

A divide and conquer strategy for the maximum likelihood localization of low intensity objects

Alexander Krull,^{1,2,*} André Steinborn,³ Vaishnavi Ananthanarayanan,¹ Damien Ramunno-Johnson,¹ Uwe Petersohn,³ and Iva M. Tolić-Nørrelykke¹

¹Max Planck Institute of Molecular Cell Biology and Genetics, 01307 Dresden, Germany

²Dresden University of Technology, Computer Graphics Institute, 01062 Dresden, Germany

³Dresden University of Technology, Artificial Intelligence Institute, 01062 Dresden, Germany

*alexander.krull@tu-dresden.de

Abstract: In cell biology and other fields the automatic accurate localization of sub-resolution objects in images is an important tool. The signal is often corrupted by multiple forms of noise, including excess noise resulting from the amplification by an electron multiplying charge-coupled device (EMCCD). Here we present our novel Nested Maximum Likelihood Algorithm (NMLA), which solves the problem of localizing multiple overlapping emitters in a setting affected by excess noise, by repeatedly solving the task of independent localization for single emitters in an excess noise-free system. NMLA dramatically improves scalability and robustness, when compared to a general purpose optimization technique. Our method was successfully applied for *in vivo* localization of fluorescent proteins.

© 2014 Optical Society of America

OCIS codes: (150.1135) Algorithms (Machine vision); (100.2960) Image analysis; (040.3780) Low light level; (110.0180) Microscopy; (170.2520) Fluorescence microscopy.

References and links

1. M. Coelho, N. Maghelli, and I. M. Tolić-Nørrelykke, "Single-molecule imaging *in vivo*: the dancing building blocks of the cell," *Integr. Biol. (Camb)* **5**, 748–758 (2013).
2. K. Irie, A. E. McKinnon, K. Unsworth, and I. M. Woodhead, "A model for measurement of noise in CCD digital-video cameras," *Meas. Sci. Technol.* **19**, 045207 (2008).
3. D. J. Denvir and E. Conroy, "Electron-multiplying CCD: the new ICCD, in *Low-Light-Level and Real-Time Imaging Systems, Components, and Applications*", Proc. SPIE **4796**, pp. 164–174 (2003).
4. B. C. Carter, G. T. Shubeita, and S. P. Gross, "Tracking single particles: a user-friendly quantitative evaluation," *Phys. Biol.* **2**, 60–72 (2005).
5. G. M. Lee, A. Ishihara, and K. A. Jacobson, "Direct observation of brownian motion of lipids in a membrane," *Proc. Natl. Acad. Sci. USA* **88**, 6274–6278 (1991).
6. R. N. Ghosh and W. W. Webb, "Automated detection and tracking of individual and clustered cell surface low density lipoprotein receptor molecules," *Biophys. J.* **66**, 1301–1318 (1994).
7. C. M. Anderson, G. N. Georgiou, I. E. Morrison, G. V. Stevenson, and R. J. Cherry, "Tracking of cell surface receptors by fluorescence digital imaging microscopy using a charge-coupled device camera. low-density lipoprotein and influenza virus receptor mobility at 4 degrees c," *J. Cell Sci.* **101** (Pt 2), 415–425 (1992).
8. G. J. Schutz, H. Schindler, and T. Schmidt, "Single-molecule microscopy on model membranes reveals anomalous diffusion," *Biophys. J.* **73**, 1073–1080 (1997).
9. R. J. Ober, S. Ram, and E. S. Ward, "Localization accuracy in single-molecule microscopy," *Biophys. J.* **86**, 1185–1200 (2004).
10. 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).

11. 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).
12. Y. Wang, T. Quan, S. Zeng, and Z.-L. Huang, "PALMER: a method capable of parallel localization of multiple emitters for high-density localization microscopy," *Opt. Express* **20**, 16039–16049 (2012).
13. K. I. Mortensen, L. S. Churchman, J. A. Spudich, and H. Flyvbjerg, "Optimized localization analysis for single-molecule tracking and super-resolution microscopy," *Nat. Methods* **7**, 377–381 (2010).
14. J. Chao, E. S. Ward, and R. J. Ober, "Fisher information matrix for branching processes with application to electron-multiplying charge-coupled devices," *Multidimens. Syst. Signal Process.* **23**, 349–379 (2012).
15. C. H. A. G. Basden, "Photon counting strategies with low light level CCDs," *Monthly Notices of the Royal Astron. Soc.* **345**, 985–991 (2003).
16. J. Chao, E. S. Ward, and R. J. Ober, "Localization accuracy in single molecule microscopy using electron-multiplying charge-coupled device cameras, in *Three-Dimensional and Multidimensional Microscopy: Image Acquisition and Processing XIX*," *Proc. SPIE* **8227**, p. 82271P (2012).
17. A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. of the Royal Stat. Soc., B* **39**, 1–38 (1977).
18. C. Rao, *Linear Statistical Inference and its Applications* (Wiley, 1975).
19. M. J. D. Powell, "An efficient method for finding the minimum of a function of several variables without calculating derivatives," *Comput. J.* **7**, 155–162 (1964).
20. J. Chao, S. Ram, E. S. Ward, and R. J. Ober, "Ultrahigh accuracy imaging modality for super-localization microscopy," *Nat. Methods* **10**, 335–338 (2013).
21. A. Steinborn, S. Taut, V. Brendler, G. Geipel, and B. Flach, "TRLFS: analysing spectra with an expectation-maximization (EM) algorithm," *Spectrochim Acta A Mol Biomol Spectrosc.* **71**, 1425–1432 (2008).
22. M. I. Schlesinger and V. Hlavac, *Ten Lectures on Statistical and Structural Pattern Recognition*, (Kluwer Academic Publishers, 2002) vol. 24 of *Computational Imaging and Vision*.
23. S. Stallinga and B. Rieger, "The effect of background on localization uncertainty in single emitter imaging," in *Proceedings of 9th IEEE International Symposium on Biomedical Imaging* (2012), pp. 988–991.
24. L. Xu and M. I. Jordan, "On convergence properties of the EM algorithm for gaussian mixtures," *Neural Comput.* **8**, 129–151 (1995).
25. R. Starr, S. Stahlheber, and A. Small, "Fast maximum likelihood algorithm for localization of fluorescent molecules," *Opt. Lett.* **37**, 413–415 (2012).
26. M. Hirsch, R. J. Wareham, M. L. Martin-Fernandez, M. P. Hobson, and D. J. Rolfe, "A stochastic model for electron multiplication charge-coupled devices – from theory to practice," *PLoS ONE* **8**, e53671 (2013).
27. I. Kalinina, A. Nandi, P. Delivani, M. R. Chacón, A. H. Klemm, D. Ramunno-Johnson, A. Krull, B. Lindner, N. Pavin, and I. M. Tolić-Nørrelykke, "Pivoting of microtubules around the spindle pole accelerates kinetochore capture," *Nat. Cell Biol.* **15**, 82–87 (2013).
28. V. Ananthanarayanan, M. Schattat, S. K. Vogel, A. Krull, N. Pavin, and I. M. Tolić-Nørrelykke, "Dynein motion switches from diffusive to directed upon cortical anchoring," *Cell* **153**, 1526–1536 (2013).
29. J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, and A. Cardona, "Fiji: an open-source platform for biological-image analysis," *Nat. Methods* **9**, 676–682 (2012).
30. S. K. Vogel, N. Pavin, N. Maghelli, F. Jülicher, and I. M. Tolić-Nørrelykke, "Self-organization of dynein motors generates meiotic nuclear oscillations," *PLoS Biol.* **7**, 0918–0928 (2009).

1. Introduction

With the advent of high time resolution microscopy and advances in the biological toolkit for microscopy, it has become possible to visualize the movement of cells and of single molecules inside a live cell [1]. For extracting information from these time-lapse images, it becomes essential to have the tools to track the motion of biological entities with high accuracy and efficiency. In cell biology, the objects of interest are often below the resolution limit of the microscope, appearing in the form of the system's point spread function (PSF). Furthermore low-light conditions are employed so as to avoid photo damage and bleaching, making the task even more challenging (Figs. 1(a) and 1(b)).

