Measuring localization confidence for quantifying accuracy and heterogeneity in single-molecule super-resolution microscopy

Hesam Mazidi[a], **Tianben Ding**[a], **Arye Nehorai**[a], **and Matthew D. Lew**[a,1]

[a]Department of Electrical and Systems Engineering, Washington University in St. Louis, MO 63130

Single-molecule localization microscopy (SMLM) measures the positions of individual blinking molecules to reconstruct images of biological and abiological structures with nanoscale resolution. The attainable resolution and accuracy of various SMLM methods are routinely benchmarked using simulated data, calibration "rulers", or secondary imaging modalities. However, these methods cannot quantify the nanoscale imaging accuracy of any particular SMLM dataset without ground-truth knowledge of the sample. Here, we show that by measuring estimation stability under a well-chosen perturbation and with accurate knowledge of the imaging system, we can robustly quantify the confidence of every individual localization within an experimental SMLM dataset, without ground-truth knowledge of the sample. We demonstrate our broadly-applicable method, termed Wasserstein-induced flux (WIF), in measuring the accuracy of various reconstruction algorithms directly on experimental data of microtubules and amyloid fibrils. We further show that the estimated confidences or WIFs can be used to evaluate the experimental mismatch of computational imaging models, enhance the accuracy and resolution of reconstructed structures, and discover sample heterogeneity due to hidden molecular parameters.

localization accuracy, statistical confidence, localization software, model mismatch, Wasserstein distance

Single-molecule localization microscopy (SMLM) has become an important tool for resolving nanoscale structures and answering fundamental questions in biology (1–5) and materials science (6, 7). SMLM uses repeated localizations of blinking fluorescent molecules to reconstruct high-resolution images of a target structure. In this way, quasi-static features of the sample are estimated from noisy individual images captured from a fluorescence microscope. These quantities, such as fluorophore positions (i.e., a map of fluorophore density), "on" times, emission wavelengths, and orientations, influence the random blinking events that are captured within an SMLM dataset. By using a mathematical model of the microscope, SMLM reconstruction algorithms seek to estimate the most likely set of fluorophore positions and brightnesses (i.e., a super-resolution image) that is consistent with the observed noisy images.

A key question left unresolved by existing SMLM methodologies is: How well do the SMLM data support an algorithm's statistical estimates comprising a super-resolved image, i.e., what is our statistical confidence in each measurement? Intuitively, one's interpretation of an SMLM reconstruction could dramatically change knowing how trustworthy each localization is.

Existing metrics for assessing SMLM image quality can be categorized broadly into two classes: those that require knowledge of the ground-truth positions of fluorophores (8, 9), and those that operate directly on SMLM reconstructions alone, possibly incorporating information from other measurements (e.g., diffraction-limited imaging) (10, 11). One popular approach is the Jaccard index (JAC) (8, 12), which measures localization accuracy, but has limited applicability for SMLM experiments as it requires exact knowledge of ground-truth molecule positions. Therefore, data-driven methods have been proposed to quantify the reliability of a localization without knowing the ground truth (13). A drawback of these methods, however, is their reliance on a user to identify accurate localizations versus inaccurate ones, which suffers from low throughput and poor accuracy in low signal-to-noise-ratio (SNR) datasets.

Methods that quantify performance by analyzing SMLM reconstructions exploit some aspect of prior knowledge of the target structure or SMLM data. Calculating a Fourier ring coefficient (FRC) utilizes correlations within SMLM datasets to measure image resolution with the expectation that SMLM reconstructions should be stable upon random partitioning (10). However, the FRC cannot detect localization biases that result in systematic distortions in the SMLM reconstruction. In contrast, other methods quantify errors between a pixelated SMLM image and a reference image, which is taken as a ground truth (11). While these methods are able to provide summary or aggregate measures of performance, none of them directly measure the accuracy of *individual* localizations. Such knowledge is critical for harnessing fully the power of SMLM for scientific discovery.

Here, we leverage two fundamental insights of the SMLM measurement process: 1) we possess highly-accurate mathematical models of the imaging system, and 2) we know the precise statistics of noise within each image. This knowledge, when combined with an analysis algorithm, enable us to assess quantitatively the confidence of each individual localization within an SMLM dataset without knowledge of the ground truth. With these confidences in hand, the experimenter may filter unreliable localizations from SMLM images without removing accurate ones necessary to resolve fine features. These confidences may also be used to detect mismatches in the mathematical imaging model that create image artifacts (14), such as misfocusing of the microscope, dipole-induced localization errors (15–17), and the presence of optical aberrations (18–20).

