



Delft University of Technology

Drift correction in localization microscopy using entropy minimization

Cnossen, Jelmer; Cui, Tao Ju; Joo, Chirlmin; Smith, Carlas

DOI

[10.1364/OE.426620](https://doi.org/10.1364/OE.426620)

Publication date

2021

Document Version

Final published version

Published in

Optics Express

Citation (APA)

Cnossen, J., Cui, T. J., Joo, C., & Smith, C. (2021). Drift correction in localization microscopy using entropy minimization. *Optics Express*, 29(18), 27961-27974. <https://doi.org/10.1364/OE.426620>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.



Drift correction in localization microscopy using entropy minimization

JELMER CNOSSEN,^{1,3} TAO JU CUI,² CHIRLMIN JOO,² AND CARLAS SMITH^{1,4} 

¹*Delft Center for Systems and Control, Delft University of Technology, Delft, Netherlands*

²*Department of BioNanoScience, Kavli Institute of Nanoscience, Delft University of Technology, Delft, Netherlands*

³*j.p.cnossen@tudelft.nl*

⁴*c.s.smith@tudelft.nl*

Abstract: Localization microscopy offers resolutions down to a single nanometer but currently requires additional dedicated hardware or fiducial markers to reduce resolution loss from the drift of the sample. Drift estimation without fiducial markers is typically implemented using redundant cross correlation (RCC). We show that RCC has sub-optimal precision and bias, which leaves room for improvement. Here, we minimize a bound on the entropy of the obtained localizations to efficiently compute a precise drift estimate. Within practical compute-time constraints, simulations show a 5x improvement in drift estimation precision over the widely used RCC algorithm. The algorithm operates directly on fluorophore localizations and is tested on simulated and experimental datasets in 2D and 3D. An open source implementation is provided, implemented in Python and C++, and can utilize a GPU if available.

© 2021 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

1. Introduction

Single molecule localization microscopy (SMLM) [1–5] has provided unique insights into molecular biology by allowing optical microscopy to be used beyond the diffraction limit (~ 250 nm). It is possible to separate the signal of individual fluorescent molecules, by using fluorescent molecules that can individually turn on and off. This way their location is estimated with high precision (~ 10 nm). Many images of the same sample are recorded, processed, and combined into a single super-resolution image.

Slight drift of the sample during acquisition results in a loss of resolution. Therefore, any super-resolution reconstructions need to be compensated for drift during the measurement. This can either be done through active compensation during the measurement or in post-processing. Early work on active drift correction using a fiducial spot created with a pinhole achieved sub-nanometer localization precision [6]. Other methods use fiducial markers embedded in the sample [7,8], or use image processing of bright field imaging from the sample itself to estimate drift [9]. If active drift correction is limited to purely z-drift (along the optical axis), commercially available interferometry autofocus systems are frequently used. This still leaves 2D drift uncorrected, but because drift within the focal plane does not deteriorate the image quality, it can be corrected afterwards without any negative consequences as long as the estimation of the drift is accurate enough. If the sample is not sufficiently static, fiducial markers can be embedded in the sample [10]. For SMLM however, although different sets of fluorophores are visible through time, the underlying structures that are being imaged are often static and thus suitable to use as input for a drift estimation algorithm. Several such approaches have been explored: [11] estimated drift by binning localizations over time, and performed image cross-correlation of the first bin against all remaining bins. Redundant cross correlation [12] (RCC) improves on this by computing the cross-correlations between all bin pairs. Due to the quadratic scaling of the number of bin-pairs, this method is fast if the number of bins remains relatively small

(<20), but can be much slower if a higher time resolution is required. RCC was further extended into 3D SMLM in ZOLA-3D [13]. To our knowledge, further improvements of post-process drift correction over RCC have been limited: [14] demonstrated Bayesian sample drift inference (BaSDI), an optimal approach using Bayesian statistics and expectation-maximization, in which the drift can be estimated per-frame. However, this approach internally uses an image-based representation of the localizations, which means that memory use and compute time do not scale well with either large fields of view, or very precise measurements. Sub-nm precise drift estimation has been demonstrated [15] specifically on DNA-PAINT origami, using markers and taking advantage of the separable clusters and known sparsity of binding sites, but this renders the approach incompatible with larger and denser biological structures such as microtubules. In practice, we note that in the absence of fiducial markers or active drift correction, RCC is currently the method of choice for researchers using SMLM. Despite its widespread use, the precision and bias of RCC and how much room is still left for improvement, is poorly understood.

Here, we propose a new method for drift estimation in both 2D and 3D datasets based on the minimization of a bound of the entropy. Our method, Drift at Minimum Entropy (DME), runs directly on localization data, supports both 2D and 3D drift estimation, and does so with a compute time similar to RCC. The method is tested on DNA-PAINT origami measurements (Fig. 2), simulated SMLM measurements of filaments Fig. 4, experimental STORM imaging of microtubules (Fig. 5) and simulated 3D microtubules (Fig. 6). The outline of the paper is as follows: in section 2.1 to section 2.4, we describe the concept and mathematical definition of the estimation method; in section 2.5 describes preparation and measurement of a DNA-origami sample that was used to compare fiducial markers and localization based drift estimation; section 3 describes the various experiments we performed, comparing the performance of DME, RCC and fiducial markers drift estimation; and section 4 discusses the results and open questions.

2. Methods

2.1. Drift at minimum entropy

A SMLM reconstruction can be regarded as a probability distribution of the true distribution of the localized molecules. The entropy of a random variable is a measure for the level of uncertainty about its outcome. Drift during the measurement will increase the uncertainty of the reconstruction and therefore its entropy. The idea behind DME is that the entropy metric can be used as an optimization goal: We can estimate the drift by choosing drift coefficients such that the entropy is minimized.

The emitter localization estimate has, under mild assumptions, a normal distributed error with respect to its true position [16]. The localization microscopy reconstruction can therefore be expressed as a probability distribution

$$p(\mathbf{r}) = \frac{1}{N} \sum_{i=1}^N p_i(\mathbf{r}), \quad (1)$$

where N is the total number of localizations and $\mathbf{r}$ is a K -dimensional position vector representing the emitter position.

$p_i(\mathbf{r})$ is the probability distribution of a single localization:

$$p_i(\mathbf{r}) = \frac{\exp\left(-\frac{1}{2}(\mathbf{r} - \boldsymbol{\mu}_i + \mathbf{d}(t_i))^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{r} - \boldsymbol{\mu}_i + \mathbf{d}(t_i))\right)}{\sqrt{(2\pi)^K |\boldsymbol{\Sigma}_i|}}, \quad (2)$$

with $\boldsymbol{\mu}_i$ as the estimated position, $\boldsymbol{\Sigma}_i$ the estimated covariance matrix (representing the uncertainty in the position estimation), and $\mathbf{d}(t_i)$ is the drift in frame t of the i -th emitter.