Image generation in low-light microscopy is affected by several types of noise, which aggravate the localization task. We will focus on shot noise, readout noise, and excess noise, which are generally considered to be the most relevant with regard to localization problems [9, 10, 11, 12, 13, 14]. Other kinds of noise include dark signal and clock induced charges [3]. The latter two will be a topic in the Discussion section.

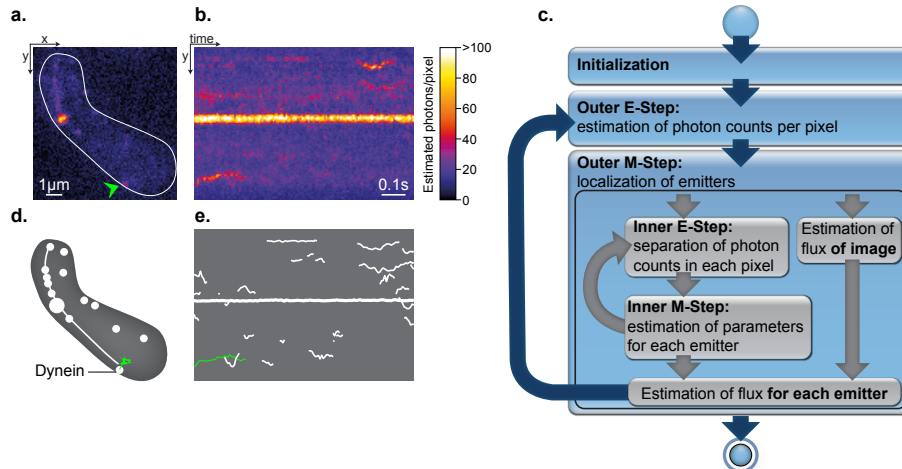


Fig. 1. The NMLA tracks proteins *in vivo* (a) Image of a fission yeast cell expressing dynein heavy chain tagged with 3 GFPs (Dhc1-3GFP). The white line marks the cell outline, the green arrowhead marks a dynein molecule. The brightest spot is the spindle pole body. (b) The movement of dyneins is visualized in consecutive maximum intensity projections onto the y-axis (see color bar). (c) Schematic representation of the NMLA. The outer loop, which addresses the excess noise, is depicted in blue while the inner loop, which performs the localization, is depicted in gray. (d) Scheme of cell with the 2D trace of a dynein (green, corresponding to the green arrowhead in a) tracked using the NMLA. (e) Projection onto the y-axis of the traces obtained using the NMLA. The green trace is the path of the molecule marked in a.

The first source of noise to be considered lies at the foundation of image generation: The number of photons detected in a pixel during an exposure is nondeterministic. The resulting effect, known as shot noise, is especially severe under low-light conditions. In the commonly used cameras with a charge-coupled device (CCD), the signal is additionally corrupted by additive Gaussian readout noise. It originates from the circuitry used to amplify the analog signal and finally convert it into a digital signal [2]. Readout noise poses a severe problem when little light is available. Cameras with electron multiplying CCDs (EMCCDs) reduce the effects of readout noise by amplifying the charge before it is read out. However, the nondeterministic amplification mechanism introduces yet another kind of noise, referred to as multiplicative or excess noise.

A variety of methods have been suggested to address the localization problem in noise affected microscopy images [4]. One simple approach is the calculation of the center of mass of the signal [5, 6]. However this approach is not useful for images with multiple emitters and high background. Another widely used approach is based on least squares fitting of an approximated PSF to the image [7, 8], but this method is suboptimal. For low-light settings, it has been shown that the theoretically motivated maximum likelihood (ML) approach to localization has superior precision [9].

The ML approach is based on a statistical model of the imaging process with its various sources of noise. ML localization methods address the problem of shot noise by viewing the number of photons in each pixel as random variable. In general, the number of photons n_i in each pixel i is assumed to follow a Poisson distribution $p_{n_i}(n_i; E_i)$. The mean of this distribution E_i is the flux at the pixel i , i.e. the expected number of photons detected at the pixel during an exposure. E_i depends on the locations of the objects which are to be localized. We will refer to

these objects as emitters and to their manifestations in the image as blobs. By assuming that the number of photons n_i at a pixel can be derived from the recorded intensity I_i at the pixel, the probability for the observation of an image is formulated as

$$\prod_{i=1}^M p_{n_i}(n_i; E_i), \quad (1)$$

where M is the number of pixels in the image. Different emitter locations change the expected photon counts E_i and thus result in different probabilities. The ML estimate is defined as the emitter locations which maximize Eq. (1).

Over the past years, this scheme and derivations have been used for localization in a number of studies, e.g. [9, 10, 11, 12]. In [9] Ober et al. formulated the problem and calculated an absolute limit for accuracy in the presence of additional readout noise. In [10] the ML approach was implemented in a fast graphics processing unit (GPU) based method. Later, similar techniques were used to simultaneously localize multiple emitters [11, 12].

The ML approach has recently been applied with a different noise model including excess noise caused by EMCCD amplification [13]. We will refer to the ML location estimator based on this model as excess noise aware ML estimator. The statistics of the amplification and its consequences regarding ML localization have been thoroughly discussed by Chao et al. [14]. We will base our considerations on the simplified model suggested by Basden et al. [15], which was recommended for high amplification gains in [14] and used in [13].

During EMCCD amplification, photo-electrons are shifted through a multiplication register. According to Basden et al. [15] the number of output electrons S_i given a certain number of input electrons n_i approximately follows a discrete version of an Erlang distribution:

$$p_{S_i|n_i}(S_i|n_i; g) = \frac{g^{n_i-1} \exp(-\frac{S_i}{g})}{g^{n_i} (n_i - 1)!}, \quad (2)$$

where g is the mean amplification gain, which we will refer to as EMCCD gain. In settings with low photon counts, such as in biological applications, a high EMCCD gain is beneficial as it leads to better localization quality [16]. The effects of readout noise become insignificant with high EMCCD gains and we therefore omit them [3, 10]. Furthermore, it will be assumed that the observed signal in a pixel is identical to the number of output electrons S_i . We will use the term signal as opposed to intensity in the context of this model, where there is no deterministic mapping from observable pixel output to the number of photons in a pixel.

When multiple emitters in close proximity are to be localized, the search for the ML estimate (excess noise aware as well as unaware) of their positions is a challenging high dimensional optimization task, because it has to be done jointly for all emitters. While [14] and [13] discuss only localization of single emitters, the localization tools from [10, 11, 12] achieve parallelization only by prior division of the image into subregions, which are then processed independently. It has not been recognized that the ML localization problem has a structure that allows for the use of a divide and conquer (DC) strategy to achieve higher stability and parallelization on a different level.

Here, we reformulate the optimization task to utilize this property in a novel nested expectation maximization (EM) algorithm that we refer to as Nested Maximum Likelihood Algorithm (NMLA), see Fig. 1(c). The EM meta algorithm was introduced by Dempster et al. [17] and is an established scheme for ML estimation in a setting with hidden variables. We show in simulation and experiments that our NMLA closely approaches the best theoretically possible precision, the Cramer-Rao lower bound (CRLB) [18]. We compare the results and required computation time of the NMLA with a well established general purpose optimization method

[19] and find that NMLA yields dramatic improvements in scalability and robustness. In addition, the application of NMLA in a biological setting is demonstrated (Figs. 1(d) and 1(e)).

We will now describe the general outline of our method. Detailed theory and calculations will be discussed in the Method section below. The NMLA consists of an inner and an outer loop (Fig. 1(c)). Starting with the shot- and excess noise corrupted image, the outer loop addresses excess noise by calculating an expected photon count in each pixel. Based on these results, the inner loop estimates the location of the emitters. In a pure shot noise setting, the inner loop on its own is able to perform ML localization. The EM scheme of the inner loop implements a DC strategy by enabling the refinement of parameter estimates to be done independently for each emitter in each iteration. Our nested approach allows us to use this strategy even in a setting affected by excess noise.