In the present paper, we describe the principles and operation of our method, Wasserstein-induced flux (WIF), whose

[1]To whom correspondence should be addressed. E-mail: mdlew@wustl.edu

underlying algorithm is built on the theory of optimal transport (21). Given a certain mathematical imaging model, WIF reliably quantifies the confidence of individual localizations within an SMLM reconstruction. We show that these confidences yield a consistent measure of localization accuracy under various imaging conditions, such as changing molecular density and optical aberrations, without knowing the ground-truth positions *a priori*. We demonstrate that our WIF confidence map outperforms other image-based methods in detecting artifacts in high-density SMLM while revealing detailed and accurate features of the target structure. We then quantify the accuracy of various algorithms on real SMLM images of microtubules. Finally, we demonstrate the benefits of localization confidences to improve reconstruction accuracy and image resolution in super-resolution Transient Amyloid Binding imaging (22) under low SNR. Notably, WIF reveals heterogeneities in the interaction of Nile red molecules with amyloid fibrils (23).

Problem statement

In SMLM we may model a variety of physical influences on stochastic fluorescence emission using a *hidden* variable $\boldsymbol{\beta}$ (Fig. 1A). For example, $\boldsymbol{\beta}$ can encode where molecules activate, how densely they are activated, or how freely they rotate. For each frame, we may represent a set of N activated molecules as $\mathcal{M} = \sum_{i=1}^{N} s_i \delta(\boldsymbol{\eta} - \boldsymbol{\eta}_i)$, where $\boldsymbol{s}_i > 0$ and $\boldsymbol{\eta}_i \in \mathbb{R}^d$ represent the brightness and related physical parameters (i.e., a d-dimensional object space comprising position, orientation, etc.) of the ith molecule, respectively. In general, N, s_i, and $\boldsymbol{\eta}_i$ are random variables whose probability distributions depend on $\boldsymbol{\beta}$. We assume that the measured images of molecular blinks $\boldsymbol{g} \in \mathbb{R}^m$ (i.e., m pixels of photon counts captured by a camera) are generated according to a statistical model with the negative log likelihood $\mathcal{L}(q, \mathcal{M}; \boldsymbol{g})$ (see Methods and Fig. 1A). Here, q is the point spread function (PSF, or the image of an SM) of the microscope that can depend on $\mathcal{M}$. In typical SMLM, an algorithm $\mathcal{A}$ equipped with a PSF model, $\hat{q}$, is used to estimate molecular positions. Let us denote the output of such localization algorithm by $\hat{\mathcal{M}} = \sum_{i=1}^{\hat{N}} \hat{s}_i \delta(\boldsymbol{\eta} - \hat{\boldsymbol{\eta}}_i)$, where $\hat{\boldsymbol{\eta}}_i = \hat{\boldsymbol{r}}_i$ represents the estimated positions. Generally, $\boldsymbol{\beta}$ affects the accuracy with which an algorithm localizes molecules, and uncertainty in $\boldsymbol{\beta}$ can cause degraded image resolution or even bias in estimating $\mathcal{M}$. This uncertainty may arise from miscalibration of the PSF model due to optical aberrations as well as neglecting the full molecular parameters $\boldsymbol{\eta}_i$ that affect the PSF q, e.g., the dipole emission pattern of fluorescent molecules (16, 17, 24). A more subtle uncertainty may arise for difficult measurements even with a well-calibrated PSF, e.g., overcounting or undercounting molecules due to image overlap (8, 12).

As $\boldsymbol{\beta}$ is often hidden in an SMLM experiment, we must estimate the degree of uncertainty or confidence of each localization in truly representing $\mathcal{M}$ (Fig. 1B). For 2D SMLM, the fitted width $\hat{\sigma}$ of the standard PSF is commonly used; if $\hat{\sigma}$ is significantly smaller or larger than the expected width of $\hat{q}$, then the corresponding localization has low confidence. However, such a strategy fails when a localization is not a single molecule (SM), but in fact two or more closely-spaced ones. As an illustrative example, we consider two scenarios: an isotropic molecule located at $(x = 0, y = 0, z = 0)$ (Fig. 1C) and two close molecules located at $(0, 0, 0)$ and $(0, 70\text{ nm}, 0)$

