "Simultaneous multiple-emitter fitting for single molecule super-resolution imaging," *Biomed. Opt. Express* **2**, 1377–1393 (2011).
24. H. Goldstein, *Classical mechanics* (Addison-Wesley Pub. Co., 1980).
25. B. Chachuat, *Nonlinear and Dynamic Optimization: From Theory to Practice* (Laboratoire d'Automatique, Ecole Polytechnique Federale de Lausanne, Lecture Notes for Winter Semester, 2007).
26. M. Bates, B. Huang, and X. Zhuang, "Super-resolution microscopy by nanoscale localization of photo-switchable fluorescent probes," *Curr. Opin. Chem. Biol.* **12**, 505–514 (2008).
27. T. Gould and S. Hess, "Nanoscale biological fluorescence imaging: Breaking the diffraction barrier," *Methods Cell Biol.* **89**, 329–358 (2008).
28. C. S. Smith, N. Joseph, B. Rieger, and K. A. Lidke, "Fast, single-molecule localization that achieves theoretically minimum uncertainty," *Nat. Methods* **7**, 373–375 (2010).
29. C. Eggeling, J. Widengren, R. Rigler, and C. A. M. Seidel, "Photobleaching of fluorescent dyes under conditions used for single-molecule detection: evidence of two-step photolysis," *Anal. Chem.* **70**, 2651–2659 (1998).
30. G. Donnert, C. Eggeling, and S. W. Hell, "Major signal increase in fluorescence microscopy through dark-state relaxation," *Nat. Methods* **4**, 81–86 (2007).

The diffraction limit in fluorescence microscopy can be overcome the use of molecules that can switch between a fluorescent "activated" state and a non-fluorescent "dark" state [1, 2, 3, 4]. In these techniques, only a small fraction of the molecules are fluorescing at any given time, producing a sparse image consisting of (usually) non-overlapping blurs (the shapes of which are related to the point spread function (PSF) of the imaging system). A molecular position can be determined from each blur, by fitting the intensity profile to the PSF. By repeatedly activating and localizing different subsets of the molecules, one can thus build a complete map of the fluorescently-labeled structure. The precision of the fitting procedure is determined by the number of photons collected [5, 6] and if the sample is labeled at sufficiently high density [7] then the resolution of the final image can be significantly better than $\lambda/10$ for photon counts in excess of a few hundred per molecule. These "localization microscopy" techniques are now being used to study a wide range of topics of biological significance, including the organization of chromatin during mitosis [8], the organization of proteins involved in bacterial chemotaxis [9], actin dynamics [10], clustering of membrane proteins [11], localization of proteins in mitochondria [12], and interactions between mitochondria and the cytoskeleton [13].

In initial localization microscopy experiments, the fluorophores were usually controlled by two separate wavelengths, with one bringing the molecule from the dark state to the activated state and the other causing the activated state to fluoresce [2, 3, 4]. In that case, the probability p that a molecule is in the activated state is, to a good approximation, independent of the (irreversible) bleaching rate β per activated molecule. In more recent approaches [14, 15, 16, 17, 18, 19], the switching between dark and activated states is controlled by the same wavelength that excites fluorescence from the activated molecules. Such approaches have the advantage of greater simplicity in hardware, at the expense of reduced latitude of control.

During a localization microscopy experiment, the number of fluorophores n in a region of size λ decreases over time if the fluorophores bleach irreversibly. (We distinguish irreversible bleaching, which damages the molecules and permanently renders them non-fluorescent, from the reversible bleaching that is used to temporarily switch fluorophores to dark states in some implementations [16].) The effect of irreversible bleaching is to change the tolerances for controlling the activation probability p per molecule: The need for non-overlapping bright spots dictates that, in a region of size λ , on average no more than 1 of the n fluorophores should be activated, so $p(t)$ must be less than or equal to $1/n(t)$. Due to irreversible bleaching, $n(t)$ is a decreasing function of time, and so $p(t)$ can be an increasing function of time. The result of bleaching is thus to enable faster acquisition: At later times, the activation probability can increase, decreasing the probability that no molecules will be on at any given time.

For an experimenter seeking a "snapshot" image of a short-lived structural feature, this speed effect of bleaching may be advantageous and hence worth optimizing. Conversely, for studies of long-term dynamics (in which case bleaching limits the durations of processes that can be studied), quantitative studies of image acquisition in the presence of bleaching are necessary to at least minimize the negative effects of bleaching. Even for fluorophores that can go through a very large number activation/deactivation cycles before irreversibly bleaching [20], maximizing the number of usable (*i.e.* single-molecule) images obtained is still desirable if monitoring small structures during a very long process.

Given that bleaching affects the tolerances on the activation probability per molecule and thus the image acquisition rate, the question that we study here is how to control $I(t)$ (and hence the activation probability p and bleaching rate β) to optimize the portion of the time in which exactly 1 of the n fluorophores is activated. We previously showed that if β and p can be varied independently (the 2-wavelength case) then the number of single-fluorophore images is maximized by varying the activation probability in such a way that the number of molecules decreases as a linear function of time: $n(t) = n(0) - \dot{n}t$, where the derivative \dot{n} is constant in time [21]. In this acquisition scheme, the 2-molecule error rate E_2 , defined as the ratio of the number of 2-molecule images obtained (and accepted by the analysis software) to the number of 1-molecule images obtained (and accepted by the analysis software) [22], is also constant. Deviations from the optimal scheme (which manifest as perturbations of the linear dependence of $n(t)$ on time) cause the number of single-molecule images to decrease. Interestingly, for fast acquisition (corresponding to larger p and E_2), deviations from the optimal scheme also decrease the number of 2-molecule images, partially mitigating the effects of a deviation on the ratio of 1-molecule to 2-molecule images.