We will first describe an alternate view of the pure shot noise model followed by the description of our algorithm's inner loop and its application in a pure shot noise setting. Finally we will describe how the full NMLA implements an excess noise aware ML estimator.

2. Method

2.1. An alternate view of the shot noise model

Our shot noise model is equivalent to the one formulated in Eq. (1). In order to motivate the inner loop of the NMLA, we suggest an alternative view of the image generation process, focusing on the distribution of photons in the image instead of the number of photons in each pixel. A similar view has been described for image simulation in [9, 20], but was not utilized for localization. A comparable view has been applied to spectra analysis [21].

Image generation is seen as a two step process: First, a number N is randomly determined from a Poisson distribution $p_N(N; E)$ which is parametrized by the total flux of the image $E = \sum_{i=1}^M E_i$. Next, N photons are randomly distributed on the pixel grid of the image $\vec{I} = (I_1, \dots, I_M)$ according to a probability distribution $p_i(i; \vec{\theta})$. Each photon increases the intensity I_i at the pixel i that it hits by the constant c . A pixel's intensity is thus $I_i = n_i c$. The vector $\vec{\theta} = (\vec{\theta}_k, p_k, p_0 | k = 1, \dots, m)$ summarizes the parameters of m emitters and additional stray background light. While the vector $\vec{\theta}_k$ holds the location of emitter k , the term p_k stands for the percentage of light from this emitter. It can also be seen as the a priori probability of a photon to have the origin k . The term p_0 denotes the percentage of background light. The numbers p_k are under the constraint that their sum is equal to one.

Our perspective of image generation allows us to write the probability for the generation of an image \vec{I} as the product:

$$p_{\vec{I}}(\vec{I}; \vec{\theta}, E) = p_N(N; E) p_{\vec{I}|N}(\vec{I}|N; \vec{\theta}). \quad (3)$$

Please refer to the Appendix for a detailed deduction. The term $p_{\vec{I}|N}(\vec{I}|N; \vec{\theta})$ stands for the conditional probability of the image given that it contains a total of N photons. It is a multinomial distribution and can be written as $p_{\vec{I}|N}(\vec{I}|N; \vec{\theta}) = \hat{c} \prod_{i=1}^M p_i(i; \vec{\theta})^{\frac{I_i}{c}}$. The constant \hat{c} is the accordant multinomial coefficient.

Let us now examine the probability distribution $p_i(i; \vec{\theta})$ according to which the photons are distributed on the pixel grid. We consider it to be a mixture of $m + 1$ components:

$$p_i(i; \vec{\theta}) = \sum_{k=0}^m p_{i|k}(i|k; \vec{\theta}_k) p_k. \quad (4)$$

The term $p_{i|k}(i|k; \vec{\theta}_k)$ denotes the probability for a photon to be recorded at a specific pixel i , given it is of origin k . For photons originating from the background ($k = 0$) it is considered

a uniform distribution. The distribution for photons coming from one of the emitters ($k > 0$) is derived by integrating the PSF over the pixel's area and normalizing it with the PSF's integral over the image area. We use a Gaussian approximation of the PSF for this purpose, which was evaluated to be suitable for most applications in [13] and is a conventional choice used among others in [10, 11, 12].

2.2. Localization with the inner loop

We will now describe the inner loop of our algorithm which is capable of performing ML localization on multiple emitters. By utilizing Eq. (3), it is possible to independently acquire ML estimates for the model parameters

$$\vec{\theta}^* = \arg \max_{\vec{\theta}} p_{\vec{I}|N}(\vec{I}|N; \vec{\theta}) \quad (5)$$

and the total image flux

$$E^* = \arg \max_E p_N(N; E). \quad (6)$$

While the total image flux can be simply estimated as $E^* = \sum_{i=1}^M I_i/c$, estimation of the model parameters is not trivial. In order to find the optimal set of parameters defined above, we employ an EM algorithm. The algorithm starts with an initial set of parameters $\vec{\theta}^0$, which is determined by the user or the result from a previous frame, and generates a sequence of ever-improving models by executing two steps repeatedly:

- The inner E-step: Regard the current parameters $\vec{\theta}^t$ of the mixture distribution as correct and use it to estimate how much of the intensity of each pixel belongs to each light source and how much to the background.
- The inner M-step: Regard the estimates of the E-step as correct and use them to independently determine the parameters $\vec{\theta}_k^{t+1}$ and p_k^{t+1} of each light source.

It has been shown that the estimates obtained by the EM scheme improve with each iteration until convergence [22]. We will now describe the two steps in detail.

The inner E-step. Using Bayes' theorem, it is possible to calculate for each pixel i , the probability $p_{k|i}(k|i; \vec{\theta}^t)$ that a photon measured at this position has a specific origin k :

$$p_{k|i}(k|i; \vec{\theta}^t) = \frac{p_{i|k}(i|k; \vec{\theta}^t) p_k}{\sum_{k'=0}^m p_{i|k'}(i|k'; \vec{\theta}^t) p_k} \quad (7)$$

The E-step uses these probabilities to calculate a set of numbers $I_{i,k}^t$.

$$I_{i,k}^t = I_i p_{k|i}(k|i; \vec{\theta}^t) \quad (8)$$

This can be seen as separating the intensity I_i at each pixel i into fractions $I_{i,k}^t$, assumed to originate from the different sources k .

The inner M-step. The estimates for the a priori probabilities are acquired in the following way:

$$p_k^{t+1} = \frac{\sum_{i=1}^M I_{i,k}^t}{\sum_{i=1}^M I_i} \quad (9)$$

The a priori probability of the background can be calculated as $p_0^{t+1} = 1 - \sum_{k=1}^m p_k^{t+1}$. Based on the results of the E-step the new parameters for each emitter $\vec{\theta}_k^{t+1}$ of the new model are independently estimated as:

$$\vec{\theta}_k^{t+1} = \arg \max_{\vec{\theta}_k} \sum_{i=1}^M \left(I'_{i,k} \ln p_{i|k}(i|k; \vec{\theta}_k) \right) \quad (10)$$

In order to maximize the expression we use the general purpose optimization method known as Powell's method [19]. Optimization can be performed independently for each emitter.

Estimating the flux. For some applications, it may be desired to estimate not only the above mentioned parameters of each light source but also the amount of light that is emitted by it. The average number of photons \hat{E}_k recorded per exposure, originating from a source k , will be called the flux of the source. It can be calculated from the total flux of the image E :

$$\hat{E}_k = E p_k \quad (11)$$

While the inner loop described above provides an estimate for the a priori probability p_k of each light source, it is straight-forward to calculate an ML estimate for the total image flux E . It can be done independently of the rest of the parameters $\vec{\theta}$. The result is proportional to the sum of all pixel intensities:

$$E^* = \frac{\sum_{i=1}^M I_i}{c} \quad (12)$$

Accordingly, the flux of a specific light source can be estimated up to the factor c as $\hat{E}_k^* = p_k E^*$. The gain factor c of the system is required in order to compute absolute estimates of the photon counts. If it is unknown, the results obtained still allow for comparison between different light sources.

2.3. Localization with the complete NMLA

Let us now examine the complete model including the probabilistic EMCCD amplification as described in [15]. We observe an image of EMCCD signals $\vec{S} = (S_1, \dots, S_M)$ generated from an unknown photon image $\vec{n} = (n_1, \dots, n_M)$. During amplification, each pixel's signal S_i is taken from the distribution described in Eq. (2). The new task is now to concurrently find the model parameters $\vec{\theta}^*$ and the flux of the image E^* , which maximize the probability to observe the image \vec{S} of signals. In this model, it is not possible to find the estimates independently:

$$(\vec{\theta}^*, E^*) = \arg \max_{(\vec{\theta}, E)} p_{\vec{S}}(\vec{S}; \vec{\theta}, E; g) \quad (13)$$

$$= \arg \max_{(\vec{\theta}, E)} \sum_{\vec{n} \in \vec{n}_{set}} p_{\vec{S}, \vec{n}}(\vec{S}, \vec{n}; \vec{\theta}, E; g), \quad (14)$$

where the function $p_{\vec{S}, \vec{n}}(\vec{S}, \vec{n}; \vec{\theta}, E; g)$ denotes the joint probability to acquire a photon image \vec{n} and observe \vec{S} . The term \vec{n}_{set} denotes the set of possible photon images of the same size.