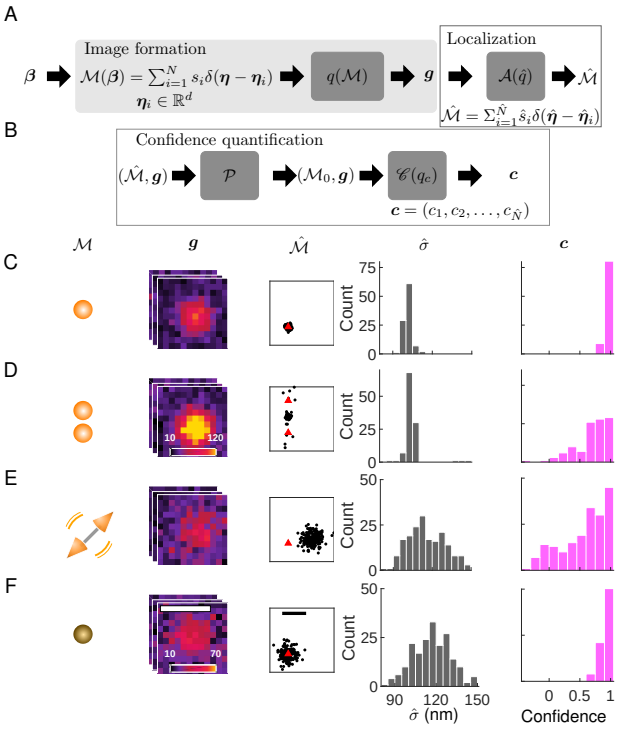

Fig. 1. Quantifying confidence in single-molecule localization microscopy (SMLM). (A) Image formation and localization. Here, $\boldsymbol{\beta}$ is a hidden variable that describes parameters that affect molecular fluorescence, including blinking rates, molecular density, etc. For each frame, activated molecules are represented by $\mathcal{M}$ in which N, s_i, and $\boldsymbol{\eta}_i$ denote number of molecules, photons emitted, and related physical parameters of the ith molecule, respectively. q denotes the PSF of the imaging system that can vary with $\mathcal{M}$. $\boldsymbol{g} \in \mathbb{R}^m$ represents the vectorized image quantifying the number of photons detected consisting of m pixels. Localization refers to estimating $\hat{\mathcal{M}}$ from $\boldsymbol{g}$ via an algorithm $\mathcal{A}$ that uses a PSF model $\hat{q}$. (B) Proposed confidence quantification framework. $\mathcal{P}$ is a perturbation operator that applies a small distortion to $\hat{\mathcal{M}}$. The perturbed molecules $\mathcal{M}_0$ and the measurements $\boldsymbol{g}$ are then analyzed via a confidence analysis algorithm $\mathscr{C}$ that uses its own PSF model q_c. The estimated confidences are represented by $\boldsymbol{c} = (c_1, c_2, \ldots, c_{\hat{N}})$ taking values within 1 (highest confidence) and -1 (lowest confidence). (C) Example of localizing and quantifying confidence using 100 simulated images of an isotropic SM analyzed by ThunderSTORM (TS). Scatter plot: localizations (black dots) and the true positions of molecules (red triangles). Black histogram: fitted widths of the PSF ($\hat{\sigma}$) estimated by TS. Magenta histogram: estimated confidences using the proposed method. (D) Similar to (C) but for two closely-spaced molecules. (E) Similar to (C) but for focused, dipole-like molecule. (E) Similar to (C) but for a dim isotropic molecule. Colorbars: photons per 58.5×58.5 nm^2. Scalebars: (C-F) left: 500 nm, right: 50 nm.

(Fig. 1D). We use ThunderSTORM (TS (25)) to localize the molecules, which also provides fitted widths $\hat{\sigma}$. Due to significant image overlap, TS almost always localizes one molecule for both scenarios, such that in the latter, the estimated positions exhibit a significant deviation from the true ones (Fig. 1C,D). However, the distributions of $\hat{\sigma}$ in both cases are virtually identical, suggesting that $\hat{\sigma}$ is a poor method for quantifying confidence and detecting localization errors due to overlapping molecules (Fig. 1C,D).

More fundamentally, mismatches in SMLM between model and measurement generally depend on $\boldsymbol{\beta}$ in a way that cannot be quantified via simple image-based features such as PSF width. We illustrate this situation by localizing a rotationally fixed molecule located at $(0, 0, 200\text{ nm})$. The anisotropic emission pattern induces a significant bias in TS localizations (Fig. 1E). The distribution of fitted widths is noisy due to photon-shot noise and broadening of the PSF (Fig. 1E). Unfortunately,

this rather wide distribution is comparable to that of a dim, isotropic molecule whose localizations have no systematic bias (Fig. 1F). These observations suggest that quantifying subtle model mismatches in SMLM and, thus, localization confidence, requires a new mathematical "metric."