In this work, we assume that the bleaching rate β and activation probability p are both functions of the same excitation intensity I , and are hence no longer independent quantities. We consider 4 related scenarios, based on plausible kinetic models of bleaching in switchable fluorophores used for superresolution microscopy. We show that the optimal data acquisition scheme depends very sensitively on the bleaching mechanism, highlighting the critical need for a detailed understanding of bleaching mechanisms in fluorophores used for superresolution localization microscopy. We will use a quasi-steady state model for the activation probabil-

ity, assuming that at any instant in time the probability of a fluorophore being in a given state depends only on the excitation intensity I and the rate constants for different upward and downward transitions. We will assume that the fluorophores are independent of each other (*i.e.* we are not considering processes such as Forster Resonance Energy Transfer).

Our focus will be on maximizing the number of single-fluorophore images while minimizing multi-fluorophore images. Recent work has shown that multi-fluorophore images can also be analyzed to obtain fluorophore position information [23], in which case one would want to maximize the number of images with m_{\max} or fewer activated fluorophores. While we do not consider this situation directly, we expect that many of the techniques developed here will carry over to the multi-fluorophore case, as one of the key results below (that in many cases the relevant integrals are stationary if the expected number of activated molecules per frame is kept constant) does not require the assumption that we only obtain information from single-fluorophore images.

1. Formalism and essential concepts

1.1. Activation probabilities

If we assume that the molecules are independent of each other, the probability of m molecules being simultaneously activated in a region of size λ is given by the binomial distribution:

$$p_m = \frac{n(n-1)\dots(n-m+1)}{m!} p^m (1-p)^{n-m} \quad (1)$$

where n is the number of molecules in a region of size λ . If the sample is labeled densely enough to resolve features of size $\lambda/10$ or smaller [7], then n will be greater than 100 in 2D, or 1000 in 3D. We can thus assume $n \gg 1$, which simplifies Eq. (1) considerably. The fractional error in approximating $n(n-1)\dots(n-m+1)$ as n^m is small for $n \gg 1$, so Eq. (1) becomes:

$$p_m = \frac{(np)^m}{m!} (1-p)^{n-m} \quad (2)$$

We can set a bound on the activation probability p and derive two useful results for this work, by invoking a result derived previously [22]:

$$p = \frac{E_2}{f_2/2f_1} \frac{1}{n} \quad (3)$$

where E_2 is the 2-molecule error rate discussed above. The parameter f_1 is the probability that the image analysis algorithm being used to process the data will correctly identify an image of a single-molecule and determine its position, while f_2 is the probability that the image analysis algorithm will correctly recognize 2-molecule overlaps as such and not analyze them. Consequently, p is bounded, and the upper bound decreases as n increases.

We also showed previously that maximizing the number of single-fluorophore images in a single cycle requires that p be less than $1/n$ [22]. Increasing p above this level actually decreases the number of 1-molecule images obtained (which can be shown by differentiating p_1 with respect to p in Eq. (1)) while increasing the number of 2-molecule images. The result is that there is a maximum error rate. In the case of non-bleaching fluorophores the maximum error rate is $f_2/2f_1$ [22], while in the case of bleachable fluorophores it is f_2/f_1 [21]

We will normalize the error rate to simplify our notation for p :

$$p = \frac{2f_1 E_2 / f_2}{n} \equiv \frac{\tilde{E}}{n} \quad (4)$$

where \tilde{E} is the normalized error rate $\tilde{E} = 2f_1 E_2 / f_2$. Note that in this notation, $np = \tilde{E}$.

With these results, it is possible to further simplify Eq. (2). Using $p = \tilde{E}/n$, and the identity $(1 + \frac{x}{n})^n \rightarrow e^x$ for large n and fixed x , we get:

$$p_m = \frac{\tilde{E}^m}{m!} \left(1 - \frac{\tilde{E}}{n}\right)^{n-m} = \frac{\tilde{E}^m}{m!} \frac{\left(1 - \frac{\tilde{E}}{n}\right)^n}{\left(1 - \frac{\tilde{E}}{n}\right)^m} \\ \rightarrow \frac{\tilde{E}^m}{m!} \frac{e^{-\tilde{E}}}{(1-0)^m} = \frac{\tilde{E}^m}{m!} e^{-\tilde{E}} \quad (5)$$

1.2. Expected times

We will be computing the expected amount of time in which exactly m molecules are activated. We thus consider the integral:

$$t_m = \int_0^{t_f} p_m(t) dt = \int_0^{t_f} \frac{(n(t)p(t))^m}{m!} e^{-n(t)p(t)} dt \quad (6)$$

If we wish to pick $p(t)$ in such a way to maximize this integral (for $m = 1$) or minimize it (for $m \neq 1$), we have a problem in variational calculus. The most commonly-used tools of variational calculus, the Euler-Lagrange equations [24, 25], require formulating the integral in terms of a time-dependent function and its first derivative, and then varying that function to make the integral an extremum. Note that while satisfaction of the Euler-Lagrange equations makes the integrals in Eq. (6) stationary, this is only a first-order condition that is satisfied by maxima, minima, and saddle points alike. Later, we will consider second-order conditions to determine when t_1 is maximized.

We will express t_m in terms of $n(t)$ and $\dot{n}(t)$. Physically, it may seem natural to pick $p(t)$ as the function to be varied, since that is the experimentally-controllable parameter. However, the Euler-Lagrange equations apply to problems that are formulated in terms of functions and their derivatives. As we show in the next section, if we have a kinetic model of the bleaching process we can formulate the problem in terms of $n(t)$ and $\dot{n}(t)$, and use the kinetic model to express $p(t)$ in terms of n and \dot{n} .