The drift is estimated by minimizing an upper bound on the statistical entropy of $p(\mathbf{r})$. We note that $p(\mathbf{r})$ is a Gaussian Mixture Model (GMM) with equal weights for each component [17]. The entropy of a GMM does not have a closed form solution, but various closed form upper bounds have been found [18–20]. DME uses a variational approximation of the entropy from [20] for its computational simplicity:

$$H(\mathbf{D}) \leq U_H(\mathbf{D}) = \sum_{i=1}^N H_i(\mathbf{D}) - \frac{1}{N} \sum_{i=1}^N \log \left(\sum_{j=1}^N \frac{1}{N} e^{-D_{\text{KL}}(p_i(\mathbf{r})||p_j(\mathbf{r}))} \right), \quad (3)$$

where $D = \{d(1), \dots, d(L)\}$ is the drift at all frames (1 to L), $H(\mathbf{D})$ is the entropy of distribution $p(\mathbf{r})$, $H_i(\mathbf{D})$ is the entropy of a single GMM component $p_i(\mathbf{r})$. The entropy of a single GMM component is given by $H_i(\mathbf{D}) = \frac{1}{2} \ln \det(2\pi e \Sigma_i)$, which will stay constant under drift change. Therefore, the first term of $U_H(\mathbf{D})$ can be ignored during the drift estimation. $D_{\text{KL}}(p_i(\mathbf{r})||p_j(\mathbf{r}))$ is the Kullback-Leibler divergence [20] between the probability distributions for localization i and j . In our implementation, we assume localization errors to be uncorrelated between axes, so Σ is a diagonal matrix. The method is not limited to diagonal Σ matrices, but the calculation is faster and to our knowledge, no localization software package actually stores the full covariance matrix for each localization. $D_{\text{KL}}(i,j)$ can then be expressed as follows:

$$D_{\text{KL}}(p_i(\mathbf{r})||p_j(\mathbf{r})) = \frac{1}{2} \sum_{k=1}^K \left(\log \left(\frac{\sigma_{j,k}^2}{\sigma_{i,k}^2} \right) + \frac{\sigma_{i,k}^2}{\sigma_{j,k}^2} + \frac{(\mu_{i,k} - d_k(t_i) - \mu_{j,k} + d_k(t_j))^2}{\sigma_{j,k}^2} \right) - \frac{K}{2}, \quad (4)$$

Here $\sigma_{i,k}^2$ is the k -th diagonal of the covariance matrix Σ_i for localization i . To compute the covariance matrix we use the Cramer Rao Lower Bound (CRLB), which defines a lower bound on localization precision of a fluorophore [21]. The achieved localization precision is estimated by evaluating the CRLB for the estimated single-molecule parameters [22,23]. This information is used in DME to convert the list of localizations into a probability density function. While Eq. (3) appears to be very computationally demanding, scaling quadratically with the number of localizations, this is effectively not the case in typical SMLM datasets. The exponential term $e^{-D_{\text{KL}}(i,j)}$ goes to zero with increasing distance between localizations. To decide which localizations to include in the inner sum of Eq. (3), we set a fixed distance threshold based on the average CRLB times 3 and ignore any localizations further than that. At that point, assuming equal CRLB, the entropy contribution of a localization pair is at the 3σ level (0.3%). An efficient Kd-tree [24] implementation is used to find localizations within this distance. Whenever the drift estimate has moved more than half the distance threshold, the localization neighbors are recomputed.

Runtimes of DME are on the same order of magnitude as those of RCC, typically less than a minute (see Fig. 2(b)). For large datasets, if compute time is still too long or too much memory is used, the number of localizations can be reduced to include only the brightest spots. The default maximum is set to 1M localizations, which we found to be more than sufficient for drift estimation in all tested datasets.

2.2. Cubic spline based drift parameterization

To reduce the number of estimated parameters, we define the drift $\mathbf{d}(t)$ as a Catmull-Rom spline i.e. a piecewise cubic spline where the value at time t is determined by the 4 control points closest to t . Such a spline function is computationally efficient to evaluate and differentiate, and tunable to the specific dataset by changing the number of control points. Cubic splines are chosen over higher-order splines, as higher order splines would complicate calculation and have a similar effect as simply reducing the frame binning size. To do this, the global frame index t has to be

translated into a local coordinate on the spline τ and an index j , defining which of the spline segments corresponds to frame t . The drift at frame t is defined as:

$$\mathbf{d}(t) = \mathbf{g}(t/b - \lfloor t/b \rfloor, \lfloor t/b \rfloor) \quad (5)$$

where

$$\mathbf{g}(\tau, j) = \sum_{i=1}^4 w_i(\tau) \mathbf{u}_{j+i-2} \quad (6)$$

with

$$\mathbf{w}(\tau) = \begin{bmatrix} \tau^3 & \tau^2 & \tau & 1 \end{bmatrix} \begin{bmatrix} -0.5 & 1.5 & -1.5 & 0.5 \\ 1 & -2.5 & 2 & -0.5 \\ -0.5 & 0 & 0.5 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad (7)$$

here j is the spline segment index, τ is an interpolating fraction between 0 and 1, $\mathbf{u}_j$ is the j -th spline control point, and $w_i(\tau)$ the i -th element of $\mathbf{w}(\tau)$ the Catmull-Rom spline. The matrix coefficients in $\mathbf{w}(\tau)$ follow from the definition of a Catmull-Rom spline [25]. We define M drift spline control points, $\mathbf{u}_j$ with $j \in \{0, M-1\}$ and $M = \lceil L/b \rceil$, with b being the binning size, i.e. the number of frames in a single spline segment and L being the total number of frames.

2.3. Gradient descent

The drift is estimated by finding a set of values for $\mathbf{u}_k$ that minimize $H(\mathbf{D}(\mathbf{u}_k))$ (Eq. (3)), using gradient descent (see Supplement 1). Fig. 1 illustrates the general idea of the algorithm. By dynamically adjusting the gradient step size we optimize the speed of convergence (see Fig. 1). Because DME uses gradient descent in a complex optimization landscape, it is susceptible to converge to a local minimum. To prevent this, we use 2D RCC with a small number of (4 to 10 depending on sparsity) bins to quickly compute a rough initial estimate. In the case of datasets with too few localizations (typically >10000 frames, < 50k localizations), such as microtubules, convergence of the algorithm can be difficult (see Supplement 1, Fig. S4). To further improve the initial RCC estimate, DME can also be performed with an initial coarse estimation step. In this step, Σ_i is increased (typically 2x larger) during drift estimation and using a larger binning size (typically 20x larger). After this coarse estimation, the DME is performed using the true Σ_i . This approach attempts to smooth the optimization landscape and results in a more robust convergence (see Supplement 1).

2.4. Spot detection and localization

All conversion of raw images to 2D and 3D localizations has been done using a custom implementation of conventional SMLM methods. Details can be found in Supplement 1.

2.5. Comparison with BaSDI and RCC

We compared DME with the implementation of RCC in Picasso [26], and BaSDI (using the BaSDI source provided with the paper). DME also uses an internal python based implementation of RCC to create an initial 2D drift estimate, but for fair comparison to existing methods currently in use, we have chosen to use the Picasso implementation, which is widely used and accepted. In this paper, we used the Picasso master branch version of 21-11-2020. The RCC drift estimation was performed by directly invoking the 'picasso undrift' command from our own python code, and subsequently loading the resulting drift values from the text file.