Proposed methodology

Our goal is to quantify the confidence c_i of each localization in $\hat{\mathcal{M}}$ produced by a given algorithm, given the measurements $\boldsymbol{g}$ (Fig. 1B). In simplest form, we can formulate the localization task as minimizing the negative log-likelihood $\mathcal{L}$ of observing an unknown number of molecules, N, each with a photon count s_i and a position $\boldsymbol{r}_i$. Obviously, if we know N, then, localization task reduces to simultaneously fitting $\{s_1, \boldsymbol{r}_1, \ldots, s_N, \boldsymbol{r}_N\}$. The difficulty stems from not knowing N *a priori*, which renders the localization as a *non-convex* optimization. Non-convexity makes algorithms susceptible to being practically trapped in a saddle point of the negative log-likelihood landscape, while correct localizations are associated with its *global minima* (26). In our example of localizing two closely-spaced molecules (Fig. 1D), almost all position estimates lie near $(0, 35\text{ nm})$, which exactly matches the centroid of the two true molecules. At the same time, the photon count estimates are twice as large as the ground-truth photons, and this point represents a saddle point of the negative log-likelihood (Fig. S1A).

A pivotal observation here is that these saddle points are unstable (in the sense of being a minimizer of the negative log-likelihood) upon a well-chosen perturbation. Put differently, for an accurate localization, the negative log-likelihood surface has a convex curvature as a function of the estimated position of a molecule. Therefore, if we locally perturb the position as well as photon count of a particular estimated molecule, relaxing this perturbation along the likelihood surface will most likely result in a localization very "close" to the unperturbed one. On the other hand, for an unreliable localization, we expect that the negative log-likelihood landscape changes arbitrarily in a local neighborhood (Fig. S1B). As a result, relocalizing most likely will *alter* the original localization. The stability in the position of a molecule upon a careful perturbation is precisely what we denote as the *quantitative confidence* of an SMLM localization. Motivated by this observation, we devise a robust method to measure the stability and therefore statistical confidence of *each localization* within an SMLM dataset.

Localization stability for measuring confidence. Intuitively, stability is a measure of discrepancy between a source point and a perturbed instance of this point after following a certain trajectory. To clarify, consider a strongly convex, differentiable function f over some open set $\Omega \in \mathbb{R}$ taking its minimum at $\omega^* \in \Omega$. Since we are mostly interested in minimizers of some functional, as they are in a sense the best "fit" to the ground truth, we think of the confidence of a point estimate $\hat{\omega}$ as a measure of its distance to ω^*. Since ω^* is unknown, we seek to measure the confidence of $\hat{\omega}$ without knowing ω^*. To this end, we construct a simple single-step gradient-descent update and find a representation of stability to quantify the said confidence.

Consider the following gradient descent update given by the gradient-descent step with a small step size $\epsilon > 0$:

$$\omega_1 = \omega_0 - \epsilon \nabla f(\omega_0), \quad \omega_0 = \mathcal{P}(\hat{\omega}), \tag{1}$$

where ω_0 is a local perturbation of $\hat{\omega}$ according to the operator $\mathcal{P}(\hat{\omega}) = \hat{\omega} + (1 - 2e)\Delta\hat{\omega}$ with $e \sim Bern(0.5)$ and perturbation distance $\Delta\hat{\omega} = |\hat{\omega} - \omega_0|$. Eq. (1) describes the movement of ω_0 in the gradient vector field, ∇f, transporting ω_0 in the direction of decreasing f. If the estimate $\hat{\omega}$ is stable, we have $|\omega_1 - \hat{\omega}| < |\omega_0 - \hat{\omega}|$ as a result of our gradient-decent update, while for an unstable estimate, we can find a perturbation that results in $|\omega_1 - \hat{\omega}| > |\omega_0 - \hat{\omega}|$. Since ω^* is the minimizer of f, we have $|\omega_1 - \omega^*| < |\omega_0 - \omega^*|$ for *any* local perturbation of ω^*. In other words, the gradient vector field pushes the perturbed point ω_0 *toward* ω^*. This observation tells us that we may quantify the confidence of $\hat{\omega}$ by measuring the average convergence of ω_0 toward $\hat{\omega}$. We may define the confidence of a point $\hat{\omega}$ simply as

$$c = \frac{\mathbb{E}\left\{\operatorname{sgn}\left[(\hat{\omega} - \omega_0) \cdot (\omega_1 - \omega_0)\right] \cdot |\epsilon \nabla f(\omega_0)|\right\}}{\mathbb{E}\left[|\epsilon \nabla f(\omega_0)|\right]}, \tag{2}$$

where $\mathbb{E}$ denotes expectation over random perturbations and $\operatorname{sgn}(x)$ takes the sign of a real number x. We call c in Eq. (2) the normalized gradient flux, for reasons that become apparent later. A stable point has the maximum inward gradient flux, i.e., $c = 1$, while an unstable point has some degree of outward gradient flux, i.e., $c < 1$. Thus, c represents a confidence score for any point in Ω without knowing ω^*. As an example, for $f(\omega) = \omega^2$ thus implying $\omega^* = 0$, we find $c = \frac{2\Delta\hat{\omega}}{|\hat{\omega} - \Delta\hat{\omega}| + |\hat{\omega} + \Delta\hat{\omega}|}$. Obviously, $\hat{\omega} = \omega^* = 0$ is the most stable point with highest confidence, and the further away $\hat{\omega}$ is from 0, the worse the confidence.