It is important to note that we are *not* trying to maximize the number of single-fluorophore images obtained in a single activation cycle. As discussed above, the number of single-fluorophore images in a given cycle is maximized when $\tilde{E} = 1$ [22]. Rather, we are trying to maximize the total number of single-fluorophore molecules imaged over a fixed time period (generally longer than a single cycle), subject to the constraint that a given number of molecules bleach in that time. The use of the Euler-Lagrange equations contains an implicit assumption that the numbers of unbleached molecules $n(t)$ at the beginning and end of the experiment are fixed. Given that constraint, we are trying to obtain as many single-fluorophore images as possible while bleaching a given number of molecules in a given time. However, a person following the prescriptions given below can pick the time interval and number of molecules bleached in that time interval (*i.e.* pick the constraints to impose) and then pick the appropriate error rate to bleach the designated number of molecules in the designated time.

2. Bleaching models

Because single-wavelength superresolution methods are a rapidly evolving area, we will consider several different models that might describe plausible methods and bleaching mechanisms for different fluorophores. In each case, we can write down a kinetic model of one of the following forms:

$$\dot{n} = -\beta(I(t))n(t) \begin{cases} p(I(t)) \\ 1 - p(I(t)) \end{cases} \quad (7)$$

where the first case corresponds to a bleaching mechanism with a rate proportional to the occupation probability of the activated state, and the second case corresponds to a bleaching mechanism proportional to the occupation probability for the dark state. The parameter $\beta(I(t))$ is the intensity-dependent rate at which molecules bleach. In either case, we can divide both sides by $n(t)$ and get:

$$-\dot{n}/n = \beta(I(t)) \begin{cases} p(I(t)) \\ 1 - p(I(t)) \end{cases} \quad (8)$$

Because the right hand side depends only on the intensity I in either case, it follows that I can be expressed as a function of $-\dot{n}/n$, *i.e.* there is a one-to-one relationship between the bleaching rate per molecule and the excitation intensity I . Therefore, p is also a function of $-\dot{n}/n$. We can thus write our integrals as:

$$t_m = \int_0^{t_f} \frac{(np(-\dot{n}/n))^m}{m!} e^{-np(-\dot{n}/n)} dt \quad (9)$$

Once we have determined the form of $p(-\dot{n}/n)$ via a model of the bleaching process, we use the Euler-Lagrange equations to obtain a differential equation for n . Our procedure is therefore:

1. Using a model of bleaching kinetics, express I in terms of $-\dot{n}/n$.
2. Using a model of the activation process, express $p(I)$ in terms of $-\dot{n}/n$.
3. From the relationship between p and $-\dot{n}/n$, express the integrands p_m in terms of n and $-\dot{n}/n$.
4. Obtain a differential equation for n via the Euler-Lagrange equations.

We now consider four cases:

2.1. Excitation from the dark to activated state, followed by bleaching from the excited state

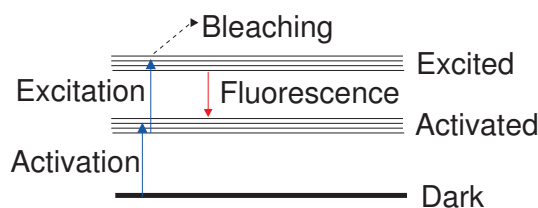


Fig. 1. Schematic of states and transitions for a fluorophore in which the dark state is the default state and bleaching occurs from the activated state. We assume multiple vibrational sublevels in the activated and excited states, to account for Stokes shifts of the absorption and emission spectra. The bleaching process depicted occurs from the excited state, and is assumed to not require the absorption of an additional molecule from the excited state.

Our first case is similar to initial superresolution experiments, in which the default state of a molecule is the dark state and light is needed to raise the molecule to the activated state. A schematic is given in Fig. 1. While such fluorophores are generally controlled with multiple wavelengths in experiments, in principle a single wavelength could be used (for simplicity) if a fluorophore has strongly overlapping activation and excitation bands. The existence of spontaneous activation [26, 27] when only the longer-wavelength (excitation) beam is turned

on suggests that control via a single wavelength may be feasible in some cases. We therefore assume, initially, that the activation probability is given by:

$$p = \frac{I}{1+I} \quad (10)$$

Note that throughout this work we will be measuring I in units of a saturation intensity chosen so that when $I = 1$ the probability of being in the higher state is $1/2$. Eq. (10) can be derived by setting the rate of upward transitions (proportional to I and $(1-p)$) equal to the rate of downward transitions (proportional to p).

It is important to not take Fig. 1 too literally. It is a schematic, and the key point is a sequence of steps: dark \rightarrow activated \rightarrow repeated excitation and fluorescence \rightarrow eventual deactivation or bleaching. We make no assumptions about short-lived or transient intermediate steps; our key assumptions are that (1) these processes have reached a steady state and (2) upward transitions proceed at a rate proportional to the excitation intensity.

After activation, bleaching requires the absorption of a second photon (to go to a more reactive excited state, which may be either the fluorescent single state or a long-lived triplet state), with a rate that is proportional to the intensity, so our kinetic model is:

$$-\dot{n}/n = \frac{k_b I^2}{1+I} \quad (11)$$

This problem is easily solved in the experimentally relevant limit that $n \gg 1$, in which case $p \ll 1$, meaning that the I term in the denominator is negligible. We then have the following results:

$$I = \sqrt{\frac{-\dot{n}}{k_b n}} \quad (12)$$

$$np \approx nI = \sqrt{-\frac{\dot{n}n}{k_b}} \equiv c_1 n \cdot (-\dot{n}/n)^{\frac{1}{2}} \quad (13)$$

where $c_1 \equiv \sqrt{1/k_b}$. We will now show that when p is a power law $p = c(-\dot{n}/n)^a$, it follows that $\tilde{E} = np$ is a constant if $n(t)$ is chosen to make the integrals in Eq. (6) stationary.