To compare with BaSDI, we used MATLAB 2020a to run BaSDI 1.1. Prior to executing the BaSDI code, the localization frame numbers are divided by frame binning size, and positions are

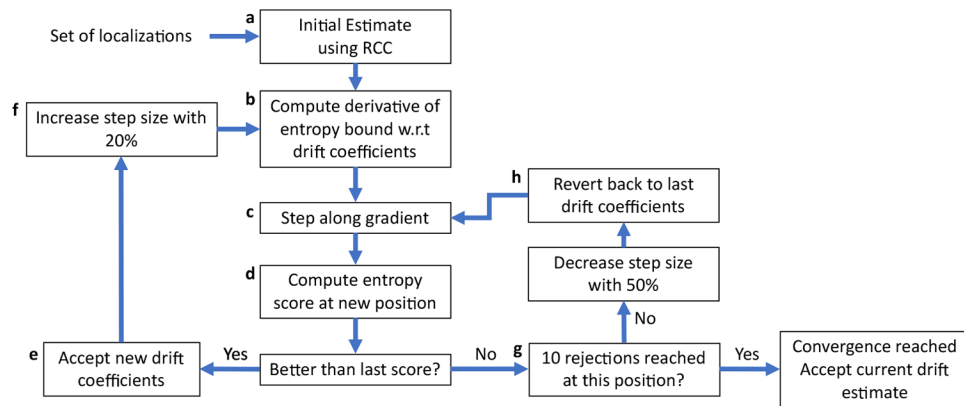


Fig. 1. Schematic representation of the DME algorithm. **a)** The RCC algorithm is used in X,Y with a minimal number of bins to create an initial estimate of drift. This makes DME more robust in cases of large drift displacement, and can alternatively be replaced with DME running on a large bin size. **b).** Given the spline-based definition of drift, we compute the derivative of the entropy upper bound w.r.t all of the spline coefficients that describe the drift. **c)** Using the computed gradient and current step size, we compute a new set of spline coefficients. **d)** The entropy score is defined as upper bound of the entropy, except for the parts that stay constant during change of the drift. If the entropy score has improved, we accept it as a new drift estimate (**e**). **f)** Gradient descent can converge very slow if a sub-optimal step size is used. To maintain a close to optimal step size, we increase it on successful steps and decrease it whenever the entropy score is not improving. **g)** The algorithm is stopped when there are no improvements in the score, even after decreasing the step size 10 times. At that point, we can conclude we have reached a (global or local) minimum in the entropy score.

rescaled (BaSDI uses a discrete pixel grid). Specific details for the BaSDI pixel size can be found in the figure description. BaSDI's 'viterbi' method is used to compute the final drift estimate from posterior drift distributions, without considering a prior for the drift distribution.

2.6. DNA origami experiments

2.6.1. Sample preparation

The DNA origami plate (Tilibit nanosystems, Munich, Germany) was designed using CADNano [27] on a square lattice design as described previously [28]. Assembly occurred in a 100 μ L reaction volume containing 10 nM p8064 scaffold strand, 100 nM core staple strands, 100 nM target handles, 100 nM biotin handles in 1x TE folding buffer with 11 mM MgCl₂. The DNA plate was annealed by setting the thermocycler to 65°C for 10 min, after which a temperature gradient was applied from 60°C to 40°C in steps of 1°C/hour. Subsequently, the assembled plate was separated from the staple, target and biotin handles using an Amicon spin filter (100 kDa). The plates were kept inside T50 buffer (50mM Tris-HCl, pH 8.0, 50 mM NaCl) supplemented with 11 mM MgCl₂ (origami buffer).

Quartz slides were pegylated according to Chandradoss et al. [29] to avoid non-specific interactions. After assembly of microfluidic plates, slides were incubated with streptavidin (0.1 mg/mL, ThermoFisher) for 1 min. followed by washing with origami buffer. DNA origami plates were introduced inside the chamber at a concentration of approx. 50 pM and incubated for 1 min after which the chamber was rinsed with origami buffer. DNA imager strand was subsequently introduced into the chamber with imaging buffer (50 mM Tris-HCl, pH 8.0, 50 mM NaCl, 1 mM

MnCl₂, 5 mM MgCl₂, 0.8% glucose, 0.5 mg/mL glucose oxidase (Sigma), 85 ug/mL catalase (Merck) and 1 mM Trolox (Sigma)).

2.6.2. Microscope setup and data acquisition

Experiments were performed on a custom built prism-TIRF setup. An inverted microscope (IX-73, Olympus) is used with a 532 nm (Sapphire 532-100 mW CW CDRH) laser to excite the DNA imager strand. The resulting fluorescence signal was collected using a 60x water immersion objective (UPLSAPO60XW, Olympus). A longpass filter blocks the excitation light (LDP01 - 532RU - 25, Semrock) after which the fluorescence signal is then projected onto the EM-CCD camera (iXon3, DU - 897 - C00 - #BV, Andor Technology). The images were recorded using Andor Solis. Prior to acquisition, we calibrated the camera photon counts as described in [Supplement 1](#), using 100 frames to estimate the dark camera offset, and 1000 frames to estimate per-pixel variance.

3. Results

3.1. DNA-PAINT based nanorulers

We quantified the performance of DME in a range of experimental conditions and simulations. An experimental comparison was performed on a DNA-PAINT sample that contains both DNA origami nanostructures and fiducial markers (Fig. 2). The recorded images were processed in our localization pipeline (see [Supplement 1](#)), and Picasso Render [26] was used to manually mark all the beads to separate them from the DNA-paint localizations. 18 beads were removed from the set of localizations, and 4 best beads (with the least noisy traces) were used for fiducial marker based drift correction. The non-bead localizations from this experiment were drift-corrected using using RCC, BaSDI and DME at different frame bin sizes. It can be seen that DME is able to reach an equal or better precision as fiducial markers at small bin sizes. BaSDI is also able to do so, but at a much longer compute time and adds discretization error due to discrete grid in which the drift is represented. The estimated precision of localization clusters agrees with the CRLB, and is even slightly lower (potentially due to camera calibration model bias of photon counts and/or PSF model bias). This indicates that for DME, BaSDI, and fiducial markers, the drift is fully corrected and the remaining cluster width can be attributed to localization error. In this particular DNA-PAINT acquisition, RCC will make cluster precision a few nanometer worse, and can correct drift only at a time resolution that is orders of magnitude worse than DME. A visual comparison between the 3 algorithms based on zoomed-in regions can be found in [Supplement 1](#). There, we can see a clear visual improvement for DME over RCC, but visual quality appears the same for BaSDI and DME.