We can gain more insight if we consider the recursive variational form of Eq. (1) as

$$\omega_k = \arg\min_{\omega \in \Omega} \left\{ \frac{1}{2} \|\omega - \omega_{k-1}\|_2^2 + \epsilon_k f(\omega) \right\}, \quad k > 0. \tag{3}$$

Informally, Eq. (3) defines a discrete trajectory $\{\omega_k\}$ by minimizing f while preserving a "local Euclidean distance" constraint. In the limit of $\epsilon_k \to 0$, i.e., considering continuous trajectories, we recover the *Cauchy Problem*, that is, $\frac{d\omega(t)}{dt} = -\nabla f(\omega(t))$, which defines the evolution of $\omega \in \Omega$ from an initial point ω_0. The resulting curve $\{\omega(t)\}_{t \geq 0}$ is called a *gradient flow*.

Wasserstein-induced flux. Molecular brightnesses $\boldsymbol{s}_i > 0$ and positions $\boldsymbol{r}_i \in \mathbb{R}^2$ in a single SMLM frame are expressed as $\mathcal{M} = \sum_{i=1}^{N} s_i \delta(\boldsymbol{r} - \boldsymbol{r}_i)$, which is a multi-parameter *distribution* in the space of non-negative finite measures $\mathrm{M}(\mathbb{R}^2)$. To extend our discussion to SMLM, we must define the distance between two candidate "guesses" $\mathcal{S}$ and $\mathcal{Q} \in \mathrm{M}(\mathbb{R}^2)$ for molecular parameters. We utilize the elegant theory of optimal transport, where roughly speaking, the optimal transport distance between any two measures is the minimum cost of transporting mass from one to the other as measured via some ground metric (21). The Wasserstein distance is particularly suitable, because its ground metric is simply Euclidean distance. The type-2 Wasserstein distance between two measures $\mathcal{S}, \mathcal{Q} \in \mathrm{M}(\mathbb{R}^2)$ is defined as

$$\mathbb{W}(\mathcal{S}, \mathcal{Q}) = \min_{\pi \in \Pi(\mathcal{S}, \mathcal{Q})} \sqrt{\int_{\mathbb{R}^2 \times \mathbb{R}^2} \|\boldsymbol{r} - \boldsymbol{r}'\|_2^2 \, d\pi(\boldsymbol{r}, \boldsymbol{r}')}, \tag{4}$$

where $\Pi(\mathcal{S}, \mathcal{Q})$ is the set of all couplings or *transportation plans* between $\mathcal{S}$ and $\mathcal{Q}$ satisfying a mass-conservation constraint (21). Equipped with Wasserstein distance, let us re-write the recursive dynamics of Eq. (3) as

$$\mathcal{S}_k = \underset{\mathcal{S}\in \mathrm{M}(\mathbb{R}^2)}{\arg\min} \left\{ \frac{1}{2}\mathbb{W}^2(\mathcal{S}, \mathcal{S}_{k-1}) + \epsilon_k \mathcal{L}(\mathcal{S}) \right\}, \quad k > 0, \tag{5}$$

where $\mathcal{L} : \mathrm{M}(\mathbb{R}^2) \to \mathbb{R}$ is the negative Poisson log-likelihood, which is a convex functional.

