Supplemental Materials

Molecular Biology of the Cell

Smith et al.

Supplementary Note

Single molecule detection

A hypothesis test for the presence of a single molecule is performed for each camera pixel. In this test H_1 is the presence of a single molecule, i.e. the intensity of the single molecule being non-zero ($\theta_I \neq 0$), and H_0 is the absence of a single molecule, i.e. the intensity being equal to zero ($\theta_I = 0$):

$$H_0: \theta_I = 0,\ \theta_{bg},$$
$$H_1: \theta_I \neq 0,\ \theta_{bg}.$$

The intensity θ_I and background θ_{bg} are generally unknown and have to be estimated. The background needs to be estimated, but does not need to be tested (nuisance parameter), in contrast to the intensity. Values, intensity and background, are found by Maximum Likelihood Estimation (MLE), one estimate for each of the two hypotheses. The Maximum Likelihood values found by this procedure are used as input for the likelihood ratio test. The MLE procedure is described in earlier work (Smith et al., 2010). The position of the hypothetical emitter is fixed at the center of the pixel. Averaging over multiple positions inside the pixel area did not lead to an appreciable improvement. Including the position in the MLE procedure as an additional fit (nuisance) parameter led to a less robust behavior.

The Image Formation Model

The Point Spread Function (PSF) is approximated by a Gaussian distribution, which is known to be a valid approach in the context of 2D single emitter localization(Stallinga and Rieger, 2010):

$$PSF(x,y) = \frac{1}{2\,\pi\,\sigma_0^2}\, e^{-\frac{1}{2\sigma_0^2}[(x-x_0)^2+(y-y_0)^2]}.$$

This PSF must be integrated over the pixel area to arrive at the expected photon count at each pixel k:

$$\mu_k = \theta_I \Delta E(x_k - x_0)\Delta E\left(y_j - y_0\right) + \theta_{bg},$$

with

$$\Delta E(u) = \frac{1}{2}\left[\mathrm{erf}\left(\frac{u+\frac{1}{2}}{\sqrt{2}\,\sigma_0}\right) - \mathrm{erf}\left(\frac{u-\frac{1}{2}}{\sqrt{2}\,\sigma_0}\right)\right],$$

where (x_k, y_k) are the pixel coordinates in unit ($[pixel]$) of pixel k, (x_0, y_0) is the location of the center of the PSF in unit $[pixel]$ and, σ_0 is the PSF width, depending on the numerical aperture ($[NA]$), magnification ($[M]$), pixel size ($[\Delta p]$)and the wavelength of the light ($[\lambda]$).

Generalized Likelihood Ratio Test Statistic

The hypothesis test that best approximates the optimal Neyman-Pearson test (Kay, 1998) is the Generalized Likelihood Ratio Test (GLRT), where the Maximum Likelihood estimates of the parameter vectors ($\hat{\theta}$) for the two hypotheses are used instead of their true values (θ). The GLRT Statistic (T_G) is given by:

$$T_G = 2\log\left(\frac{\max_{\theta_I^1,\theta_{bg}^1}\left[P\left(\theta_I^1,\theta_{bg}^1;x\right)\right]}{\max_{\theta_{bg}^0}\left[P\left(0,\theta_{bg}^0;x\right)\right]}\right) = 2\log\left(\frac{P\left(\hat{\theta}_I^1,\ \hat{\theta}_{bg}^1;x\right)}{P\left(0,\ \hat{\theta}_{bg}^0;x\right)}\right),$$

where $P\left(\hat{\theta}_I^1,\ \hat{\theta}_{bg}^1;x\right)$ is the maximum likelihood of the measured data (x) under hypothesis, H_1 and $P\left(0,\ \hat{\theta}_{bg}^0;x\right)$ is the maximum likelihood of the measured data (x) under hypothesis, H_0. Here $P(.)$ is the likelihood function, which described the noise model of the camera used. We assume Poisson noise, which is a good assumption for EMCCD cameras (Smith et al., 2010), and can be easily be modified for sCMOS (Huang et al., 2013). The two MLE fits that are performed for each pixel must include photon counts from surrounding pixels where the size of this subregion depends on the width of the PSF (σ_0). We choose the size of this subregion as small as possible without jeopardizing the localization precision (Smith et al., 2010), $s = 3(2\sigma_0 + 1)$. For pixels near the border of the image that do not have enough neighbors to fill the subregion no MLE and therefore no GRLT is performed.