The Euler-Lagrange equations that must be satisfied to make the integrals in Eq. (6) stationary are:

$$\frac{d}{dt}\Pi = \frac{\partial}{\partial n} p_m \quad (14)$$

$$\Pi = \frac{\partial}{\partial \dot{n}} p_m \quad (15)$$

In the terminology of classical mechanics, Π is a momentum, and p_m is our Lagrangian. In the case where p is a power law, we get:

$$\Pi = p'_m(nc \cdot (-\dot{n}/n)^a) \frac{\partial}{\partial \dot{n}} nc \cdot (-\dot{n}/n)^a = -p'_m(nc \cdot (-\dot{n}/n)^a) ac \cdot (-\dot{n}/n)^{a-1} \quad (16)$$

where p'_m is the derivative of p_m with respect to its argument np .

Rather than using our result in Eq. (16) to derive the Euler-Lagrange equations and then solve them, we will instead use an approach analogous to energy conservation in classical mechanics: Because the Lagrangian in Eq. (6) has no explicit time-dependence (*i.e.* the time-dependence

of p_m is solely due to the time-dependence of n and \dot{n} , if we pick $n(t)$ to satisfy the Euler-Lagrange equations then the Hamiltonian H will be a constant (*i.e.* time-independent)[24]:

$$\begin{aligned} H &= \dot{n}\Pi - p_m(nc \cdot (-\dot{n}/n)^a) \\ &= ac \frac{(-\dot{n})^a}{n^{a-1}} p'_m \left(c \frac{(-\dot{n})^a}{n^{a-1}} \right) - p_m \left(c \frac{(-\dot{n})^a}{n^{a-1}} \right) \end{aligned} \quad (17)$$

Because H is time-independent and is a function of a single argument $(-\dot{n})^a/n^{a-1}$, it therefore follows that its argument $(-\dot{n})^a/n^{a-1} = \tilde{E}/c$ is also time-independent, and hence \tilde{E} is a constant, even as n and p change.

The requirement of a constant error rate gives us a simple differential equation to solve:

$$\dot{n} = - \left(\frac{\tilde{E}}{c} \right)^{\frac{1}{a}} n^{1-\frac{1}{a}} \quad (18)$$

For the case considered here, where $a = 1/2$ and $c = c_1 = \sqrt{1/k_b}$, the differential equation can be written as:

$$\begin{aligned} \dot{n} &= - \left(\frac{\tilde{E}}{1/\sqrt{k_b}} \right)^{\frac{1}{1/2}} n^{1-\frac{1}{1/2}} = k_b \tilde{E}^2 / n \\ n\dot{n} &= -k_b \tilde{E}^2 = \frac{1}{2} \frac{d}{dt} n^2 \end{aligned} \quad (19)$$

with solution:

$$n(t) = \sqrt{n^2(0) - 2\tilde{E}^2 k_b t} \quad (20)$$

Note that, as in our previous work [21], a higher error rate causes a faster decline in the number of unbleached molecules. However, in this case the dependence on time involves the square root of time rather than a linear function of time.

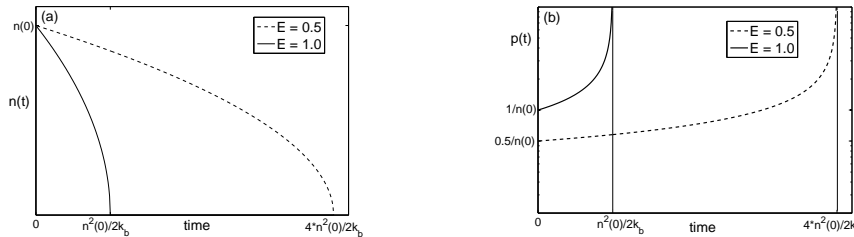


Fig. 2. (a) $n(t)$ and (b) $p(t)$ for acquisition at different constant error rates, under scenario 1.

The time-dependent activation probability and illumination intensity are easy to obtain. Because this is a constant error rate scheme, $p(t) = \tilde{E}/n(t)$ (from Eq. (4)), and from Eq. (10) we know that (for small I) $p = I$, so we get:

$$I(t) \approx p(t) = \frac{\tilde{E}}{n(t)} = \frac{\tilde{E}}{\sqrt{n^2(0) - 2\tilde{E}^2 k_b t}} \quad (21)$$

Results for $n(t)$ and $p(t)$ are shown in Fig. 2.

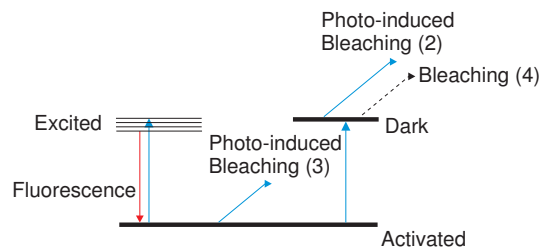


Fig. 3. Schematic of a fluorophore whose default state is activated (*i.e.* fluorescent). Three plausible bleaching pathways are illustrated, numbered in the order in which they are considered here. Blue upward arrows indicate absorption of a photon, solid diagonal lines indicate bleaching upon the absorption of an additional photon, and diagonal dashed lines indicate bleaching without the absorption of an additional photon.

2.2. Excitation from the activated state to the dark state, followed by photo-induced bleaching

In many single-wavelength superresolution experiments, the ground state of a molecule is actually *not* the dark state; the dark state is reached by the absorption of a photon [14, 15, 16]. Typically, this dark state is a long-lived triplet state. A schematic of this process is shown in Fig. 3. We assume that the rate of transitions from the activated state to the dark state is proportional to the illumination intensity I and the occupation probability p for the activated state, while the rate of transitions from the dark state to the activated state is proportional to $1 - p$ (the probability of being in the dark state). By setting the dark state probability equal to $1 - p$ we are implicitly assuming that fluorophores spend a negligible amount of time in the excited state. This assumption is valid if the typical fluorophore yields of order 10^3 photons per second (a common number in superresolution experiments, *e.g.* [3, 28, 14, 19] and has an excited state lifetime of order 10^{-9} seconds, for a total excited state time of order 10^{-6} seconds, while the time in the dark state is of order milliseconds to tens of milliseconds [14, 15]. Putting these assumptions together, we can do some algebra to get the following for the activated state probability:

$$p = \frac{1}{1 + I} \quad (22)$$

where I is again normalized so that the probability of being in the higher-energy dark state is $1/2$ when $I = 1$.