We recall that our goal is to obtain a useful representation of point stability in the space of measures. Since stability is coupled with evolution of the measures, we analyze the properties of Wasserstein gradient flows, i.e., $\{\mathcal{S}\}_{t\geq 0}$. A set of intriguing results from the theory of Wasserstein gradient flow assert that if $\mathcal{S}$ has a smooth density, 1) there exists a unique transport map, $T_k : \mathrm{M}(\mathbb{R}^2) \to \mathrm{M}(\mathbb{R}^2)$, such that $\mathbb{W}_2^2(\mathcal{S}, \mathcal{S}_{k-1}) = \int_\Omega |T_k(\boldsymbol{r}) - \boldsymbol{r}|^2 \, d\mathcal{S}_{k-1}(\boldsymbol{r})$; that is, the mass-weighted displacement distance for the transport plan is given by the type-2 Wasserstein distance; and 2) the *backward* velocity field $\mathbf{v}(\boldsymbol{r})$, i.e., the ratio between the displacement $T(\boldsymbol{r}) - \boldsymbol{r}$ and the time step ϵ_k, obtained in the transport of $\mathcal{S}_k$ to $\mathcal{S}_{k-1}$ is given by $\nabla\big(\frac{\delta\mathcal{L}}{\delta\mathcal{S}}(\mathcal{S})\big)(\boldsymbol{r})$ (in the limit of $\epsilon_k \to 0$) (27). The functional $\frac{\delta\mathcal{L}}{\delta\mathcal{S}}(\mathcal{S})$ is the so-called *first variation* of $\mathcal{L}$. In fact, the gradient flows satisfy the *continuity equation* (27):

$$\frac{\partial \mathcal{S}}{\partial t} - \nabla \cdot \left(\mathcal{S} \nabla \big(\frac{\delta \mathcal{L}}{\delta \mathcal{S}}(\mathcal{S}) \big) \right) = 0. \tag{6}$$

We now invoke the *divergence theorem* and define the *Wasserstein-induced flux* (WIF) corresponding to the perturbation volume around a molecule $\mathcal{V}$ as:

$$\begin{aligned} \mathrm{WIF} &\triangleq \int_{\mathcal{V}} \left(\nabla \cdot \left(\mathcal{S} \nabla \big(\frac{\delta \mathcal{L}}{\delta \mathcal{S}}(\mathcal{S}) \big) \right) \right) d\mathcal{V} \\ &= \int_{\mathfrak{S}} \left(\mathcal{S} \nabla \big(\frac{\delta \mathcal{L}}{\delta \mathcal{S}}(\mathcal{S}) \big) \cdot \boldsymbol{n} \right) d\mathfrak{S}, \end{aligned} \tag{7}$$

where $\mathfrak{S}$ and $\boldsymbol{n}$ represent the closed surface on the boundary of $\mathcal{V}$ and its normal vector, respectively. We posit that WIF serves as a mathematically-grounded representation for stability that accounts for local interactions of point sources on the likelihood surface. We note source molecules have various brightnesses; photons are the conserved mass under our perturbation. We therefore normalize WIF w.r.t. the flux associated with an isolated source in $\mathcal{V}$. Henceforth, we denote WIF as its normalized quantity, which means that it takes on values within $[-1, 1]$ with 1 representing the maximum statistical confidence.

As stated previously, we can write the *gradient field* $\nabla\big(\frac{\delta\mathcal{L}}{\delta\mathcal{S}}(\mathcal{S})\big)(\boldsymbol{r}) = \mathbf{v}(\boldsymbol{r}) \approx [T_1(\boldsymbol{r}) - \boldsymbol{r}]/\epsilon$ in transporting $\mathcal{S}_1$ to $\mathcal{S}_0$. This equivalence effectively gives us a strategy to approximate WIF by finding an estimate of T_1, that is, the transport map. Unfortunately, it is computationally expensive to solve for the infinite-dimensional measure $\mathcal{S}_1$. In addition, molecules are in actuality point sources, which means that our object space $\mathrm{M}(\mathbb{R}^2)$ consists of discrete measures and not smooth densities. Even though the uniqueness condition of the transport map requires measures with smooth densities, we show that even with these approximations, our WIF dynamics mirror those predicted by Eq. (6). We designed an efficient, iterative algorithm to approximately compute T_1, which ultimately allows us to compute WIF (SI section 2, Fig. S2) using 1) raw SMLM images of blinking molecules and 2) a computational model of the imaging system (Fig. 1B).

Returning to our earlier examples (Fig. 1C-F), we see that when estimated localizations are close to the ground-truth positions, their estimated confidences or WIFs are concentrated close to 1 (Fig. 1C,F). On the other hand, for inaccurate estimates, localization confidences become significantly smaller, implying their unreliability (Fig. 1D,E). Note that knowledge of the ground truth molecule location is not needed to compute these confidence values.

Results

Localization confidence of an isolated molecule. To test our confidence metric, we analyze images of fluorescent molecules, generated using a vectorial image formation model (24), having various hidden physical parameters such as defocus and rotational mobility. These analyses characterize not only how well WIF measures mismatches introduced by these parameters, but also its limitations due to statistical shot noise, especially for low photon counts. As a baseline, we fix the PSF model in our confidence analysis to that of an isotropic molecule with zero defocus. To determine our confidence metric's robustness to shot noise, we use RoSE to localize an isotropic emitter from 200 noisy, independent realizations of its image for a wide range of detected photons. Computing WIF for these localizations, we observe that the confidences are mostly close to 1 for all photon counts, taking values in [0.95, 1] (Fig. S3). There is a slight reduction in estimated confidences for large photon counts, most likely due to the first-order approximation in our PSF model (SI section 2B).