Fig. 2(d) shows a smoother error distribution for DME compared to BaSDI, likely caused by the discretization error of BaSDI. At a binning size of 1 frame/bin, DME shows a wider and smoother distribution than at 5 frames/bin. This is not surprising, considering this dataset has, on average, 30 localizations per frame. The best possible (unbiased) estimate of drift on a particular frame, if you would know exactly which localizations correspond to which binding sites, would then be limited by the combined precision of those localizations ($CRLB/\sqrt{30} \approx 2.4$ nm for 30 localizations, or ~ 1 nm for 5 frames/bin). Interestingly however, the cluster precision in **b** is slightly better at 1 f/b than at 5 f/b. We hypothesize that this could be caused by overfitting: If there are not enough localizations per frame to constrain the optimization problem, it can be that DME finds a set of drift coefficients that improve the cluster precision but are further away from the true drift. This idea is tested further in [Supplement 1](#). Here, we observe that some amount of overfitting occurs, but that it remains minor even for an intentionally sparse dataset with only 1 emitter per frame on average. In more realistic datasets, with >10 emitters per frame, overfitting is likely negligible.

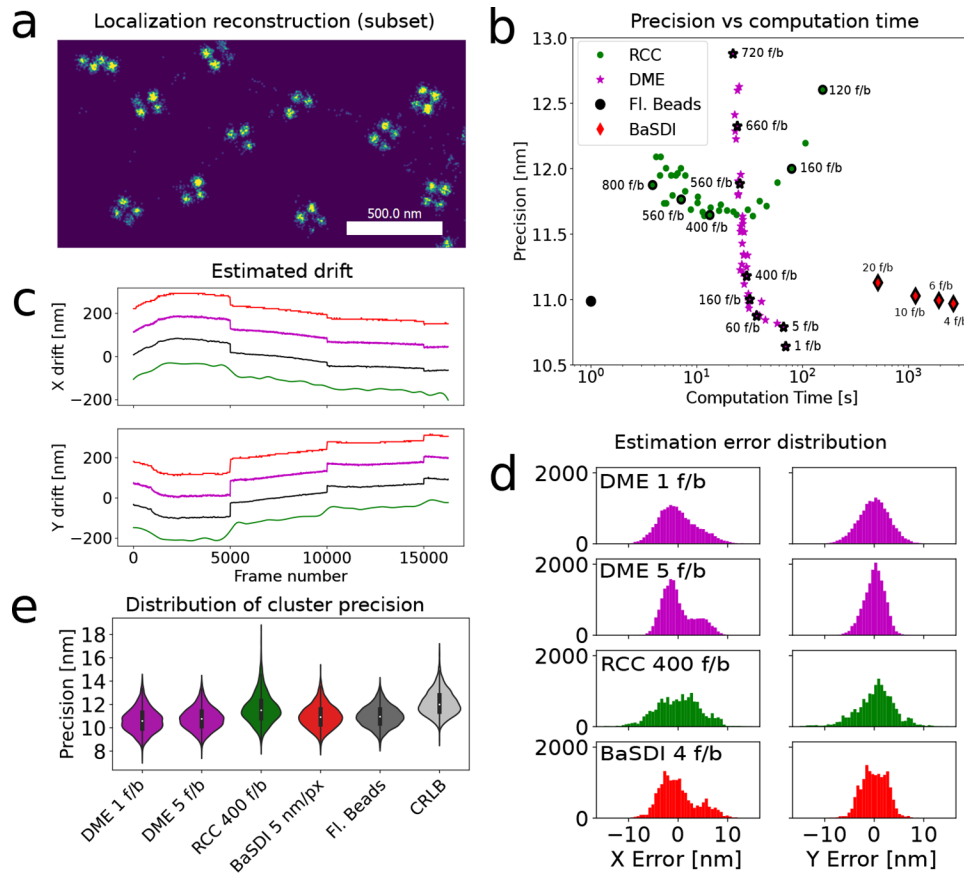


Fig. 2. Drift correction using 3 different methods for a DNA-PAINT experiment done on a TIRF microscope, imaging 16256 frames at 200 ms exposure time. The sample consists of square DNA origami binding sites of 60 nm x 80 nm, with fluorescent tetraspeck beads as fiducial markers. A subset of the binding sites are shown in **a**, visualized using Picasso Render [26]. Converting the image data to localizations was done as described in methods. The DNA-PAINT localizations have an average emitter photon count of 462 photons, with 4.4 photons/pixel background, whereas the bead localizations have an average photon count of 12635 photons, and 28.7 photons/pixel background. The disparity in background is most likely caused by PSF model mismatch with the 2D Gaussian PSF. **b**) compares drift estimation precision with required computation time. The drift estimate from fiducial markers (black) is compared to DME (purple), RCC (green) and BaSDI (red) on the localizations without beads. The dataset was chosen as it has several discrete jumps in it due to pausing of the recording, which is suitable to demonstrate the improved time resolution of the DME algorithm. The precision is estimated from the standard deviation of localization positions corresponding to a binding site (cluster), with the mean standard deviation of the clusters plotted. The plotted drift traces (**c**) are shifted by one pixel each to prevent overlap in the plot. 5 frames/bin was used for DME (showing the sharpest peak in **d**) and 400 frames/bin for RCC (lowest RCC cluster precision in **b**). For BaSDI, a subset of the dataset was used, cropped to 120x120 pixels, and the number of frames was reduced by 4x using binning. 5.414 nm/pixel was used as BaSDI grid spacing. The difference between the estimate from fiducial markers and the estimate from the algorithms (**d**) indicates potential drift bias on the shape of resolved structures in a sample. Noticeable is that in the X axis the DME drift difference does not follow a Normal distribution. Since this shape is also seen in the BaSDI drift, it could be caused by a bias in the fiducial marker localization due to PSF aberrations. Panel **e** shows standard deviation of localizations in 2604 clusters. Clusters with < 30 localizations are discarded.

As it is hard to estimate true drift estimation precision without the ground truth drift, we performed an additional simulation (Fig. 3) with clusters of localizations (similar to DNA-PAINT measurements). Here, it is clear that DME has a better precision versus compute time ratio, and outperforms BaSDI and RCC on a typical SMLM localization reconstruction.