We obtain an expression of the same form if we assume that the dark state is reached by first passing through the excited state (*e.g.* a transition from a single ground state S_0 to a first singlet excited state S_1 , from which some fraction of the molecules are transferred to a triplet state T_1). Because a variety of microscopic models give the same result, it is important to not take Fig. 3 too literally; it is a schematic illustrating that upon absorption of a photon the molecule can either go to a state from which it will fluoresce and return to the ground state (called “activated” here for convenience), or a long-lived state from which it will not fluoresce. The key assumptions are that the dark state is longer-lived than the state producing fluorescence, and that it is reached via photon absorption from the ground state (which we call “activated” here).

While bleaching mechanisms in different fluorophores are an area of continued investigation, if the dominant bleaching process occurs from the dark state and is induced by the absorption of a second photon [29, 30](pathway 2 in Fig. 3), the bleaching rate per molecule is given by:

$$-\dot{n}/n = k_b(1 - p) = \frac{k_b I^2}{1 + I} \quad (23)$$

If the activation probability per molecule is assumed to be less than $1/n$ for $n \gg 1$, it follows that $p \ll 1$ and so $I \gg 1$. We get the following relationship between $-\dot{n}/n$ and I :

$$I = -\frac{\dot{n}}{k_b n} \quad (24)$$

The activation probability is then:

$$p \approx 1/I = -\frac{k_b n}{\dot{n}} \quad (25)$$

This is again a power law in $-\dot{n}/n$ with exponent -1 , and so it follows that \tilde{E} is again constant. Our differential equation is:

$$\frac{1}{np} = -\frac{\dot{n}}{k_b n^2} = \frac{1}{k_b} \frac{d}{dt} \frac{1}{n} = \frac{1}{\tilde{E}} \quad (26)$$

with solution:

$$n(t) = \frac{n(0)}{1 + n(0)k_b t / \tilde{E}} \quad (27)$$

Note that in this case, lower error rates actually cause the number of molecules to deplete more rapidly. This is because achieving a low error rate requires a high excitation intensity to place more fluorophores in the dark state. At the same time, increasing the intensity increases the rate at which dark molecules are bleached as well as the number of molecules that are in the dark state and hence available to be bleached.

Given $n(t)$, it is again straightforward to determine $p(t)$ and $I(t)$. For a constant error rate, $p = \tilde{E}/n$ (Eq. (4)), and for this energy level scheme $p = 1/I$, so we get:

$$p(t) = \frac{\tilde{E}}{n(0)} + k_b t \quad (28)$$

$$I(t) = \frac{n(0)}{\tilde{E} + n(0)k_b t} \quad (29)$$

2.3. Photo-induced bleaching from the activated state

Next, let us suppose that bleaching can only happen if activated molecules absorb a photon, at a rate proportional to the excitation intensity. (Pathway 3 in Fig. 3) The bleaching rate per molecule is given by:

$$-\dot{n}/n = k_b I p = \frac{k_b I}{1 + I} \quad (30)$$

In this case, $I = -\dot{n}/(\dot{n} + k_b n)$, so $p = 1/(1 + I) = 1 + \dot{n}/k_b n$ and we get the following for the error rate:

$$\tilde{E} = np = n + \dot{n}/k_b \quad (31)$$

Before we solve this model, we will examine one more case, and show that it is equivalent.

2.4. Bleaching from the dark state without the absorption of a second photon

Alternatively, let us consider the case where the dark state is reached via absorption of a photon, and bleaching occurs from the dark state without the absorption of a second photon. (Pathway 4 in Fig. 3) Such a scenario would correspond to a first order bleaching process, in which a

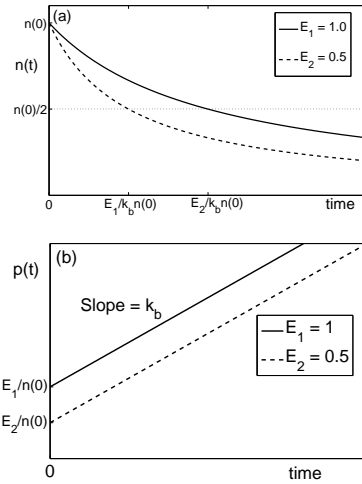


Fig. 4. (a) $n(t)$ and (b) $p(t)$ for acquisition at constant error rate, under scenario 2. The time at which the number of molecules has decreased by half is shown for each plot in (a).

molecule in the dark state has a constant probability per unit time of undergoing an irreversible bleaching reaction. In this case, the bleaching rate is proportional to the probability of being in the dark state:

$$-\dot{n}/n = k_b(1 - p) = k_b \frac{I}{1 + I} \quad (32)$$

We can solve Eq. (32) for I in terms of \dot{n}/n , and get $I = -\dot{n}/(k_b n + \dot{n})$. This gives $p = 1/(1 + I) = 1 + \dot{n}/k_b n$, so we again have for \tilde{E} :

$$\tilde{E} = np = n + \dot{n}/k_b \quad (33)$$

Interestingly, in this case the error rate is not constant. We can show, however, that it is a decreasing function of time. To see this, we need to use the Euler-Lagrange equations and some properties of our Lagrangian. The momentum Π for this case is:

$$\Pi = \frac{\partial}{\partial \dot{n}} p_m(n + \dot{n}/k_b) = \frac{1}{k_b} p'_m(n + \dot{n}/k_b) \quad (34)$$

where p'_m is evaluated with respect to its argument $n + \dot{n}/k_b$. The time derivative of Π is:

$$\frac{d}{dt} p'_m/k_b = (\dot{n} + \ddot{n}/k_b) \cdot p''_m/k_b = \frac{\partial}{\partial n} L(n + \dot{n}/k_b) = p'_m \quad (35)$$