Up to now we have calculated the GRLT statistic (T_G) per pixel, but this value by itself does not have a useful meaning. As for any other test statistic its value can be converted into a false positive probability ($\mathrm{P}_{FA} = \mathrm{P}(\mathrm{H}_1;\mathrm{H}_0)$). This is done using the probability distribution of the test statistic. We have found (proof in the next section) that for this specific problem the false positive probability is given by: $P_{FA} = 2CDF(-\sqrt{T_G})$, where CDF is the cumulative distribution function of the standard normal distribution defined by:

$$CDF(x) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{x}{\sqrt{2}}\right)\right].$$

The distribution of the Generalized Likelihood Ratio Test Statistic

Any hypothesis test returns a so-called p-value that measures the likelihood of the current value of the test statistic to be a false positive, i.e. that the value is wrongly classified as signal where it should have been background (probability of H_1 (signal) given H_0 (background)). Using this p-value we decide if the measured difference is significant or not. Apart from the false positive probability, which is the distribution of T_G under the null hypothesis H_0 ($\mathrm{P}_{FA} = \mathrm{P}(\mathrm{H}_1;\mathrm{H}_0)$), we can also calculate the detection probability, which is the distribution of T_G under H_1 ($\mathrm{P}_D = \mathrm{P}(\mathrm{H}_1;\mathrm{H}_1)$). Below we show that the false positive probability is given by: $P_{FA} = 2CDF(-\sqrt{T_G})$.

The GLRT uses estimated parameters instead of the true parameters and therefore we have to derive the distribution of the test statistic based on the properties of the estimator. The GRLT is formulated using MLE which is unbiased (if it exists and is unrestricted) and asymptotically attains, as the number of data points go to infinity ($n \to \infty$), the minimal possible variance in the parameter estimate, the so called Cramer Rao Lower Bound (CRLB)[4]. Using this property we can derive the distribution of our test statistic. The CRLB is given by the inverse of the Fisher information matrix (Kay, 1993):

$$I(\theta) = -E\left[\frac{\partial^2 \log P(\theta; x)}{\partial \theta^2}\right],$$

where the $E[\cdot]$ is the expectation operation and $P(\theta; x)$ is the likelihood of the parameters (θ) given the data (x); and for an MLE that attains the CRLB ($n \to \infty$) we can make the following Taylor expansion around the true value (θ): $\log P(\hat{\theta}; x) = -\frac{1}{2}(\hat{\theta} - \theta)^T I(\theta)(\hat{\theta} - \theta) + c(\theta)$, which is consistent with the fact that in expectation the second derivative results in the Fisher information. This is an important expression, because equivalently:

$$P(\hat{\theta}; x) = P(\theta; x)\exp\left[-\frac{1}{2}(\hat{\theta} - \theta)^T I(\theta)(\hat{\theta} - \theta)\right],$$

which shows us that for large data records ($n \to \infty$) the MLE is normally distributed with a covariance equal to that of the CRLB.

In our problem there are two unknown parameters that have to be estimated (θ_I, θ_{bg}) but only one parameter to test ($\theta_I; H_0: \theta_I = 0, H_1: \theta_I \neq 0$), which means that we can make a separation of θ into the parameters we need to estimate, but do not need to test (nuisance parameters, θ_n) and parameters we need to estimate and test (parameters of interest, θ_t) as the concatenated vector $\theta = [\theta_t \ ; \ \theta_n]$. Here we test for $H_0: \theta_t = \theta_{t_0}, H_1: \theta_t \neq \theta_{t_0}$, where the parameters to be tested are θ_t against θ_{t_0}.We need two MLEs ($\hat{\theta}^0 = [\theta_{t_0}; \hat{\theta}_n^0], \hat{\theta}^1 = [\hat{\theta}_t^1; \hat{\theta}_n^1]$), because there are nuisance parameters that have to be estimate under both hypotheses (H_0, H_1). The GLRT statistic follows as

$$T_G = 2\log\left(\frac{P(\hat{\theta}_t^1, \hat{\theta}_n^1; x)}{P(\theta_{t_0}, \hat{\theta}_n^0; x)}\right).$$