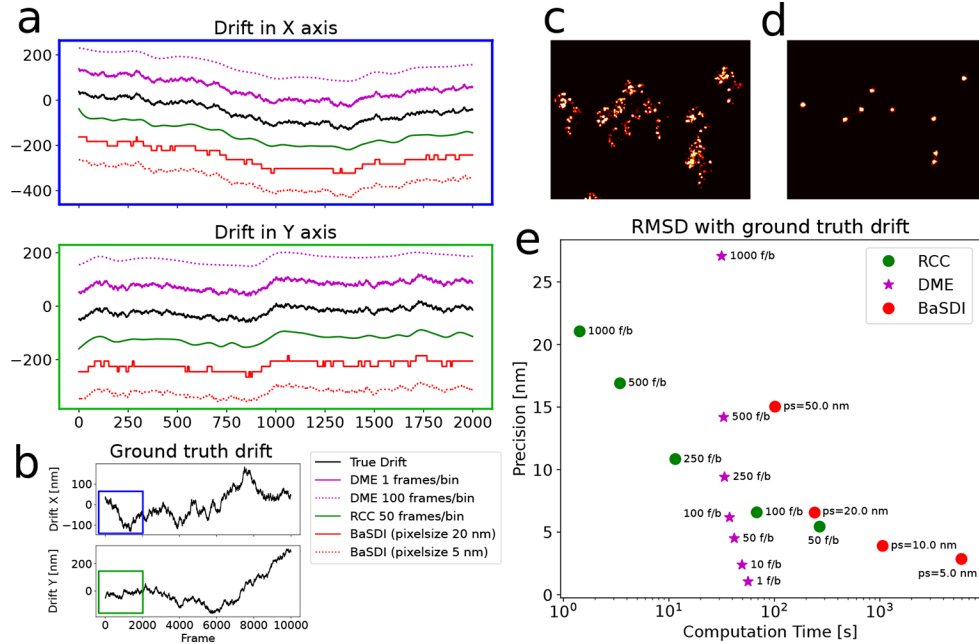


Fig. 3. Drift correction using various algorithms on a simulation of blinking spots at fixed positions, similar to DNA-PAINT measurements. The drift (b) was simulated using a cumulative normal distribution, sampling a new offset from a $\mathcal{N}(\mu = 0.02\text{nm}, \sigma^2 = 4\text{nm}^2)$. A sub-interval of the drift is shown in (a). 1000 binding sites were simulated within an area of 300x300 pixels (100 nm/px), with 3000 photons per localization and 10 photons/pixel background. 10000 frames were simulated resulting in a total of 353901 localizations. A subsection of the localization dataset is shown before drift correction (c), and after drift correction (d). The root-mean-square deviation w.r.t the ground truth drift is plotted versus the required computation time (e). Picasso RCC with bins smaller than 50 frames fails to complete due to too few localizations per bin. It can be seen that with this particular set of parameters, such as image size, localization precision and underlying structure, DME is computationally more efficient than BaSDI and achieves a better estimation precision. The best precision for DME, RCC and BaSDI plotted here are 1.05 nm, 5.44 nm and 2.85 nm resp.

We explored the effect of a mismatch between the CRLB used by DME and the true localization uncertainty. [Supplement 1](#) shows the effect of simulated emitters that produce localization errors with a full covariance matrix (one with significant off-diagonal components), while the DME algorithm is provided with average and isotropic localization errors that do not include correlation between the axes. The resulting increase in drift estimation error remains small even when the assumed localization error already has a strong mismatch with the true localization error. From this we conclude that DME is likely insensitive to a small CRLB mismatch.

3.2. Experimental and simulated microtubules

DNA origami samples as demonstrated in Fig. 2 have a relatively large number of localization events per binding site. Real biological structures typically use a similar number of frames in

the acquisition, but the localizations from these frames are spread over many more fluorescent molecules in the sample. We therefore tested the algorithm performance on simulated microtubule acquisitions, shown in Fig. 4.

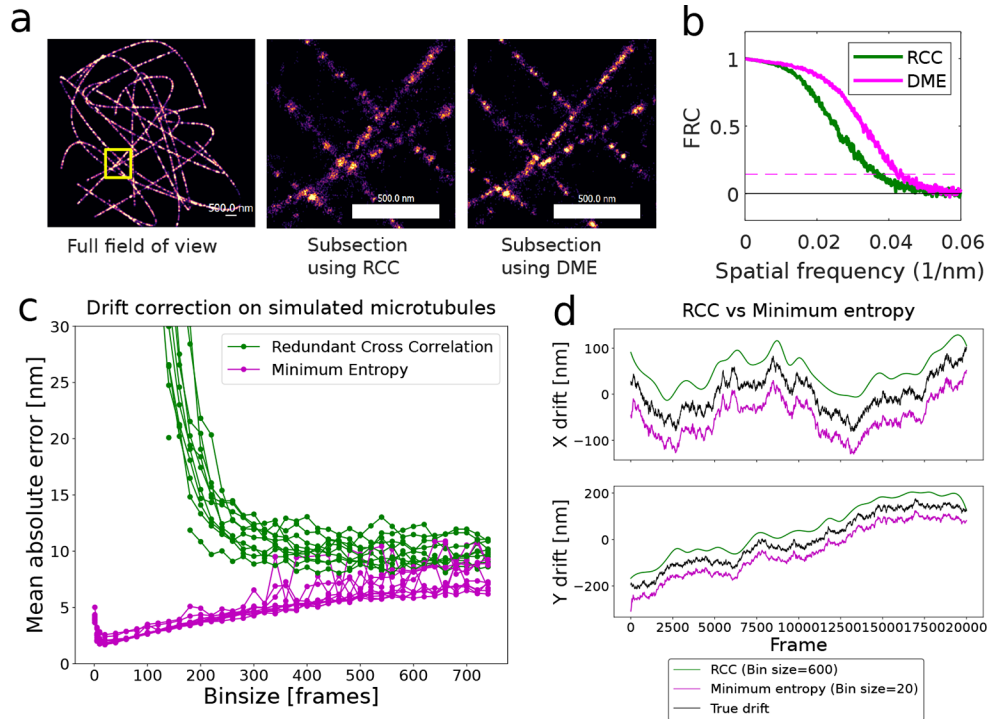


Fig. 4. Drift correction on simulated microtubule measurements. To estimate the true error of drift correction methods, we generated 10 datasets with binding sites positioned along randomly generated 2D splines (see panel **a**). Each dataset has 20000 frames of 100x100 pixels (100 nm/pixel), each sampled from Poisson distributed noise. The simulation was done with 20 photons/pixel of fluorescence background, between 500 and 1500 photons per emitter per on frame, and an average of 10 emitters per frame with an average on-time of 10 frames. Drift was simulated as cumulative normally distributed noise, typically around a few pixels as is common in real measurements. **(c)** For each dataset, both RCC and DME drift correction was performed for a range of binning sizes. The mean absolute error (MAE) in the estimate is plotted, with each line representing one of the 10 datasets. **(d)** The true drift trace plotted next to both estimates. Bin sizes are chosen to be optimal for each the algorithm in terms of MAE (20 fr/bin for DME, 600 fr/bin for RCC). **(b)** The FRC for both RCC and DME is plotted for the same dataset as **(b)**, showing a resolution improvement from 27.15 nm (RCC) to 23.35 nm (DME). **(a)** The dataset used in **(b)** visualized, with a subsection shown in yellow indicating the zoomed in area.

It can be seen that DME outperforms RCC on (sparser) microtubule localization microscopy data: when the number of localizations within one bin becomes a limiting factor for RCC, DME will still perform well. The effect of improved drift correction on this particular dataset can be clearly seen in the two zoom-in sub panels. The visual improvement can also be partly due to the shape of the drift trace, as the image blurring might be more significant for the simulated cumulative Gaussian noise in Fig. 4 than for the thermal drift seen in Fig. 2.