Note that $\dot{n} + \ddot{n}/k_b$ is just the time derivative of \tilde{E} , so we get that:

$$\dot{\tilde{E}} = k_b p'_m / p''_m \quad (36)$$

To go further, we will assume that our Lagrangian is p_1 (given in Eq. (6) as $\tilde{E}e^{-\tilde{E}}$), *i.e.* we are trying to maximize the number of single-molecule images. The derivatives of p_1 are:

$$p'_1 = (1 - \tilde{E})e^{-\tilde{E}} \quad (37a)$$

$$p''_1 = -(2 - \tilde{E})e^{-\tilde{E}} \quad (37b)$$

The time derivative of \tilde{E} is then:

$$\dot{\tilde{E}} = -k_b \frac{1 - \tilde{E}}{2 - \tilde{E}} \quad (38)$$

This differential equation has an unstable fixed point at $\tilde{E} = 1$, and a singularity at $\tilde{E} = 2$. The most interesting cases for our purposes are initial error rates less than 1, for which \tilde{E} decreases as a function of time.

We can get the time-dependence of \tilde{E} from Eq. (38), which can be solved analytically:

$$\int_{\tilde{E}(0)}^{\tilde{E}(t)} \left(1 + \frac{1}{1 - \tilde{E}} \right) d\tilde{E} = - \int_0^t k_b dt'$$

$$\tilde{E}(t) - \tilde{E}(0) + \ln \frac{1 - \tilde{E}(0)}{1 - \tilde{E}(t)} = -k_b t \quad (39)$$

Because the initial conditions show up additively with the time, changing the initial condition merely shifts the plot in time. Also, \tilde{E} reaches 0 at a finite time $t_f = (\tilde{E}(0) - \log 1 - \tilde{E}(0))/k_b$, which increases as $\tilde{E}(0)$ increases. Solutions of Eq. (39) are plotted for different initial errors

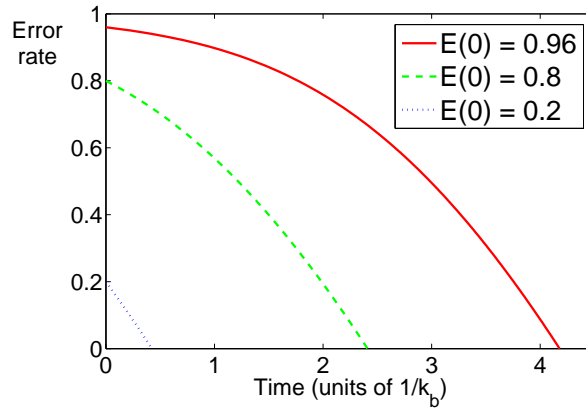


Fig. 5. Solution to Eq. (38) for different initial error rates.

in Fig. 5.

Once we have $\tilde{E}(t)$, we can solve for $n(t)$ using $\tilde{E} = n + \dot{n}/k_b$. Because $\tilde{E} < 1$ and $n \gg 1$, the time dependence of $n(t)$ is, to an excellent approximation, an exponential decay with rate k_b . The difference between \dot{n} and $-k_b n$ is very small. Fortunately, however, the quantity that needs to be controlled with high precision is $I(t)$, not n or \dot{n} . Also, because $I \gg 1$, there is considerable latitude in the control of I .

To obtain $I(t)$, we recall that $p = 1/(1 + I)$, and solving Eq. (32) gave $I = -\dot{n}/(k_b n + \dot{n}) = -\dot{n}/k_b \tilde{E}$. The time-dependence of n is approximately $n(0)e^{-k_b t}$, so $-\dot{n} = k_b n(0)e^{-k_b t}$, and we get the following for $I(t)$:

$$I(t) = n(0)e^{-k_b t} / \tilde{E}(t) \quad (40)$$

Solutions to Eq. (40) are plotted in Fig. 6. Because large relative changes in $I(t)$ are required to obtain the optimal scheme, the excitation intensity does not need to be finely-tuned. We show $I(t)$ for 2 pairs of initial error rates, each pair differing by 10%. In each case, the intensity vs. time graphs differ by approximately 10% initially, and the percentage difference in I increases substantially over time. We thus conclude that the optimal acquisition scheme is achievable without delicate fine-tuning. This issue of robustness is further explored in the next section.

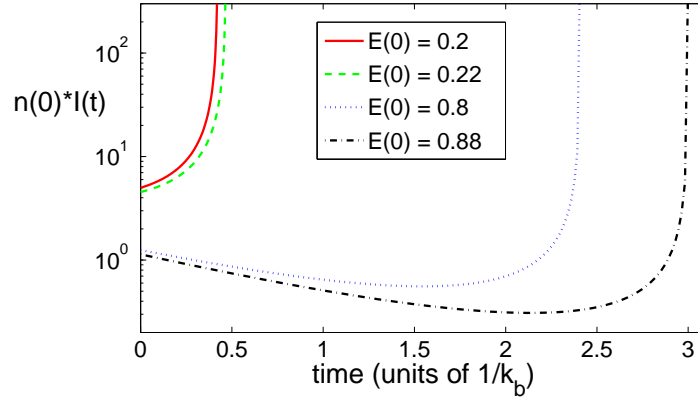


Fig. 6. Excitation intensity $I(t)$ for different initial error rates \tilde{E} .