Next, we quantify how hidden variables that are not accounted for within the model affect the confidences. For a dim molecule (800 photons and 20 background photons/pixel) at modest defocus values ($z \in [0, 200$ nm$]$), we observe that the confidences mostly remain above 0.9 (Fig. S4A,B). As defocus increases beyond 200 nm, approximately 50% of localizations exhibit confidence lower than 0.9. In particular, for $z = 300$ nm, the median confidence decreases to 0.62, a reduction of approximately 40% from $z = 0$ (Fig. S4A). Our confidence metric is remarkably more sensitive to defocus compared to estimates of normalized PSF width (w.r.t. the PSF width at focus), which fluctuate mostly within 10% of their nominal values. For $z = 300$ nm, the median width reduces somewhat counter-intuitively by 13% from its nominal value (Fig. S4A), most likely because of the low SNR.

To explore how shot noise affects WIF and width estimates, we consider a bright molecule (2000 photons) (Fig. S4D). Interestingly, as soon as the defocus increases beyond 140 nm, the confidences sharply drop below 0.9 such that at $z = 200$ nm the median confidence approaches 0.3. In contrast, normalized width estimates remain mostly within 5% of their nominal values with their medians consistently close to 1 (Fig. S4C). Therefore, WIF even detects subtle defocus-induced model mismatches for brighter molecules with sufficient SNRs.

Lastly, we study how well WIF can quantify dipole-induced imaging errors, further exacerbated by defocus. We consider a molecule inclined at 45° with respect to the optical axis and with various degrees of rotational motion: effectively unconstrained or isotropic (uniform rotation within a cone of half angle $\alpha = 90°$), moderate confinement ($\alpha = 30°$), and

strong constraint ($\alpha = 15°$) (Fig. S5C). For a photon count of 1000, notably, we observe consistent decreases in median confidences (below 0.85) for both $\alpha = 30°$ and $\alpha = 15°$ across all z, while for the isotropic molecule, the median confidence drops below 0.9 only for z greater than 160 nm. In addition, confidences for $\alpha = 15°$ are smaller than those of $\alpha = 30°$, which shows our confidence metric's consistency, trending smaller as the degree of mismatch increases (Fig. S5A). On the other hand, normalized width estimates are practically indistinguishable for all α and z values (Fig. S5B).

We next consider a brighter molecule (2000 photons) (Fig. S6C), and observe that confidences for both $\alpha = 30°$ and $\alpha = 15°$ significantly decrease below 0.5 for almost all z positions (Fig. S6A). Surprisingly, the normalized width estimates for $\alpha = 30°$ and $\alpha = 15°$ converge to their nominal (in focus) value as z approaches 200 nm (Fig. S6B). Therefore, confounding molecular parameters (e.g., a defocused dipole-like emitter) may cause estimates of PSF width to appear unbiased, while our WIF metric consistently detects these image distortions and yields small confidence values, resulting in a quantitative, interpretable measure of image trustworthiness. Collectively, these analyses demonstrate that WIF provides a consistent and reliable measure of localization confidence for various forms of experimental mismatches.

Quantifying localization accuracy without ground truth. Assume we are given a series of camera frames of blinking SMs and the corresponding set of localizations returned by an arbitrary algorithm. Our main objective is to assess the trustworthiness of each localization and to quantify the aggregate accuracy of the said algorithm. Obviously, the difficulty is that, in practice, we cannot access an oracle that knows the ground-truth positions of the SMs for comparison.

We propose average confidence $\mathrm{WIF}_{\mathrm{avg}}$ as a novel metric for quantifying the collective accuracy of these localizations (with confidences $\{c_1, \ldots, c_N\}$):

$$\mathrm{WIF}_{\mathrm{avg}} \triangleq \frac{1}{N} \sum_{i=1}^{N} c_i. \tag{8}$$

We can gain insight into Eq. (8) by examining its correspondence to the well-known Jaccard index (JAC), which uses an oracle to determine the credibility of a localization based on its distance to the ground-truth SM. In particular, we may define $\mathrm{JAC} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FN} + \mathrm{FP})$, where TP, FN, and FP denote number of true positives, false negatives, and false positives, respectively. An undetected molecule, that is, a false negative, would increase the denominator of JAC, thereby reducing its value. We posit that this same undetected molecule adversely affects the confidence of a nearby localized molecule, thereby reducing $\mathrm{WIF}_{\mathrm{avg}}$. This intuitive connection between JAC and $\mathrm{WIF}_{\mathrm{avg}}$ suggests that the average confidence may serve as a good surrogate for localization accuracy.