To quantify DME performance on experimental microtubule data, we used a publically available STORM microtubule dataset from SIMFLUX [30]. Results of this are shown in Fig. 5. The dataset is split in two subsets with equal number of localizations. The drift is estimated

independently on each of these subsets, which allows us to compute an estimate of the drift estimation precision. The histograms of the difference between these two drift traces for RCC show a significantly non-Gaussian shape of errors, which could indicate potential artifacts of several nanometers in size. This non-Gaussian distribution is caused by the combination of having few time points in the RCC result, and the spline interpolation to interpolate between those points. In typical STORM measurements with a localization precision of >10 nm these artifacts will be hidden, but they could become apparent for more precise localization microscopy methods.

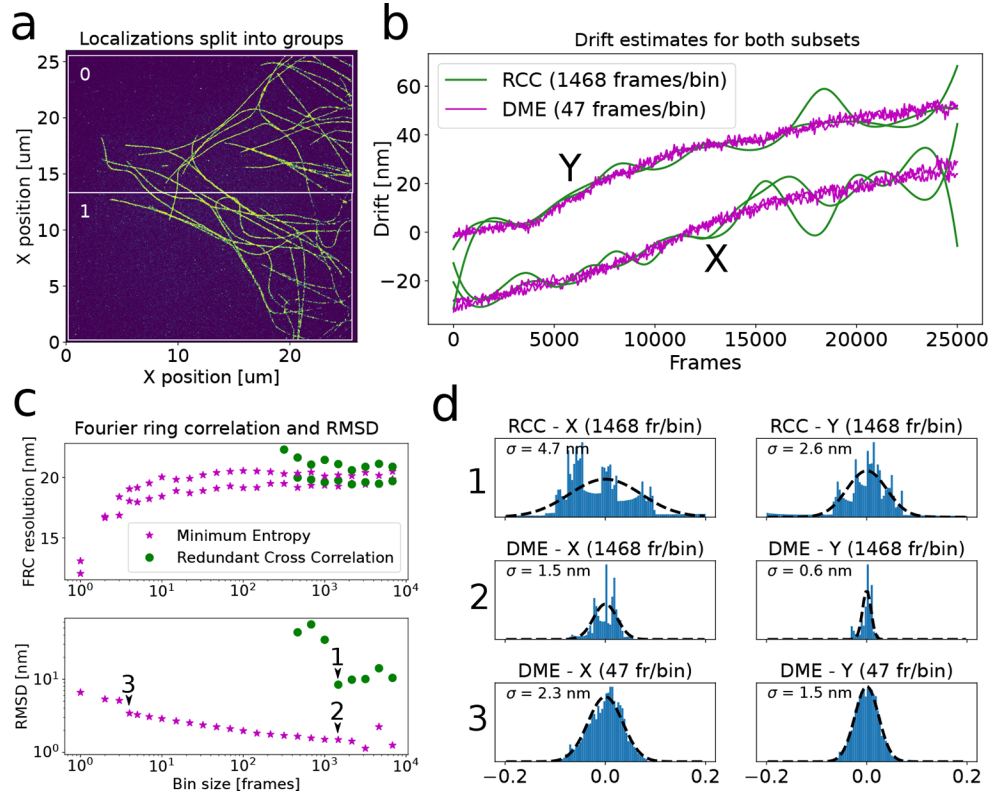


Fig. 5. Drift correction comparison done on a STORM acquisition of microtubules from SIMFLUX [30]. The comparison was done on uniform excitation localizations, by summing every 6 frames to get raw images equivalent to uniform excitation measurements. The resulting localization dataset was split up in two sets of 410401 localizations each (a), and both RCC and DME was performed on each subset using different binning sizes (b). The Fourier ring correlation (FRC [31]) of both subsets, and root mean square difference (RMSD) between the two drift estimates is plotted for a range of drift estimation bin sizes (c). (d) Histograms of the difference between the two drift estimates, for the two methods in both axes. Apparent is that DME errors are normally distributed at small binning sizes (d.3), or have a low variance compared to RCC at equal binning size (d.1 and d.2). RCC is likely creating drift related artifacts on the scale of ~ 2 nm in the result due to errors not being normally distributed.

We have performed a separate comparison including BaSDI in Supplement 1, with additional zoomed-in regions showing visual quality differences. Here, due to BaSDI algorithm scalability, drift estimation is performed on a smaller subset of the data. While cross validation in Fig. 5 shows an improvement in drift estimation for DME over RCC, there does not seem to be a clear

visual improvement of DME over RCC or BaSDI in Supplement 1, Fig. S5. From this we conclude that the visual quality of this dataset is limited by labelling and localization precision, as opposed to drift.

The computed Fourier ring correlations (FRC) show a surprising drop in FRC at low bin sizes. After visual inspection of microtubule structures, no clear improvement can be seen. From this we conclude that the FRC improvement is at least partially due to overfitting. At a binsize