3. Second-order conditions and robustness

The Euler-Lagrange equations that we solved above are only first order conditions. Maxima, minima, and saddle points are distinguished by second-order conditions. Because a variational calculus problem is a calculus problem in infinite dimensions, it is often difficult to formulate necessary second-order conditions without producing an infinite set of equations (one for each direction in function space). However, there are *sufficient* second-order conditions that are straightforward to apply: If the Lagrangian p_m is everywhere a convex function (second derivative non-negative) of its inputs n and \dot{n} then the integral of p_m is minimized when n is chosen to satisfy the Euler-Lagrange equations [25]. Conversely, if p_m is everywhere a concave function (second derivative non-positive) of n and \dot{n} then the integral of p_m is maximized when n is chosen to satisfy the Euler-Lagrange equations.

3.1. Constant Error Rate Schemes

In the constant error rate scenarios considered, the Lagrangians are of the form:

$$p_m = \frac{\tilde{E}^m}{m!} e^{-\tilde{E}} = \frac{c(-\dot{n})^{am}}{m!n^{m(a-1)}} e^{\frac{c(-\dot{n})^a}{n^{a-1}}} \quad (41)$$

with $a = 1/2$ (section 2.1), -1 (section 2.2), and 1 (previous work [21]). In what follows, we choose our units of time so that $c = 1$. The second derivatives of p_m are:

$$\frac{\partial^2 p_m}{\partial \dot{n}^2} = a \frac{\tilde{E}^m}{m! \dot{n}^2} e^{\tilde{E}} (a\tilde{E}^2 + \tilde{E}(1-a-2am) + m(am-1)) \quad (42a)$$

$$\frac{\partial^2 p_m}{\partial n^2} = \frac{a-1}{m!n^2} \tilde{E}^m e^{\tilde{E}} ((a-1)\tilde{E}^2 + \tilde{E}(2m-2am-a) + (a-1)m^2 + m) \quad (42b)$$

We can use these expressions to determine which integrals are maximized or minimized for small \tilde{E} .

The second derivatives change sign when they are equal to zero, which occurs when:

$$\tilde{E} = \frac{-1+a+2am \pm \sqrt{(a-1)^2 + 4a^2m}}{2a} \quad (\text{setting } \frac{\partial^2 p_m}{\partial \dot{n}^2} = 0) \quad (43a)$$

$$\tilde{E} = \frac{2m(a-1) + a \pm \sqrt{4m(a-1)^2 + a^2}}{2(a-1)} \quad (\text{setting } \frac{\partial^2 p_m}{\partial n^2} = 0) \quad (43b)$$

Our results for the second derivatives are summarized below in Tables 1 and 2

3.1.1. The $a = \frac{1}{2}$ case

Table 1. Summary of second derivatives of integrands for the constant error rate scheme described in section 2.1.

| t_m | Second derivative for small \tilde{E} : | Sign change: | Comment |
|----------|---|--|---------------------------------|
| t_0 | positive | none | Always minimized. |
| t_1 | negative | $\tilde{E} = 1.62$ | Maximized for small \tilde{E} |
| t_2 | negative | $\tilde{E} = 3$ | Maximized for small \tilde{E} |
| t_{3+} | positive | at value of \tilde{E} that increases with m (0.697 for $m = 3$) | Minimized for small \tilde{E} |

In this case, the integrands p_m are functions of $\sqrt{m\tilde{n}}$, which is symmetric under exchange of n and \tilde{n} . It is thus only necessary to calculate second derivatives with respect to one of those variables, rather than both.

We find that t_0 satisfies sufficient conditions for a minimum for all \tilde{E} ; deviations from a constant error rate scheme will increase the number of zero-fluorophore images. This is consistent with our previous findings for two-wavelength acquisition schemes [21]. Likewise, for $m \geq 3$, t_m is minimized for small error rates. This is exactly what we'd expect from an optimal acquisition scheme. It is also not surprising that t_1 is maximized for $\tilde{E} < 1.62$, consistent with our goal of getting as many single-fluorophore images as possible.

It may seem unfortunate that t_2 is *also* maximized for small \tilde{E} . However, consider the effects of deviations from the optimal scheme: If t_1 and t_2 are both maximized, then deviations reduce the number of 1-fluorophore images and also the number of 2-fluorophore overlap images. The loss of 2-fluorophore images partially compensates for the loss of 1-fluorophore images, mitigating the effect on the 2-fluorophore error rate (which is the ratio of 2-molecule images to single-molecule images). This is hence a robust acquisition scheme.

One might wonder whether it would then be even more advantageous to also maximize the number of 3-fluorophore images, 4-fluorophore images, etc. However, when one deviates from the optimal scheme, reducing the number of images with 3 or more activated fluorophores is less important than reducing the number of images with 2 activated and overlapping fluorophores, for two reasons. First, for small \tilde{E} , the 2-fluorophore images are more common than images with more activated fluorophores, because for small \tilde{E} $p_2 > p_m$ ($m \geq 3$). Second, the 2-fluorophore images are, in general, more difficult to identify and reject than images with 3 or more fluorophores: 2-fluorophore images generally have fewer photons than images with 3 or more fluorophores, and are larger in cross-section. Also, 2-fluorophore images likely to be only slightly elliptical, while images with more activated fluorophores are more likely to have irregular and large shapes that are easier to identify. Thus, when one deviates from the optimal scheme it is most important that the number of 2-fluorophore images be reduced along with the number of 1-fluorophore images. We therefore conclude that acquisition is optimized for $\tilde{E} < 1.62$ in this scenario.