We again apply the EM scheme to find the solution. Starting again with an initial estimate by the user or from a previous frame $(\vec{\theta}^0, E^0)$, the algorithm creates an iteratively improving sequence of tuples that converge to an optimum. As in the optimization of the pure shot noise model, the new estimates are in each iteration calculated by executing two steps:

- The outer E-step: Regard the current parameters ($\vec{\theta}^t, E^t$) as correct and use them to estimate each pixel's expected photon count \hat{n}_i^t , given the pixel's observed signal S_i .
- The outer M-step: Regard the results of the E-step as correct and use them to re-estimate the parameters $\vec{\theta}^{t+1}$ as well as the total image flux E^{t+1} and the flux for each light source \hat{E}_k^{t+1} .

The two estimation tasks to be solved in the outer M-step closely resemble the ML task defined for the pure shot noise model described above and in fact we can use the same scheme to solve them. We will now describe the steps in detail:

The outer E-step. Based on the current parameters ($\vec{\theta}^t, E^t$) we calculate the expected number of photons \hat{n}_i^t for each pixel i , given the observed signal S_i . It is defined as:

$$\hat{n}_i^t = \frac{\sum_{n=0}^{\infty} n p_{S_i|n_i}(S_i|n; g) p_{n_i}(n; \vec{\theta}^t, E^t)}{\sum_{n=0}^{\infty} p_{S_i|n_i}(S_i|n; g) p_{n_i}(n; \vec{\theta}^t, E^t)} \quad (15)$$

The function $p_{n_i}(n; \vec{\theta}^t, E^t)$ stands for the a priori probability of a specific number of photons n to have hit pixel i . It is assumed to be a Poisson distribution with the flux E_i^t at the pixel as its single parameter. We compute this flux based on the last estimates of the model $\vec{\theta}^t$ and the mean total photon count E^t :

$$E_i^t = p_i(i; \vec{\theta}^t) E^t \quad (16)$$

We compute Eq. (15) efficiently as

$$\hat{n}_i^t = \sqrt{E_i^t \cdot S_i / g} \cdot \frac{I_0(2 \cdot \sqrt{S_i \cdot E_i^t / g})}{I_1(2 \cdot \sqrt{S_i \cdot E_i^t / g})}, \quad (17)$$

where I_0 and I_1 are the modified Bessel functions of the first kind of order zero and one, respectively. For a detailed discussion of this simplification please refer to the Appendix.

The outer M-step. Even though the estimates for the model and the mean photon count have to be performed together, in each iteration of the algorithm their new estimates can be found independently of one another. The tasks are essentially equivalent to tasks described in Eq. (5) and Eq. (6). We will first describe the optimization of the emitter parameters $\vec{\theta}^{t+1}$

Optimizing the estimate of the emitter parameters: Using the expected values calculated in the previous step, the next parameter estimate is defined as:

$$\vec{\theta}^{t+1} = \arg \max_{\vec{\theta}} \sum_{i=1}^M \hat{n}_i^t \ln p_i(i; \vec{\theta}) \quad (18)$$

Expression 18 can be identified with the maximization problem formulated in Eq. (5). Indeed the new parameter estimates $\vec{\theta}^{t+1}$ can be acquired by employing the inner loop described above. The only modification necessary is a substitution of the image intensities I_i by the current expected photon count \hat{n}_i^t . This constitutes a nested EM algorithm in which the M-step consists of an EM algorithm itself.

Optimizing the estimate for the total image flux: The maximum likelihood estimation of the total image flux can be identified with the task from Eq. (6). It has an analytical solution analogous to Eq. (12):

$$E^* = \sum_{i=1}^M \hat{n}_i^t \quad (19)$$

As in the pure shot noise model, the flux for the different light sources \hat{E}_k^* can be calculated accordingly as $\hat{E}_k^* = p_k^* E^*$.

3. Experiments

3.1. Evaluation with simulated images

To evaluate our method and determine what additional gain is to be expected from the extension of the pure shot noise model, the inner loop as well as the complete NMLA were applied to simulated images with and without excess noise. We compared the standard deviation of the localization results with theoretically predicted bounds for precision (Fig. 2). As expected, the inner loop of our algorithm attains the CRLB in a pure shot noise case for a range of different values of flux (Fig. 2(a)) [10]. After deriving the CRLB for the excess noise aware ML estimator, we find that it is attained by the NMLA. We confirm the finding from [20] that the approximation of the CRLB for excess noise, obtained by multiplying the CRLB for a pure shot noise setting by a factor of $\sqrt{2}$ is not appropriate for low photon counts. Surprisingly, this approximation from [13] closely corresponds to the results attained by the ML estimator, which is based on a pure shot noise model and implemented by our inner loop (Figs. 2,3).

The same general picture can be observed when different levels of Poisson background noise are applied (Fig. 2(b)). With increasing background, however, a deviation between the results of all three ML estimators and the corresponding theoretical bounds is evident. This apparent general feature of the ML localization has been observed independently in [23]. Since heavy background is a common feature in biological applications this effect might be worth further examination.

3.2. Evaluation on images of fluorescent beads

To evaluate our technique experimentally, we imaged fluorescent beads under various light conditions and tracked them using the NMLA and its inner loop (Fig. 3). As in simulation, the results are in agreement with the theoretical bound. The NMLA yielded a lower standard deviation than its inner loop for each of the observed tracks. As in simulation the difference is larger for low photon counts.

3.3. Evaluation of computation time and robustness

We compared the performance of the NMLA and a direct application of Powell's method [19] to the log likelihood as described in [13] (supplementary equation 138). To assess the scalability we applied both optimizers to a grid of overlapping emitters and measured the required computation time for different emitter numbers (Fig. 4(a)). To compare the robustness of the methods we initialized them at varying positions and measured localization success in a single emitter setting (Fig. 4(b)). In both cases our NMLA outperforms Powell's method.

Note that the set of parameters Powell's method is employed on differs from our $\vec{\theta}$ since it does not include the a priori probabilities p_k . Instead the flux of each emitter $\hat{E}_k = p_k E, \forall k = 1, \dots, m$ as well as the background flux $\hat{E}_0 = p_0 E$ are used. When performing our experiments, we used the implementation of Powell's method in the Apache Commons Mathematics Library. We set the absolute and relative threshold used as stopping criteria to 0.001, which produced results of similar quality as the NMLA.