The MLE under H_1 is unrestricted ($\hat{\theta}^1$) and therefore independent of which hypothesis is true, we will find that $E[\hat{\theta}_t^1] = \theta_t$ and $E[\hat{\theta}_n^1] = \theta_n$, however the MLE under H_0 is restricted (constrained to $\theta_t = \theta_{t_0}$) and therefore if H_1 is true we obtain a biased estimate, $E[\hat{\theta}_n^0] \neq \theta_n$, and only if H_0 is true we have that $E[\hat{\theta}_n^0] = \theta_n$. Asymptotically (as $n \to \infty$), we can make the following second order Taylor expansion around the MLE $\hat{\theta}^1$ for $P(\theta_{t_0}, \theta_n; x)$. This expansion is valid because the MLE under H_1 ($\hat{\theta}^1$) will maximize the likelihood independent of which hypothesis is true:

$$P(\theta_{t_0}, \theta_n; x) = P(\hat{\theta}_t^1, \hat{\theta}_n^1; x)\exp\left[-\frac{1}{2}\left(\begin{bmatrix}\hat{\theta}_t^1 \\ \hat{\theta}_n^1\end{bmatrix} - \begin{bmatrix}\theta_{t_0} \\ \theta_n\end{bmatrix}\right)^T I(\hat{\theta}_t^1, \hat{\theta}_n^1)\left(\begin{bmatrix}\hat{\theta}_t^1 \\ \hat{\theta}_n^1\end{bmatrix} - \begin{bmatrix}\theta_{t_0} \\ \theta_n\end{bmatrix}\right)\right].$$

However we need, $P\left(\theta_{t_0}, \hat{\theta}_n^0; x\right)$, instead of $P\left(\theta_{t_0}, \theta_n; x\right)$, which is the maximum of $P\left(\theta_{t_0}, \theta_n; x\right)$ for θ_n. To be able to perform the maximization of $P\left(\theta_{t_0}, \theta_n; x\right)$ to $\theta_n$we introduce a short hand notation, where we factorize the Fisher information matrix according to test and nuisance parameters as we have done for the parameter vector:

$$I\left(\hat{\theta}^1\right) = \begin{bmatrix} \mathrm{I}_{\theta_t\theta_t}\left(\hat{\theta}^1\right) & \mathrm{I}_{\theta_t\theta_n}\left(\hat{\theta}^1\right) \\ \mathrm{I}_{\theta_n\theta_t}\left(\hat{\theta}^1\right) & \mathrm{I}_{\theta_n\theta_n}\left(\hat{\theta}^1\right) \end{bmatrix},$$

the Fisher information matrix is symmetric, and therefore the partitioned Fisher information matrix following symmetry properties: $\mathrm{I}_{\theta_t\theta_t}\left(\hat{\theta}^1\right) = \mathrm{I}_{\theta_t\theta_t}\left(\hat{\theta}^1\right)^T$, $\mathrm{I}_{\theta_n\theta_n}\left(\hat{\theta}^1\right) = \mathrm{I}_{\theta_n\theta_n}\left(\hat{\theta}^1\right)^T$, $\mathrm{I}_{\theta_n\theta_t}\left(\hat{\theta}^1\right) = \mathrm{I}_{\theta_t\theta_n}\left(\hat{\theta}^1\right)^T$. Using this short hand notation and the symmetry properties we obtain the maximum by setting the gradient equal to zero

$$\begin{aligned} \frac{\partial \log\left(P\left(\theta_{t_0}, \theta_n; x\right)\right)}{\partial \theta_n} &= \mathrm{I}_{\theta_n\theta_t}\left(\hat{\theta}^1\right)\left(\hat{\theta}_t^1 - \theta_{t_0}\right) + \mathrm{I}_{\theta_n\theta_n}\left(\hat{\theta}^1\right)\left(\hat{\theta}_n^1 - \theta_n\right) \\ &= 0, \end{aligned}$$

and solving for $\theta_n = \hat{\theta}_n^0$. We find that the maximum is obtained at

$$\hat{\theta}_n^0 = \hat{\theta}_n^1 - \mathrm{I}_{\theta_n\theta_n}\left(\hat{\theta}^1\right)^{-1}\mathrm{I}_{\theta_n\theta_t}\left(\hat{\theta}^1\right)\left(\hat{\theta}_t^1 - \theta_{t_0}\right).$$