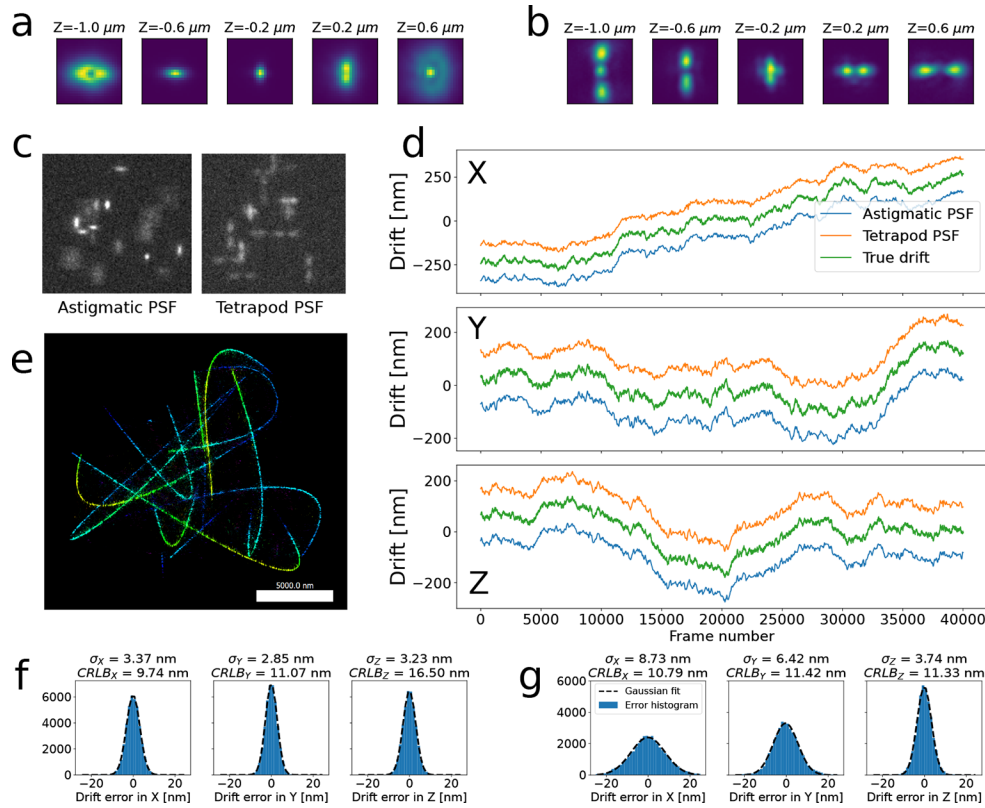


Fig. 6. 3D drift correction on 2 simulated SMLM measurements of microtubules, done the same as in Fig. 4 but spread over a $2 \mu\text{m}$ Z range. The simulation was done with 10 photons/pixel background, and an average of 3000 photons/emitter. Both datasets use the same ground truth structure and drift, but two different experimentally measured PSFs were used: an astigmatic PSF (Tubulin-A647-3D [32]) (a), and a tetrapod PSF from our custom build microscope using a deformable mirror (b). (c) An example of raw frames from the 2 datasets. The simulated camera frames were processed by our 3D localization pipeline as described in Supplement 1, section 3.C. (d) The drift was estimated first using RCC, followed by a coarse estimation using 4x larger localization Σ values and 2000 frames per bin, and finally a precise estimation using 50 frames per bin. (e) 3D rendered version of the tetrapod PSF dataset, using Picasso Render [26] and Z position encoded as color. (f & g) The drift estimation errors plotted in histograms for both astigmatic (f) and tetrapod (g) datasets, showing a clear normal distribution for the drift estimation errors. $\sigma_{x,y,z}$ and $\text{CRLB}_{x,y,z}$ indicate the standard deviation of drift estimation error and mean CRLB of individual localizations, resp. The difference in drift error can be attributed to different numbers of localizations in the datasets. Due to difference in size and shape of the PSF, spot detection efficiency varies between the datasets and the astigmatic dataset has 241910 localizations whereas the tetrapod data has 144443.

of 1 frame, each frame on average has only around 16 localizations. At such low counts, we argue that the DME algorithm will move localizations directly on top of each other and assign a non-physical drift that coincides with an FRC improvement. The increase in drift estimation precision as measured by the RMSD (Fig. 5(c)) around low binsizes also supports this conclusion. Fig. 4 possibly also shows overfitting of the DME algorithm, noticeable by the peak in mean absolute error at low binsizes.

The application of DME on 3D drift estimation is demonstrated in a simulated experiment, shown in Fig. 6, where two experimentally measured PSFs were used to generate 2 3D microtubule datasets. In 3D drift estimation, we again computed an initial estimate in 2D using Picasso's drift estimation implementation. Then, we applied a coarse drift estimation by running DME with a larger CRLB assigned to the localizations, which smoothes the optimization landscape and prevents the Z drift estimation getting stuck in a local minima. One caveat of this approach is that the number of neighboring localizations increases quickly with an increasing CRLB, so it quickly increases compute time and memory use. This could alternatively be replaced with a 3D estimate using RCC as done in ZOLA-3D.

4. Discussion and conclusion

On experimental DNA-PAINT origami measurements, DME shows an approximate five fold improvement over RCC in drift estimation precision on DNA-PAINT nanorulers, and is able to achieve the same localization precision as fiducial marker based drift correction (Fig. 2(b)). We show that DME achieves a better drift estimation precision and is able to do so at a much smaller frame binning size (Fig. 5(d)), indicating a higher time resolution. In a typical SMLM reconstruction, a binning size of around 50 frames per bin is optimal for DME, versus 500 for RCC (Fig. 4(c)). Depending on the drift in the system, this can result in an improvement in image quality as measured by the FRC (Fig. 4(b)) and localization cluster precision (Fig. 2(b)). Simulated SMLM acquisitions demonstrate that DME is able to perform precise 3D drift estimation as well, at a near-isotropic 3D precision (Fig. 6, panel f and g). Compared to the BaSDI drift estimation algorithm, DME achieves a similar or better drift estimate with a much lower compute time (Fig. 2(b)). We find that performance of the drift estimation depends very much on the characteristics of the datasets. While BaSDI performs very well on small, sparse datasets, we found it to be prohibitively slow on realistic larger datasets.

The improved performance of DME on sparser datasets also makes it feasible to estimate the precision of the drift estimation itself by cross validation (Fig. 5(c)). Our implementation reports this using the RMSD between the drift estimates of the dataset split two-fold, which is used to inspect the quality of the DME output and tune the bin size. We observed that the drift error increases both at low binsizes due to low number of localizations per frame and potential overfitting (see Fig. 3(c) and Supplement 1), and at high binsizes (see Fig. 3(c)). Therefore, we recommend to the user to search for a binsize that minimizes the RMSD.

Drift estimation in SMLM is an inverse problem with potentially many local minima. As a result, DME's gradient descent approach requires parameter tuning (especially on datasets with few localization events per binding site). We have explored this in Supplement 1, showing that the initial estimate cannot be too far away from the ground truth. In this particular simulation, initial estimates above 20nm precision would already be unable to fully converge. We expect that the algorithm's robustness could be improved further, perhaps using smoothing of the optimization landscape or by divide-and-conquer approaches. Additionally, the minimum entropy metric can potentially be used for other SMLM applications, such as alignment of different sample FOVs in image stitching, or multicolor channel alignment.

In conclusion, we developed a novel algorithm for drift estimation on point-based datasets using a minimum entropy optimization metric. We anticipate that for many experiments, especially those with sparse binding sites as found in Fig. 2, DME can replace fiducial markers and achieve

the same image resolution, simplifying sample preparation. The method scales well to 3D drift estimation, and with 3D SMLM becoming more popular and mainstream, we expect this algorithm to find widespread practical use.

Funding. Nederlandse Organisatie voor Wetenschappelijk Onderzoek (16761, 740.018.015).

Acknowledgments. J.C. and C.S.S. were supported by the Netherlands Organisation for Scientific Research (NWO), under NWO START-UP project no. 740.018.015 and NWO Veni project no. 16761. J.C. designed the algorithm, performed measurements and simulations. T.J.C. created the combined fiducial marker + DNA-PAINT samples and performed measurements. J.C., T.J.C. and C.S. wrote the paper with input from C.J.

Disclosures. The authors declare no conflicts of interest.

Data availability. Code and example data is available in Ref. [33]. Original image data is available in Refs. [33] and [34].

Supplemental document. See [Supplement 1](#) for supporting content.