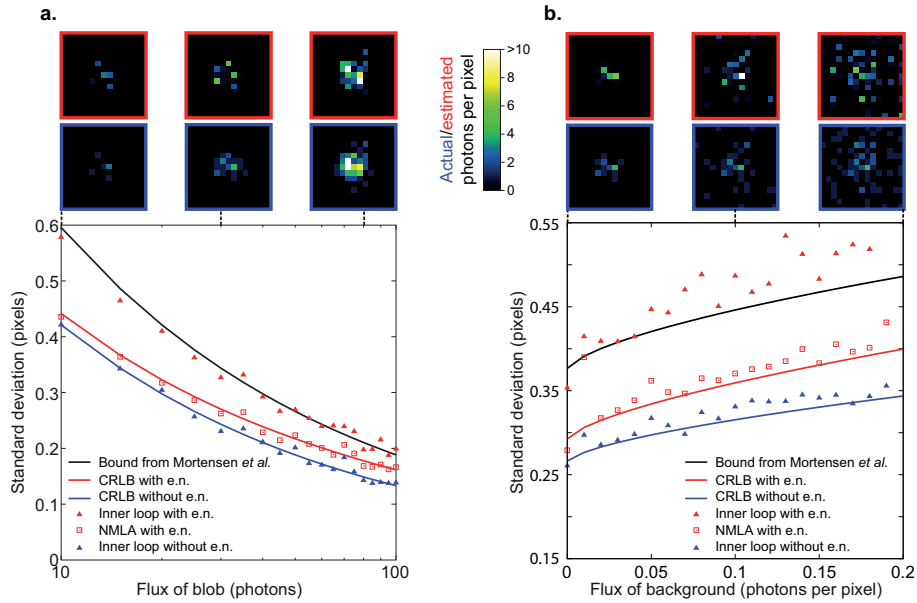


Fig. 2. The NMLA and its inner loop closely approach the theoretical bounds in simulation. (a) The standard deviation of the localization results in one dimension as a function of the flux of the blob. The complete NMLA (red squares) and the inner loop (red triangles) were applied to the simulated images. The inner loop was also applied to the same image series without excess noise (blue triangles). The results are compared with theoretical bounds: the Cramer-Rao lower bound (CRLB) for a scenario without excess noise (blue line), the CRLB considering excess noise (red line) and the approximated bound suggested in [13] (black line). The NMLA yields the highest possible precision, given by the CRLB, in settings with excess noise. (b) The standard deviation of the localization results as a function of background flux. The flux of the blob was set to 25 photons. The remaining parameters and the algorithms employed were as in a. Note that the deviation between the theoretical bounds and the results of the estimators (also seen in [23]) is a general feature of the ML localization. In a and b, simulated images (without excess noise outlined in blue, and the same image with excess noise outlined in red) correspond to different values of flux/background flux indicated by the dashed lines under the images. In the images without excess noise the colors indicate the number of photons in each pixel. In the images with excess noise colors indicate the estimated number of photons as described in the Appendix.

4. Discussion

4.1. General findings regarding the excess noise aware ML estimator

The ML estimator based on the pure shot noise model is frequently used on EMCCD images, e.g. in [10, 11, 12], likely because the theoretical understanding of the problem has only recently been achieved [14, 13]. We suggest to use the ML estimator based on the pure shot noise model as an approximate solution only if enough light is available. One should also consider the findings from [14] and determine whether EMCCD amplification is the tool of choice in such a setting. On the other hand, for the low photon count setting which is relevant in biology and astronomy, we conclude from theory, simulation, and experiments that significantly better results can be obtained when the statistical model accounts for excess noise in addition to shot noise. The NMLA is a powerful implementation of the excess noise aware ML estimator.

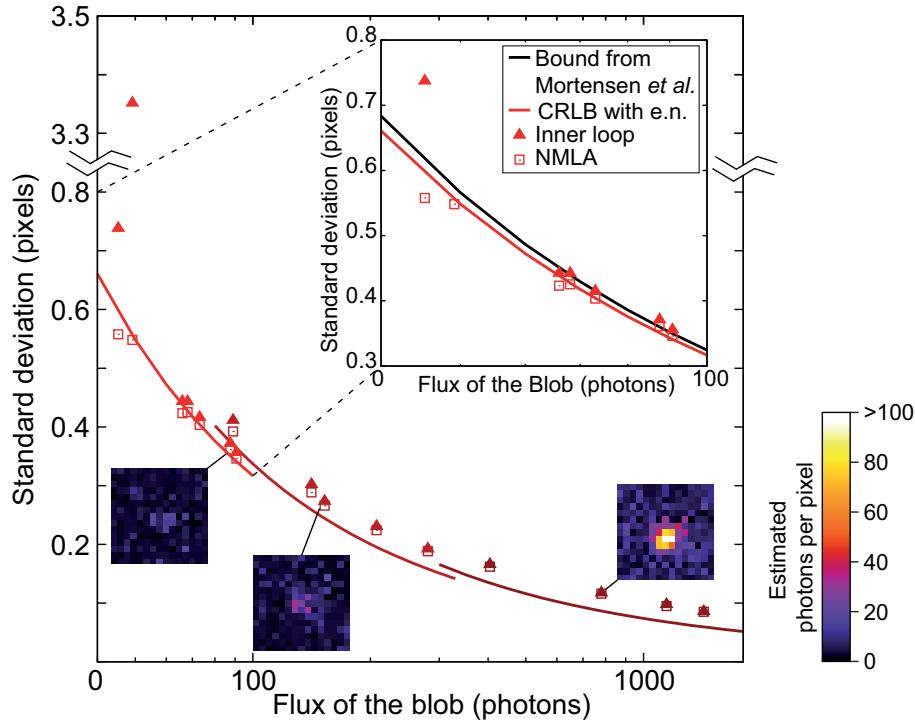


Fig. 3. The standard deviation of the localization results as a function of the measured flux of fluorescent bead images. Three 3000-frame-long movies were obtained using Total Internal Reflection Fluorescence (TIRF) microscopy with different laser powers and EMCCD gain of 300. The pixel size was 106 nm (see Appendix for details). The beads were localized with the NMLA (squares) and its inner loop (triangles). Each pair of a triangle and a square at a certain value of flux corresponds to a single bead; displayed are images of three tracked beads connected to their resulting data point. The results are compared with the CRLB considering excess noise. The background flux used in the calculation of the bound was derived from the average intensities in manually selected empty areas of the movies (see Appendix for details). The inset is an enlarged view of the section corresponding to the low flux region (dashed lines) including the theoretical bound from [13].

4.2. Practical advantages of NMLA

We see three major advantages of our method with regard to practical application:

First, our algorithm keeps computation time low even as the number of overlapping emitters is increased (Fig. 4(a)). Second, since the algorithm allows independent re-estimation of emitter locations during the inner M-step, it allows for easy coarse level parallelization. This could facilitate GPU implementations. While previous methods [10, 11, 12] rely on the division of the image into subregions in a preprocessing step, NMLA becomes faster through parallelization even when all emitters are in close proximity and have to be localized jointly (Fig. 4(a)). Third, our algorithm proved to be significantly less sensitive to random initialization than the general purpose optimizer (Fig. 4(b)).

4.3. Conceptual advances

We achieve the level of robustness and scalability shown in Fig. 4 by reformulating the optimization task in a way that allows for the use of EM. The task of localization in a shot noise

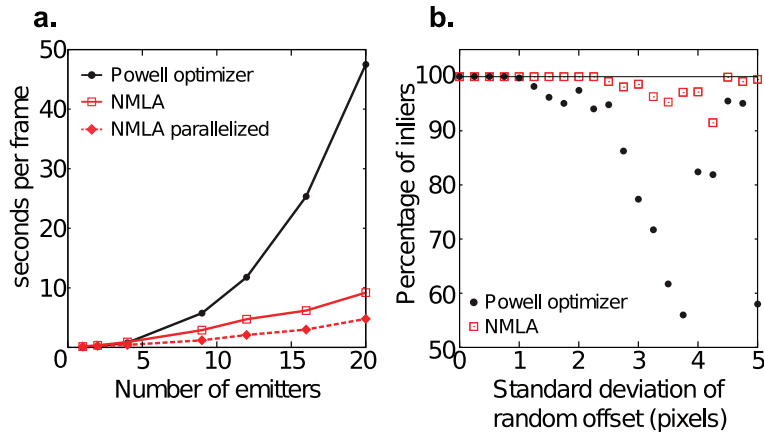


Fig. 4. The NMLA outperforms Powell's method with regard to scalability and robustness. (a) Required computation time as a function of the number of emitters. Different numbers of overlapping emitters were tracked with three different implementations of the excess noise aware ML estimator: The computation time required by Powell's method (circles) grows significantly faster than the time required by the NMLA (squares). A naively parallelized version of the NMLA (diamonds) yields a further improvement. (b) Number of inliers in localization results as a function of the standard deviation of the random initialization. In each frame the algorithms were initialized at a Gaussian distributed random location around the true position of the emitter. Every location estimate closer than 1.3 pixels to the correct location was considered an inlier. The NMLA achieves 100% inliers for higher standard deviations and does not lose as many as Powell's method.

setting is very closely related to the problem of Gaussian mixture estimation and EM is well established for this problem and demonstrates solid convergence properties [24]. In addition, it does not require the calculation of derivatives, in contrast to many general purpose optimization methods.