After back substitution into $P\left(\theta_{t_0}, \theta_n; x\right)$ we find that

$$\begin{aligned} P\left(\theta_{t_0}, \theta_n; x\right) = P\left(x, \hat{\theta}_t^1, \hat{\theta}_n^1\right)\exp\Bigg[-\frac{1}{2}\left(\hat{\theta}_t^1 - \theta_{t_0}\right)^T \Big[\mathrm{I}_{\theta_t\theta_t}\left(\hat{\theta}^1\right) \\ - \mathrm{I}_{\theta_t\theta_n}\left(\hat{\theta}^1\right)\mathrm{I}_{\theta_n\theta_n}\left(\hat{\theta}^1\right)^{-1}\mathrm{I}_{\theta_n\theta_t}\left(\hat{\theta}^1\right)\Big]\left(\hat{\theta}_t^1 - \theta_{t_0}\right)\Bigg], \end{aligned}$$

which can be simplified using the block inversion lemma because

$$\left[I\left(\hat{\theta}^1\right)^{-1}\right]_{\theta_t\theta_t} = \left[\mathrm{I}_{\theta_t\theta_t}\left(\hat{\theta}^1\right) - \mathrm{I}_{\theta_t\theta_n}\left(\hat{\theta}^1\right)\mathrm{I}_{\theta_n\theta_n}\left(\hat{\theta}^1\right)^{-1}\mathrm{I}_{\theta_n\theta_t}\left(\hat{\theta}^1\right)\right]^{-1}.$$

We observe that after quite some linear algebra we have obtained exactly what we have expected: the covariance of the parameters θ_t is given by the upper right block $[\cdot]_{\theta_t\theta_t}$ of the inverse of the Fisher information matrix which is the CRLB for the parameters (θ_t). Note that the covariance is dependent on both the nuisance and the test parameters.

After substitution of this expression into the GLRT statistic we find that the limit form follows as

$$T_G = 2\log\left(\frac{P\left(\hat{\theta}_t^1, \hat{\theta}_n^1; x\right)}{P\left(\theta_{t_0}, \hat{\theta}_n^0; x\right)}\right)$$

$$= 2\log\left(\frac{P\left(\hat{\theta}_t^1, \hat{\theta}_n^1; x\right)}{P\left(\hat{\theta}_t^1, \hat{\theta}_n^1; x\right)\exp\left[-\frac{1}{2}\left(\hat{\theta}_t^1 - \theta_{t_0}\right)^T \left(\left[I\left(\hat{\theta}^1\right)^{-1}\right]_{\theta_t\theta_t}\right)^{-1}\left(\hat{\theta}_t^1 - \theta_{t_0}\right)\right]}\right)$$
$$= \left(\hat{\theta}_t^1 - \theta_{t_0}\right)^T \left(\left[I\left(\hat{\theta}^1\right)^{-1}\right]_{\theta_t\theta_t}\right)^{-1}\left(\hat{\theta}_t^1 - \theta_{t_0}\right).$$

Asymptotically, $E\left[\left[\hat{\theta}_t^1; \hat{\theta}_n^1\right]\right] = [\theta_t;\ \theta_n]$, and therefore we may replace the values in the Fisher matrix, if H_1 is true:

$$T_G = \left(\hat{\theta}_t^1 - \theta_{t_0}\right)^T \left([I([\theta_t;\ \theta_n])^{-1}]_{\theta_t\theta_t}\right)^{-1}\left(\hat{\theta}_t^1 - \theta_{t_0}\right),$$

and if H_0 is true ($E\left[\left[\theta_{t_0}; \hat{\theta}_n^1\right]\right] = [\theta_{t_0}; \theta_n]$):

$$T_G = \left(\hat{\theta}_t^1 - \theta_{t_0}\right)^T \left(\left[I\left([\theta_{t_0}; \theta_n]\right)^{-1}\right]_{\theta_t\theta_t}\right)^{-1}\left(\hat{\theta}_t^1 - \theta_{t_0}\right).$$