References

1. 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**(5793), 1642–1645 (2006).
2. S. T. Hess, T. P. Girirajan, and M. D. Mason, "Ultra-high resolution imaging by fluorescence photoactivation localization microscopy," *Biophys. J.* **91**(11), 4258–4272 (2006).
3. M. J. Rust, M. Bates, and X. Zhuang, "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)," *Nat. Methods* **3**(10), 793–796 (2006).
4. M. Heilemann, S. Van De Linde, M. Schüttel, R. Kasper, B. Seefeldt, A. Mukherjee, P. Tinnefeld, and M. Sauer, "Subdiffraction-resolution fluorescence imaging with conventional fluorescent probes," *Angew. Chem., Int. Ed.* **47**(33), 6172–6176 (2008).
5. M. Lelek, M. T. Gyparakis, G. Belu, F. Schueder, J. Griffié, S. Manley, R. Jungmann, M. Sauer, M. Lakadamyali, and C. Zimmer, "Single-molecule localization microscopy," *Nat. Rev. Methods Primers* **1**(1), 39 (2021).
6. A. Pertsinidis, Y. Zhang, and S. Chu, "Subnanometre single-molecule localization, registration and distance measurements," *Nature* **466**(7306), 647–651 (2010).
7. S. H. Lee, M. Baday, M. Tjioe, P. D. Simonson, R. Zhang, E. Cai, and P. R. Selvin, "Using fixed fiducial markers for stage drift correction," *Opt. Express* **20**(11), 12177–12183 (2012).
8. H. Ma, J. Xu, J. Jin, Y. Huang, and Y. Liu, "A Simple Marker-Assisted 3D Nanometer Drift Correction Method for Superresolution Microscopy," *Biophys. J.* **112**(10), 2196–2208 (2017).
9. R. McGorty, D. Kamiyama, and B. Huang, "Active microscope stabilization in three dimensions using image correlation," *Opt. Nanoscopy* **2**(1), 3 (2013).
10. A. Balinovic, D. Albrecht, and U. Endesfelder, "Spectrally red-shifted fluorescent fiducial markers for optimal drift correction in localization microscopy," *J. Phys. D: Appl. Phys.* **52**(20), 204002 (2019).
11. M. J. Mlodzionoski, J. M. Schreiner, S. P. Callahan, K. Smolková, A. Dlasková, J. Šantorová, P. Ježek, and J. Bewersdorf, "Sample drift correction in 3D fluorescence photoactivation localization microscopy," *Opt. Express* **19**(16), 15009–15019 (2011).
12. Y. Wang, J. Schnitzbauer, Z. Hu, X. Li, Y. Cheng, Z.-L. Huang, and B. Huang, "Localization events-based sample drift correction for localization microscopy with redundant cross-correlation algorithm," *Opt. Express* **22**(13), 15982–15991 (2014).
13. A. Aristov, B. Lelandais, E. Rensen, and C. Zimmer, "ZOLA-3D allows flexible 3D localization microscopy over an adjustable axial range," *Nat. Commun.* **9**(1), 2409 (2018).
14. A. Elmokadem and J. Yu, "Optimal Drift Correction for Superresolution Localization Microscopy with Bayesian Inference," *Biophys. J.* **109**(9), 1772–1780 (2015).
15. M. Dai, R. Jungmann, and P. Yin, "Optical imaging of individual biomolecules in densely packed clusters," *Nat. Nanotechnol.* **11**(9), 798–807 (2016).
16. C. S. Smith, S. Stallings, K. A. Lidke, B. Rieger, and D. Grunwald, "Probability-based particle detection that enables threshold-free and robust in vivo single-molecule tracking," *Mol. Biol. Cell* **26**(22), 4057–4062 (2015).
17. M. R. B. Clarke, R. O. Duda, and P. E. Hart, "Pattern Classification and Scene Analysis," *J. Royal Stat. Soc. Ser. A (General)* **137**(3), 442 (1974).
18. J. L. Durrieu, J. P. Thiran, and F. Kelly, "Lower and upper bounds for approximation of the Kullback-Leibler divergence between Gaussian mixture models," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, (2012).
19. M. F. Huber, T. Bailey, H. Durrant-Whyte, and U. D. Hanebeck, "On entropy approximation for Gaussian mixture random vectors," in *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, (2008).
20. J. R. Hershey and P. A. Olsen, "Approximating the Kullback Leibler divergence between Gaussian mixture models," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, (2007).

21. R. J. Ober, S. Ram, and E. S. Ward, "Localization Accuracy in Single-Molecule Microscopy," *Biophys. J.* **86**(2), 1185–1200 (2004).
22. C. S. Smith, N. Joseph, B. Rieger, and K. A. Lidke, "Fast, single-molecule localization that achieves theoretically minimum uncertainty," *Nat. Methods* **7**(5), 373–375 (2010).
23. C. S. Smith, K. Jouravleva, M. Huisman, S. M. Jolly, P. D. Zamore, and D. Grunwald, "An automated Bayesian pipeline for rapid analysis of single-molecule binding data," *Nat. Commun.* **10**(1), 272 (2019).
24. J. L. Bentley, "Multidimensional Binary Search Trees Used for Associative Searching," *Commun. ACM* **18**(9), 509–517 (1975).
25. E. Catmull and R. Rom, "Class of Local Interpolating Splines," (1974).
26. J. Schnitzbauer, M. T. Strauss, T. Schlichthaerle, F. Schueder, and R. Jungmann, "Super-resolution microscopy with DNA-PAINT," *Nat. Protoc.* **12**(6), 1198–1228 (2017).
27. S. M. Douglas, A. H. Marblestone, S. Teerapittayanon, A. Vazquez, G. M. Church, and W. M. Shih, "Rapid prototyping of 3D DNA-origami shapes with caDNAno," *Nucleic Acids Res.* **37**(15), 5001–5006 (2009).
28. M. Filius, T. J. Cui, A. N. Ananth, M. W. Docter, J. W. Hegge, J. Van Der Oost, and C. Joo, "High-Speed Super-Resolution Imaging Using Protein-Assisted DNA-PAINT," *Nano Lett.* **20**(4), 2264–2270 (2020).
29. S. D. Chandross, A. C. Haagsma, Y. K. Lee, J. H. Hwang, J. M. Nam, and C. Joo, "Surface passivation for single-molecule protein studies," *J. Visualized Exp.* (2014).
30. J. Cnossen, T. Hinsdale, R. Thorsen, M. Siemons, F. Schueder, R. Jungmann, C. S. Smith, B. Rieger, and S. Stallinga, "Localization microscopy at doubled precision with patterned illumination," *Nat. Methods* **17**(1), 59–63 (2020).
31. R. P. Nieuwenhuizen, K. A. Lidke, M. Bates, D. L. Puig, D. Grünwald, S. Stallinga, and B. Rieger, "Measuring image resolution in optical nanoscopy," *Nat. Methods* **10**(6), 557–562 (2013).
32. Y. Li, M. Mund, P. Hoess, J. Deschamps, U. Matti, B. Nijmeijer, V. J. Sabinina, J. Ellenberg, I. Schoen, and J. Ries, "Real-time 3D single-molecule localization using experimental point spread functions," *Nat. Methods* **15**(5), 367–369 (2018).
33. J. Cnossen, T. J. Cui, C. Joo, and C. Smith, "Example code for "Drift correction in localization microscopy using entropy minimization"," *GitHub* (2021), <https://github.com/qnano/drift-estimation>.
34. J. Cnossen, T. J. Cui, C. Joo, and C. Smith, "Image data associated with "Drift correction in localization microscopy using entropy minimization"," *figshare* (2021), <https://doi.org/10.4121/14995089>.