It should also be noted, that our novel formulation of the optimization task defined in Eq. (5) is of lower dimensionality than the previous formulation which can be found in [9, 10, 11, 12]. Our approach still yields the same optimum. In all previous formulations, apart from the emitter locations, $m + 1$ parameters namely the flux of each emitter and the flux of the background had to be determined. Apart from the emitter locations, our formulation requires estimation of only m parameters, namely each emitter's percentage of light p_k . This is due to the fact, that p_0 can be calculated directly from the other values p_k .

By wrapping the inner loop in another EM algorithm we are able to effectively adapt these properties into a setting involving excess noise. To our knowledge, we are the first to explore the excess noise aware ML estimator for multiple emitters and from an algorithmic point of view.

4.4. Perspective

Our approach allows us to reduce the problem of ML localization in an excess noise affected setting to repeated localization in a pure shot noise setting. Therefore, it is possible to apply any new algorithmic finding for the pure shot noise setting in the presence of excess noise. An example could be the algorithm in [25], which radically reduces the complexity for the localization of a single emitter in a shot noise setting. Our framework makes it possible to incorporate such methods in the inner M-step, thereby making them applicable with excess

noise.

Even though only two types of noise, shot noise and excess noise were considered, our general scheme is not limited to this noise model. Other components could be included without structural change if they prove to be beneficial. Readout noise has little effect in our setting [3], but it could be included by calculating $p_{S_i|n_i}(S_i|n_i; g)$ as convolution of Eq. (2) and a Gaussian distribution [13]. The other parts of our method would remain unaffected. A different source of noise called dark signal is a Poisson distributed charge arising from imperfections in the production process of the chip [2]. Dark signal is pixel dependent [2] and its intensity pattern on the sensor can be measured. Consequently, one could extend our noise model by adding a measured pixel dependent term to each pixel's flux E_i . All other calculations could be performed as described above. Finally, our algorithm could include clock induced charges, which are generated by shifting electrons through the CCD [26]. According to [26] these charges should be considered in the same way as dark signal. It would thus be possible to incorporate clock induced charges into our algorithm in the same manner as dark signal.

It should be noted that since we were aiming at a qualitative demonstration, the NMLA currently relies on user initialization and our implementation has not yet been optimized with regard to performance. While an interactive rather than an automatic behavior is desirable for some applications we see no reason why our general algorithmic scheme could not be combined with initialization mechanisms and highly optimized GPU methods as used in [10, 11, 12] to produce a fully automatic system.

Implementations of the NMLA are invaluable for studying biological phenomena such as membrane diffusion and cellular kinetics of molecules. While conventional methods of studying particle properties inside the cell such as FRAP (Fluorescent Recovery After Photobleaching) or FCS (Fluorescent Correlation Spectroscopy) provide bulk measurements of a population of molecules, single particle tracking enabled by the NMLA implementation offers kinetic data at the single-molecule level. We can thus study the temporal evolution of each particle, as well as detect the existence of different populations of particles. In addition, our NMLA offers the possibility of simultaneous tracking of multiple molecules imaged with low-light in a crowded environment such as inside the cell, while at the same time ensuring a lower computation time but higher precision than other methods.

5. Conclusion

ML estimation in EMCCD images and its potential in low-light microscopy has accrued considerable theoretical understanding. [14]. We have designed and tested a new tool for ML estimation that can help to utilize these findings in cell biology. A variant of this tool has already been successfully employed by us in [27, 28].

An implementation of the NMLA is available as Open Source software under the GNU General Public License. It can be installed under the name Low Light Tracking Tool (http://fiji.sc/Low_Light_Tracking_Tool) via the image processing software Fiji [29].

Appendix

Imaging using the Total Internal Reflection Fluorescence Microscope

An inverted stand, manual XY stage Olympus IX71 microscope (Olympus, Tokyo, Japan) with custom-built TIRF condenser and manual TIRF angle adjustment was employed for imaging beads (Fig. 3) and dyniens inside the cytoplasm of a fission yeast zygote (Figs. 1(a) and 1(b)). Imaging was performed using an Olympus UApo 150x 1.45 Oil TIRFM inf/0.13-0.21 corr (Olympus, Tokyo, Japan) objective with diode-pumped solid state 491nm laser for GFP excitation and 560 nm laser for mCherry/tdTomato excitation (75mW; Cobolt, Solna, Sweden).

Laser intensity was controlled using the acousto-optic tunable filter in the Andor Revolution laser combiner (ALC, Andor Technology plc., Belfast, UK). The wavelength filter used was the BL HC 525/30 (Semrock Inc., Rochester, NY, USA). The microscope was equipped with an Andor iXon EM+ DU-897 BV back illuminated EMCCD (Andor Technology plc., Belfast, UK) with pixel size of EMCCD chip being 16 μm and image pixel size being 0.106 μm with the 150x objective. The system was controlled using the Andor iQ software version 1.9.1 (Andor Technology plc., Belfast, UK). For imaging dynein in the cytoplasm (Fig. 1), the imaging conditions used were: excitation with 80% power (18mW) of 491nm laser, exposure time of 5-9ms, with 2000 continuous repetitions. The dyneins in the strain used in this paper are labeled with triple GFP [30]. The beads (Estopor 0.103 μm fluorescent beads, Merck KGaA, Darmstadt, Germany) were imaged with 2%, 3% and 5% power (0.45mW, 0.675mW and 1.125mW respectively) of 491nm laser, exposure time of 5ms, with 3000 continuous repetitions. The EMCCD gain used was 300 and the pre-amplification gain was 4.8.

The image probability as product

In Eq. (3) we formulate the probability for the observation of a specific image \vec{I} created by a total of N photons as the product of the probability $p_N(N; E)$ that an image contains exactly N photons and the probability $p_{\vec{I}|N}(\vec{I}|N; \vec{\theta})$, that the image is observed under the condition, that it contains exactly N photons. We will now explain why this is possible. According to Bayesian theory we can write:

$$p_{\vec{I}}(\vec{I}; \vec{\theta}, E) = \sum_{\hat{N}=0}^{\infty} \left(p_N(\hat{N}; E) p_{\vec{I}|\hat{N}}(\vec{I}|\hat{N}; \vec{\theta}) \right). \quad (20)$$

Since it is impossible to generate the image \vec{I} from any number of photons other than the exact number of photons N that created it, the probability $p_{\vec{I}|\hat{N}}(\vec{I}|\hat{N}; \vec{\theta})$ will be 0 in all summands except for the single case $\hat{N} = N$. We can thus write:

$$p_{\vec{I}}(\vec{I}; \vec{\theta}, E) = p_N(N; E) p_{\vec{I}|N}(\vec{I}|N; \vec{\theta}). \quad (21)$$

Calculating the CRLB

We calculated the CRLB as the inverse of the Fisher information matrix [18]. For the pure shot noise setting the elements of the Fisher information matrix were calculated as described in [10]. For the case with excess noise, the elements of the Fisher information matrix can be computed as

$$FIM_{ab} = \sum_{i=1}^M E'_{S_i; \vec{\theta}, E} \left[\frac{\partial \ln q(S_i; E_i)}{\partial a} \cdot \frac{\partial \ln q(S_i; E_i)}{\partial b} \right] \quad (22)$$

$$= \sum_{i=1}^M \sum_{S_i} q(S_i; E_i) \cdot \frac{\partial \ln q(S_i; E_i)}{\partial a} \cdot \frac{\partial \ln q(S_i; E_i)}{\partial b}, \quad (23)$$

where $a, b \in \vec{\theta} \cup \{E\}$ are the parameters and $E'_{S_i; \vec{\theta}, E}$ stands for the expected value with regard to the signal S_i of the pixel i and $q(S_i; E_i)$ describes the probability distribution of the signal S_i , which is parametrized by the pixel's flux $E_i = p_i(i; \vec{\theta}) \cdot E$. A deduction of Eq. (22) can be

found in [14]. As done in [13] we calculate $q(S_i; E_i)$ efficiently as

$$\begin{aligned} q(S_i; E_i) &= \sum_{n=0}^{\infty} p_{S_i|n_i}(S_i|n; g) \cdot p_n(n; E_i) \\ &= \exp(-E_i) \cdot \delta(S_i) + \exp(-S_i/g - E_i) \cdot \frac{g^{-1} \cdot E_i}{\sqrt{E_i \cdot S_i/g}} \cdot I_1\left(2 \cdot \sqrt{E_i \cdot S_i/g}\right), \end{aligned} \quad (24)$$

with I_1 being the modified Bessel function of the first kind of order one.