We observe that under all circumstances, independent of the true hypothesis (H_i), we have that $T_G \geq 0$, which is in agreement with what we expect, since the MLE under hypothesis H_1, $\hat{\theta}^1$, always results a higher likelihood ($P\left(\hat{\theta}_t^1, \hat{\theta}_n^1; x\right) \geq P\left(\theta_{t_0}, \hat{\theta}_n^0; x\right)$), as there are additional fit parameters present.

Now that the distribution of the GLRT is obtained using the limit form and based on the fact that the MLE of θ is asymptotically ($n \to \infty$) normally distributed:

$$\hat{\theta}_{t_i}^1 \sim \begin{cases} N\left(\theta_{t_0}, \left[I\left([\theta_{t_0}; \theta_n]\right)^{-1}\right]_{\theta_t\theta_t}\right) & under\ \ H_0 \\ N\left(\theta_t, [I([\theta_t; \theta_n])^{-1}]_{\theta_t\theta_t}\right) & under\ \ H_1 \end{cases},$$

Here i denotes the true hypothesis H_i, $N(\mu, \Sigma)$ denotes the normal distribution with mean μ, the true parameter vector, and covariance Σ, which is the estimation uncertainty.

We continue by observing that the limit form of T_G under both hypotheses is of the form:

$$T_G = x^T \Sigma^{-1} x,$$

where $x \sim N(\mu, \Sigma)$. The covariance matrix Σ is symmetric and positive semi-definite and can hence be factorized as $\Sigma = LL^T$. Defining the variable $z = L^{-T}x$ we find that:

$$T_G = x^T L^{-1} L^{-T} x = z^T z.$$

The variable z is normally distributed with identity covariance. If a random variable z_1 follows a normal distribution with a non-zero mean and identity covariance ($z_1 \sim N(\mu, I)$) then its square ($z_1^T z_1$) follows a non-central chi-square distribution ($\chi_t^2(\delta)$) with t degrees of freedom

($rank(I) = t$) and a non-centrality parameter (Ravishanker and Dey, 2001) ($\delta = \mu^T \mu$). The non-central chi-square distribution simplifies to a central chi-square distribution when the mean of the normal distribution is zero ($z_2 \sim N(0, I) \rightarrow z_2^T z_2 \sim \chi_t^2$). Putting everything together we find that the test statistic follows a non-central chi-square distribution under H_1 and a central chi-square distribution under H_0 (as then the mean of the test parameter is zero):

$$T_G \sim \begin{cases} \chi_t^2 & under\ H_0 \\ \chi_t^2(\delta) & under\ H_1 \end{cases},$$

where the non-centrality parameter (δ) is given by:

$$\delta = \left(\theta_t - \theta_{t_0}\right)^T \left([I([\theta_t; \theta_n])^{-1}]_{\theta_t \theta_t}\right)^{-1} \left(\theta_t - \theta_{t_0}\right).$$

In our particular problem the only test parameter is the intensity of the single molecule ($\theta_t = \theta_I$). This allows for a couple of simplifications and using these we can derive a simple equation for T_G in terms of the normal distribution. This is possible because we can use the fact that the square root of a random variable having a chi-square distribution with one degree of freedom is normally distributed (Ravishanker and Dey, 2001). However, we have to be careful: since T_G is always positive there are two possible values that could be the source of the obtained value of $z = \sqrt{T_G}$. Therefore, the probability of a false positive detection if a significance boundary γ is placed on T_G follows as:

$$\begin{aligned} P_{FA} &= P(T_G > \gamma\,; \mathrm{H}_0) \\ &= P(z > \sqrt{\gamma}\,; H_0) + P\left(z \leq -\sqrt{\gamma}; H_0\right) \\ &= \frac{1}{\sqrt{2\pi}} \left[\int_{-\infty}^{-\sqrt{\gamma}} dz + \int_{\sqrt{\gamma}}^{\infty} dz \right] exp\left(-\frac{z^2}{2}\right) \\ &= 1 + erf\left(-\sqrt{\gamma/2}\right) \\ &= 2CDF\left(-\sqrt{\gamma}\,\right), \end{aligned}$$

where CDF is the so called cumulative distribution function of the standard normal distribution function. Therefore, reversely, the significance boundary can be calculated from the false positive rate by:

$$\gamma = CDF^{-1}(P_{FA}/2)^2$$

Similarly, the detection probability $T_G \sim \chi_1^2(\delta)$ can be transformed into $z = \sqrt{T_G} \sim N(\sqrt{\delta}, 1)$, where the detection probability can be calculated as:

$$\begin{aligned} P_D &= P(T_G > \gamma; H_1) \\ &= P(z > \sqrt{\gamma} - \sqrt{\delta}\,; H_1) + P\left(z \leq -\sqrt{\gamma} - \sqrt{\delta}; H_1\right) \\ &= \frac{1}{\sqrt{2\pi}} \left[\int_{-\infty}^{-\sqrt{\gamma}} dz + \int_{\sqrt{\gamma}}^{\infty} dz \right] exp\left(-\frac{\left(z - \sqrt{\delta}\right)^2}{2}\right) \end{aligned}$$

$$= 1 - \frac{1}{2}\left[erf\left(\frac{-\sqrt{\delta} + \sqrt{\gamma/2}}{\sqrt{2}}\right) - erf\left(\frac{-\sqrt{\delta} - \sqrt{\gamma/2}}{\sqrt{2}}\right)\right].$$

Multiple Comparison Problem

Now we are able to calculate the false positive probability as a function of the outcome of T_G and based on this value we can make the decision if a pixel is significant or not, H_1 versus H_0, respectively. Recall that for instance a $P_{FA} = 0.05$ means that there is a 5% probability that the current value is a false positive (decide H_1, while actually the hypothesis H_0 is true).

For a single test a probability of 5% for a false positive might seem acceptable, however, if we perform many tests simultaneously the actual number of false positives can be extremely large. In our case we have to compute millions of tests (256x256 pixels for 1000 frames already results in more than 65 million hypothesis tests) which results in millions of false positives, which is not acceptable in practice. In the field of statistics this problem is known as the multiple comparison problem (Miller, 1981). In the case of performing multiple hypotheses simultaneously, we would like to have a control on the probability that a positive declared test is false (#false positives/(#true positives+#false positives)), instead of the probability that a test gives a false positive (#false positives/#hypothesis tests).

False Discovery Rate Control

There exist a number of approaches to overcome this multiple comparison problem and they are all based on adjusting the p-values for the number of hypothesis tests executed simultaneously. The method of choice to overcome the multiple comparison problem is the procedure for so-called False Discovery Rate (FDR) control devised by Benjamini and Hochberg (Benjamini and Hochberg, 1995). The FDR is defined as the expected value of the proportion of false positives among total positives (FDP). This FDP is an (unobserved) random variable:

$$FDP = \frac{V}{V + S},$$

where V and S are two random variables with outcome equal to the number of false positives and true positives, respectively. Therefore,

$$FDR = E[FDP] \leq \alpha,$$

where the $E[\cdot]$ is the expectation operation and α is the value at which the FDR is controlled. The proof for independent tests can be found in reference 9, appendix A.

In our case the pixel tests are not independent because the PSF extends over multiple pixels. Therefore we apply a modified version of the FDR(Benjamini and Yekutieli, 2001). The procedure to control the FDR consists of the following steps:

1. The P_{FA} values for the pixels $1,2,\cdots,m$ $(P_1,\ldots,P_m)$ are ordered from smallest p-value to highest: $0 \leq P_{(1)} \ldots \leq P_{(m)}$.
2. Find the largest k for which $P_{(k)} \leq (k/m\, c(m))\alpha$, where m is the total number of tests and $c(m) = \sum_{i=1}^{m}(1/i)$, and α is the FDR.
3. Finally, declare all $H_{(i)}$ significant for $i = 1 \ldots k$ and calculate the adjusted p-values using $P^*_{(k)} = (m\, c(m)/k)P_k$.