3.1.2. The $a = -1$ case

In this case, zero-fluorophore images are actually *maximized* for small error rates ($\tilde{E} < 0.5$), while single-fluorophore and multi-fluorophore images are minimized. Specifically, t_1 satisfies sufficient conditions for a minimum for $\tilde{E} < 0.219$. Acquisition at constant error rate can only be considered optimal for higher error rates ($\tilde{E} > 0.586$), in which case the integrand satisfies

Table 2. Summary of second derivatives of integrands for the constant error rate scheme described in section 2.2.

| t_m | $\partial^2 p_m / \partial \dot{n}^2$ | | $\partial^2 p_m / \partial n^2$ | | Comment |
|----------|---------------------------------------|---|---------------------------------|---|---|
| | Small \tilde{E} : | Sign change: | Small \tilde{E} : | Sign change: | |
| t_0 | negative | $\tilde{E} = 2$ | negative | $\tilde{E} = 0.5$ | Minimized if $\tilde{E} > 2$, maximized if $\tilde{E} < 1/2$ |
| t_1 | positive | $\tilde{E} = 0.586$ | positive | $\tilde{E} = 0.219$ | Maximized if $\tilde{E} > 0.586$, minimized if $\tilde{E} < 0.219$ |
| t_{2+} | positive | at value of \tilde{E} that increases with m (1.27 for $m = 2$) | positive | at value of \tilde{E} that increases with m (0.81 for $m = 2$) | Minimum for small \tilde{E} |

sufficient conditions for maximizing t_1 . At intermediate error rates it is difficult to say whether t_1 is a minimum, maximum, or saddle point for the constant error rate acquisition scheme. If one wishes to maximize t_2 to make the constant-error scheme more robust, as discussed above, it is necessary to work at $\tilde{E} = 1.27$ (a very large \tilde{E} value).

3.2. The exponential case

In cases 3 and 4 from Fig. 3, we found that $n(t)$ decays approximately exponentially in the optimal acquisition scheme. In both of these cases, $\tilde{E} = n + \dot{n}/k_b$, so the second derivatives of p_m with respect to n and \dot{n} have the same form (up to a factor of $1/k_b^2$) and the concavity or convexity of the integrand is easy to determine. The quantity that we need to consider is:

$$p_m''(n + \dot{n}/k_b) = \frac{d^2}{d\tilde{E}^2} \tilde{E}^m e^{-\tilde{E}} = \frac{\tilde{E}^2 - 2m\tilde{E} + m^2 - m}{m!} e^{-\tilde{E}} \quad (44)$$

For $m = 0$, the right side of Eq. (44) is positive for any \tilde{E} , so the optimal acquisition scheme minimizes the number of zero-fluorophore frames for any value of the error rate. For $m = 1$, the right side of Eq. (44) is negative as long as $\tilde{E} < 2$, which means that even for very high initial error rates the number of single-fluorophore frames is maximized. Since we established in Eq. (38) that the error rate decreases monotonically if $\tilde{E} < 1$, it follows that the bound on the error rate set by the requirement of a decreasing error rate is stronger than the bound set by the second order conditions.

For $m \geq 2$, the righthand side of Eq. (44) is always positive at $\tilde{E} = 0$ and has zeros at $\tilde{E} = m \pm \sqrt{m}$. One consequence is that for $\tilde{E} < 1$ the number of images with 3 or more activated fluorophores is always minimized. The case of $m = 2$ is interesting: The second derivative of p_2 is positive for $\tilde{E} < 0.587$ and negative for $0.587 < \tilde{E} < 2.414$, so that for $\tilde{E} < 0.587$ the number of 2-fluorophore images is minimized, while for larger \tilde{E} the number of 2-fluorophore images is maximized.

As discussed above for acquisition at constant error rate, maximizing the number of 2-fluorophore images along with the number of 1-fluorophore images makes the scheme more robust. The key difference between this case and the $a = 1/2$ case is that acquisition here is actually optimized at larger error rates $\tilde{E} > 0.587$, while in the other case acquisition is optimized for all $\tilde{E} < 1.62$. While working at higher error rates might seem problematic, if one uses good rejection algorithms to remove multi-fluorophore images, a large normalized error rate \tilde{E} can still correspond to a small absolute error rate $E_2 = 2f_2\tilde{E}/f_1$.

4. Conclusions

A major goal in any superresolution localization microscopy experiment is to maximize the number of single-fluorophore images obtained. When the activation process is controlled by the same light source as the bleaching process, it is necessary to balance bleaching effects (which reduce the number of usable fluorophores but also reduce the probability of a nearby fluorophore emitting light concurrent with the fluorophore of interest) against activation effects (which determine the relative probabilities of obtaining single-fluorophore and multi-fluorophore images). Some of the details of the bleaching process therefore have significant effects on the optimal acquisition scheme. While short-lived intermediate states do not affect our results, the following aspects of the activation and bleaching kinetics are of critical importance:

1. Whether bleaching occurs from the dark or activated state
2. Whether bleaching requires the absorption of an additional photon after excitation
3. Whether the activated state is a default state or is reached via absorption of a photon.

We have analyzed 4 plausible models of the bleaching process in different superresolution localization microscopy experiments, and have shown that in each case the optimal acquisition scheme either involves acquisition at constant error rate or with a decreasing error rate. In each case, only two numbers must be known to implement the optimal scheme: a saturation intensity and a bleaching rate constant. In addition, we have shown that the robustness of the scheme (and whether substantial robustness is achieved at low error rates or high error rates) also depends on the details of the bleaching process. Finally, although new fluorophores are being rapidly developed for use in localization microscopy, our methods are general, and can be used to investigate almost any bleaching and activation process controlled by a single wavelength, as well as to predict optimal acquisition schemes for techniques that extract information from multi-fluorophore images [23].

Acknowledgments

This work was supported by a seed award from the California State University Program for Education and Research in Biotechnology (CSUPERB) and a Teacher Scholar Award from California State Polytechnic University. We thank Tijana Jovanovic-Talisan for very useful comments on the manuscript, and Kai S. Lam for useful conversations about the variational approach. We also thank a reviewer of a previous paper for urging us to take up the single-wavelength case.