Deduction for the use of the Bessel function

In this section the step from the inconvenient Eq. (15) to the convenient Eq. (17) is derived. Eq. (15) is inconvenient with regard to computation, because of the infinite number of summands. Eq. (17) is convenient because the Bessel functions are very well investigated and fast implementations are available.

Reiterating, Eq. (2) describes the probability that a signal S_i is observed in a pixel given a photon count of n_i . Since the factorial of a negative number as well as the division by 0 is not defined we assume that this probability $p_{S_i|n_i}(S_i|n_i = 0; g) = 0$ for all $S_i > 0$ and that $p_{S_i|n_i}(S_i = 0|n_i = 0; g) = 1$. We furthermore assume that $p_{S_i|n_i}(S_i = 0|n_i; g) = 0$ for all $n_i > 0$. These are reasonable assumptions when considering the underlying model of electron amplification. At least one electron has to be present to generate an output of $S_i > 0$. Let us furthermore remember the assumption that the number of photons in each pixel is Poisson distributed:

$$p_n(n; E_i) = \frac{E_i^n}{n!} \cdot \exp(-E_i). \quad (26)$$

We will now examine the denominator in Eq. (15):

$$\begin{aligned} &\sum_{n=0}^{\infty} p_{S_i|n_i}(S_i|n; g) \cdot p_n(n; E_i) \\ &= \exp(-E_i) \cdot \delta(S_i) + \sum_{n=1}^{\infty} p_{S_i|n_i}(S_i|n; g) \cdot p_n(n; E_i) \\ &= \exp(-E_i) \cdot \delta(S_i) + \sum_{n=1}^{\infty} \frac{S_i^{(n-1)} \cdot \exp(-S_i/g)}{g^n \cdot (n-1)!} \cdot \frac{E_i^n}{n!} \cdot \exp(-E_i) \\ &= \exp(-E_i) \cdot \delta(S_i) + \exp(-S_i/g - E_i) \cdot \sum_{n=1}^{\infty} \frac{S_i^{(n-1)} \cdot E_i^n}{g^n \cdot n! \cdot (n-1)!} \end{aligned} \quad (27)$$

substituting $n = l + 1$ we have

$$\begin{aligned}
& \exp(-E_i) \cdot \delta(S_i) + \exp(-S_i/g - E_i) \cdot \sum_{l=0}^{\infty} \frac{S_i^l \cdot E_i \cdot E_i^l}{g \cdot g^l \cdot l! \cdot (l+1)!} \\
&= \exp(-E_i) \cdot \delta(S_i) + \exp(-S_i/g - E_i) \cdot g^{-1} \cdot E_i \cdot \sum_{l=0}^{\infty} \frac{1}{l! \cdot (l+1)!} \cdot (S_i \cdot E_i/g)^l \\
&= \exp(-E_i) \cdot \delta(S_i) + \exp(-S_i/g - E_i) \cdot g^{-1} \cdot E_i \cdot \sum_{l=0}^{\infty} \frac{1}{l! \cdot (l+1)!} \cdot (S_i \cdot E_i/g)^l \cdot \frac{\sqrt{E_i \cdot S_i/g}}{\sqrt{E_i \cdot S_i/g}} \\
&= \exp(-E_i) \cdot \delta(S_i) + \exp(-S_i/g - E_i) \cdot \frac{g^{-1} \cdot E_i}{\sqrt{E_i \cdot S_i/g}} \cdot \sum_{l=0}^{\infty} \frac{1}{l! \cdot (l+1)!} \cdot (S_i \cdot E_i/g)^l \cdot \sqrt{E_i \cdot S_i/g} \\
&= \exp(-E_i) \cdot \delta(S_i) + \exp(-S_i/g - E_i) \cdot \frac{g^{-1} \cdot E_i}{\sqrt{E_i \cdot S_i/g}} \cdot \underbrace{\sum_{l=0}^{\infty} \frac{1}{l! \cdot (l+1)!} \cdot \left(\frac{2 \cdot \sqrt{E_i \cdot S_i/g}}{2}\right)^{2l+1}}_{I_1(2 \cdot \sqrt{E_i \cdot S_i/g})} \\
&= \exp(-E_i) \cdot \delta(S_i) + \exp(-S_i/g - E_i) \cdot \frac{g^{-1} \cdot E_i}{\sqrt{E_i \cdot S_i/g}} \cdot I_1\left(2 \cdot \sqrt{E_i \cdot S_i/g}\right) \tag{28}
\end{aligned}$$

where

$$I_1\left(2 \cdot \sqrt{E_i \cdot S_i/g}\right) = \sum_{l=0}^{\infty} \frac{1}{l! \cdot (l+1)!} \cdot \left(\frac{2 \cdot \sqrt{E_i \cdot S_i/g}}{2}\right)^{2l+1} \tag{29}$$

is the modified Bessel function of the first kind of order one:

$$I_1(z) = \sum_{l=0}^{\infty} \frac{1}{l! \cdot (l+1)!} \cdot \left(\frac{z}{2}\right)^{2l+1} \tag{30}$$

Let us now consider the numerator from Eq. (15):

$$\begin{aligned}
& \sum_{n=1}^{\infty} p_{S_i|n_i}(S_i|n;g) \cdot p_n(n;E_i) \cdot n \\
&= \sum_{n=1}^{\infty} \frac{S_i^{(n-1)}}{g^n \cdot (n-1)!} \cdot \exp(-S_i/g) \cdot \exp(-E_i) \cdot \frac{E_i^n}{n!} \cdot n \\
&= \exp(-S_i/g - E_i) \cdot \sum_{n=1}^{\infty} \frac{S_i^{(n-1)} \cdot E_i^n}{g^n \cdot ((n-1)!)^2} \tag{31}
\end{aligned}$$

substituting again $n = l + 1$ we get

$$\begin{aligned}
& \exp(-S_i/g - E_i) \cdot \sum_{l=0}^{\infty} \frac{S_i^l \cdot E_i \cdot E_i^l}{g \cdot g^l \cdot (l!)^2} \\
&= \exp(-S_i/g - E_i) \cdot g^{-1} \cdot E_i \cdot \sum_{l=0}^{\infty} \frac{(S_i \cdot E_i/g)^l}{(l!)^2} \\
&= \exp(-\alpha \cdot S_i - E_i) \cdot g^{-1} \cdot E_i \cdot \underbrace{\sum_{l=0}^{\infty} \frac{1}{(l!)^2} \cdot \left(\frac{2 \cdot \sqrt{S_i \cdot E_i/g}}{2}\right)^{2l}}_{I_0(2 \cdot \sqrt{S_i \cdot E_i/g})} \\
&= \exp(-S_i/g - E_i) \cdot g^{-1} \cdot E_i \cdot I_0\left(2 \cdot \sqrt{S_i \cdot E_i/g}\right) \tag{32}
\end{aligned}$$

where

$$I_0\left(2 \cdot \sqrt{S_i \cdot E_i / g}\right) = \sum_{l=0}^{\infty} \frac{1}{(l!)^2} \cdot \left(\frac{2 \cdot \sqrt{S_i \cdot E_i / g}}{2}\right)^{2l} \quad (33)$$

is the modified Bessel function of the first kind of order zero:

$$I_0(z) = \sum_{l=0}^{\infty} \frac{1}{(l!)^2} \cdot \left(\frac{z}{2}\right)^{2l} \quad (34)$$