This procedure returns the adjusted false positive probabilities, $P^*_{(k)}$, and these adjusted probabilities are declared significant using a user defined significance level which equals the value at which the FDR is controlled. The probabilities below the target FP are the regions where single molecules are detected.

Receiver operating characteristic

For a statistical detection approach like the GLRT it is possible to present an alternative performance measure to Fig. 2 called a receiver operating characteristic (ROC). A ROC is created by plotting the true positive rate against the false positive rate for varying threshold settings. From such ROC curve the sensitive of a detection algorithm can be judged, as the true positive rate for different false positive rates are plotted. For the GLRT the ROC curve is shown in Fig. S1 for varying background (bg) and constant single molecule intensities (I).

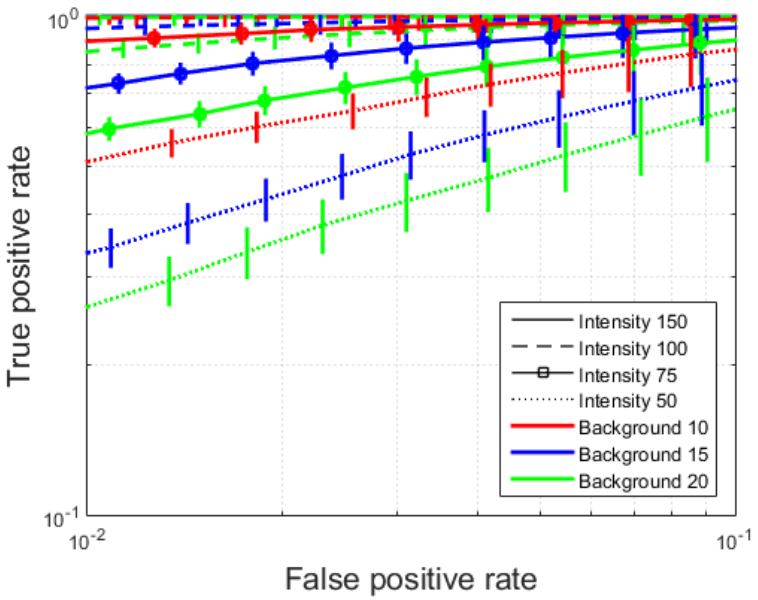

Figure S1 | Receiver operating characteristic (ROC) of the GLRT for varying background (bg) and single molecule intensities (I) relevant to the simulation presented in Fig 1. The average and standard deviations are calculated over 512 samples, where an area of 13x13 [pixel] and $\sigma_{\mathrm{psf}} = 1.39$ [pixel] is used.

Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*. 57:289-300.

Benjamini, Y., and D. Yekutieli. 2001. The control of the false discovery rate in multiple testing under depencency. *The Annals of Statistics*. 29:1165-1188.

Huang, F., T.M. Hartwich, F.E. Rivera-Molina, Y. Lin, W.C. Duim, J.J. Long, P.D. Uchil, J.R. Myers, M.A. Baird, W. Mothes, M.W. Davidson, D. Toomre, and J. Bewersdorf. 2013. Video-rate nanoscopy using sCMOS camera-specific single-molecule localization algorithms. *Nat Methods*. 10:653-658.

Kay, S. 1993. Fundamentals of Statistical Signal Processing. Prentice Hall.

Kay, S.M. 1998. Fundamentals of statistical signal processing. . Signal Processing, Upper Saddle River, NJ.

Miller, R.G. 1981. Simultaneous Statistical Inference. 299 pp.

Ravishanker, N., and K.D. Dey. 2001. A First Course in Linear Model Theory. Chapman and Hall/CRC.

Smith, C.S., N. Joseph, B. Rieger, and K.A. Lidke. 2010. Fast, single-molecule localization that achieves theoretically minimum uncertainty. *Nat Methods*. 7:373-375.

Stallinga, S., and B. Rieger. 2010. Accuracy of the gaussian point spread function model in 2D localization microscopy. *Opt Express*. 18:24461-24476.