Combining Eq. (15) with Eq. (28) and Eq. (32) we get:

$$\begin{aligned} \hat{n}_i &= \frac{\sum_{n=0}^{\infty} P_{S_i|n_i}(S_i|n;g) \cdot p_n(n;E_i) \cdot n}{\sum_{n=0}^{\infty} P_{S_i|n_i}(S_i|n;g) \cdot p_n(n;E_i)} \\ &= \frac{\exp(-S_i/g - E_i) \cdot g^{-1} \cdot E_i \cdot I_0\left(2 \cdot \sqrt{S_i \cdot E_i / g}\right)}{\exp(-E_i) \cdot \delta(S_i) + \exp(-S_i/g - E_i) \cdot \frac{g^{-1} \cdot E_i}{\sqrt{E_i \cdot S_i / g}} \cdot I_1\left(2 \cdot \sqrt{E_i \cdot S_i / g}\right)} \\ &= \begin{cases} 0 & \text{if } S_i = 0 \\ \sqrt{S_i \cdot E_i / g} \cdot \frac{I_0\left(2 \cdot \sqrt{S_i \cdot E_i / g}\right)}{I_1\left(2 \cdot \sqrt{S_i \cdot E_i / g}\right)} & \text{if } S_i > 0 \end{cases} \quad (35) \end{aligned}$$

It follows from the considerations above, that Eq. (15) can be calculated by equation Eq. (17) and that it should be interpreted as zero in the special case where $S_i = 0$.

Simulation of images for precision measurement

To calculate the standard deviations shown in Fig. 2(a) we generated a series of 250 simulated 15x15 pixel images for each value of flux. For Fig. 2(b) a series of 1000 simulated 15x15 pixel images was created for each value of background flux. In both cases a single emitter was placed in the center of each image. The standard deviation of the Gaussian used as PSF was set to be 1.3 pixels. Photon images were generated by randomly drawing each pixel's photon count n_i from a Poisson distribution parametrized by the pixel's flux E_i . We calculated the flux as combination of the Gaussian integral over the pixel's area and the per pixel background flux. The algorithms were initialized in the center of the first frame. The initial value of p_k in the first frame was 10% for all emitters. All algorithms subsequently used the results of one frame as initialization for the next one.

After the inner loop was applied to these photon images, the EMCCD amplification was simulated. An EMCCD gain of 300 was used. As in [13], we drew random numbers from a continuous gamma distribution, approximating for the discrete distribution described by Eq. (2). The results were then rounded to determine the pixel's final signal S_i . We use the Apache Commons Mathematics Library implementation of the Poisson distribution and the Colt Library implementation of the gamma distribution.

Measuring computation time

The simulated images used to produce Fig. 4(a) were generated in the same way as the images used for precision measurement, but using a grid of emitters instead of a single one. The image sequences had a size of 11x11 pixels (1 emitter), 11x16 pixels (2 emitters), 16x16 pixels (4 emitters), 21x21 pixels (9 emitters), 21x 26 pixels (12 emitters), 26x26 pixels (16 emitters) and 26x31 pixels (20 emitters). The first emitter was set at pixel location (5,5) with all other emitters shifted in x and y direction by multiples of 5 pixels. The Gaussians used to model the PSF of

the emitters was set to have a standard deviation of 1.3 pixels. Each emitter was set to have a flux of 100 photons. Background flux was set to 1 photon per pixel. A sequence of 10 frames was generated for each number of emitters. The images were post processed to account for EMCCD excess noise as was done for precision measurement.

To perform tracking, the appropriate number of emitters was initialized in the first frame on their true position and the initial p_k^0 set to 5%. It should be noted that this initialization is problematic when applied on the grid with 20 emitters, because background is implicitly set to 0. We found however, that localization was performed correctly in practice.

The localization results from one frame were used as initialization for the next. Computation time was averaged over all frames. The calculations were performed on a Intel® Core™ i7-3820 Processor. The program was run on a virtual machine with 4 simulated cores (using VMware Player). The guest system was Ubuntu 12.04 hosted by Windows 7.

Measuring robustness

Sequences of 1000 simulated 15x15 pixel images with were generated and post processed as for precision measurement. The flux of the single emitter was set to 100 photons. The background flux was set to 0.25 photons per pixel. For each frame the initial p_k^0 for the estimators was set to 5%. The initialization of the location was randomly determined in each frame using a normal distribution centered around the true position. Different standard deviations were used for each sequence.

Estimation of photon counts per pixel

The estimated photon counts per pixel indicated by color in Figs. 1 and 3 were computed according to the manufacturer's information. First, the bias of 100 counts was subtracted from the signal (resulting values below zero were interpreted as zero). Then, the resulting values were divided by the EMCCD gain of 300 and finally multiplied by a sensitivity factor (given in the Andor iXon EM+ DU-897 performance sheet) of 12.17 electrons per analog/digital count.

To calculate the background flux used in the calculation of the theoretical bounds in Fig. 3 an area without beads was manually selected in each movie. The pixel value was then averaged over all frames and pixels in the area. Finally, as for the photon count estimation described above, the bias was subtracted, the value was divided by the EMCCD gain and finally multiplied by the sensitivity factor.

The photon count per pixel for the simulated EMCCD images in Fig. 2 was calculated by dividing the pixel's signal by the EMCCD gain of 300.

Tracking of fluorescent beads

Isolated beads were selected from each movies first frame. The beads were tracked independently, starting from a manual initialization in the first frame. The initial value of p_k in the first frame was 10% for all emitters. The standard deviation of the Gaussian to be used as approximate PSF was experimentally determined and set to be 1.3 pixels for all emitters. The cameras bias of 100 counts in each pixel was subtracted for the complete NMLA as well as for the inner loop alone. In case of the complete NMLA, each pixel's signal was subsequently divided by the sensitivity factor (given in the Andor iXon EM+ DU-897 performance sheet) of 12.17 electrons per analog/digital count. This factor can be ignored when the inner loop alone is applied.

In order to correct the drift of the beads under the microscope during the 3000 frames long movies, the standard deviations displayed in Fig. 3 were calculated in the following way:

$\left(\frac{1}{3000} \sum_{l=1}^{3000} (f(l) - x_{est}^l)^2\right)^{\frac{1}{2}}$, where x_{est}^l is the estimated x-position in frame l and $f(l) = a \cdot l + b$ is a linear least squares fit to all estimated x-positions and their frame number.

Implementation details for NMLA

Each iteration of the NMLA was performed in a window around the last estimated position. The width of the square window was $6\sigma + 6$ pixels, where σ is the standard deviation of the Gaussian approximating the emitter's PSF. After each iteration, the window was shifted according to the new position estimate. The iteration was continued until the improvement of the log likelihood (as described in [13] in supplementary equation 138) between two outer iterations dropped below a threshold of 0.001 or the maximum number of 30 iterations was reached.

The inner loop was employed in a window like the complete NMLA. In order to solve the optimization task in the M-step of the inner loop, we employed Powell's method [19] implemented in the Apache Commons Mathematics Library.

For the experiments regarding computation time and robustness we used the center of mass of all $I_{k,i}^t$, as initialization in every M-step, provided it yielded an improvement regarding the current optimization task. After this initialization, Powell's method was employed.

After each iteration, we calculated the change of parameters as $\max(|p_k^t - p_k^{t-1}|, \|\vec{\theta}_k^t - \vec{\theta}_k^{t-1}\|)$. The iterations in the inner loop were continued until the change of parameters dropped below a threshold of 0.001 or the maximum of 1000 iterations was reached.

Acknowledgments

We thank Britta Schroth-Diez, Daniel White and Nicola Maghelli for the discussion and help regarding EMCCD microscopy; Stephan Saalfeld, Stefan Gumhold, Ivo Sbalzarini and Pavel Tomancak for comments on the manuscript; Thomas Kühn, Tobias Pietzsch and Johannes Schindelin for advice regarding the software development with Imglib and Fiji; and Ivana Šarić for help with the preparation of the figures.