Using $\mathrm{WIF}_{\mathrm{avg}}$, we quantify the performance of two algorithms, RoSE (28) and ThunderSTORM (25), for localizing emitters at various blinking densities (defined as number of molecules per $\mu\mathrm{m}^2$, see Methods) (Fig. 2A,B). Examining the localizations returned by the algorithms, we calculate the Jaccard index using ground-truth information from the oracle and $\mathrm{WIF}_{\mathrm{avg}}$ using only the simulated images of SM blinking (see Methods). For both RoSE and TS, we observe excellent agreement between $\mathrm{WIF}_{\mathrm{avg}}$ and Jaccard index for densities as high as 5 mol./$\mu\mathrm{m}^2$. For higher densities, $\mathrm{WIF}_{\mathrm{avg}}$ monotonically decreases at a rate differing from that of Jaccard index. For instance, at high densities JAC for TS saturates to 0.1, whereas $\mathrm{WIF}_{\mathrm{avg}}$ further decreases due to high FN and low TP, thereby demonstrating the non-convexity of the negative log-likelihood landscape (Fig. 2C).

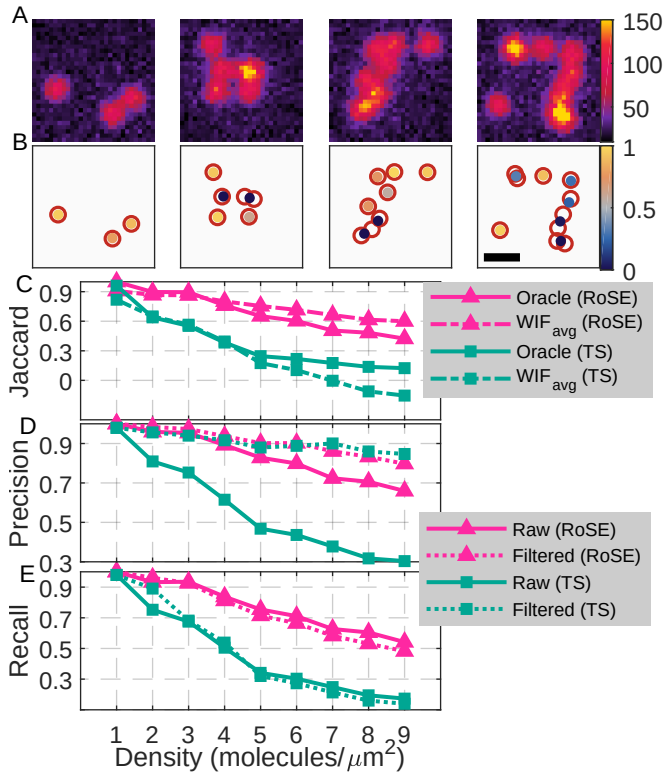

Fig. 2. Wasserstein-induced flux ($\mathrm{WIF}_{\mathrm{avg}}$) quantifies localization accuracy without ground truth. (A) From left to right: images of molecules for blinking densities of 3, 5, 7, and 9 mol. per $\mu\mathrm{m}^2$, respectively. (B) RoSE localizations (colored dots represent calculated confidence) corresponding to images in (A). Open red circles represent ground-truth positions. (C) Jaccard index for RoSE (solid, red) and TS (solid, green) at various blinking densities. The dashed lines represent $\mathrm{WIF}_{\mathrm{avg}}$ for RoSE (red) and TS (green). For each blinking density, 200 independent realizations were used. (D) Precision for all localizations (solid) and localizations with confidence greater than 0.5 (dotted) using RoSE (red) and TS (green), respectively. (E) Recall for all localizations (solid) and localizations with confidence greater than 0.5 (dotted) using RoSE (red) and TS (green), respectively. Colorbars: (A) photons per 58.5×58.5 nm^2; (B) confidence. Scalebar: 500 nm.

A natural application of our confidence metric is to remove localizations with poor accuracy. We filter localizations with confidence smaller than 0.5, corresponding to half of the perturbed photons "returning" toward a particular localization, and calculate the resulting precision $= \mathrm{TP}/(\mathrm{TP} + \mathrm{FP})$ and recall $= \mathrm{TP}/(\mathrm{TP} + \mathrm{FN})$. If the filtered localizations truly represent false positives, we expect to see an increase in precision and a relatively unchanged recall after filtering. Our results show a precision enhancement as high as 180% for TS and a desirable increase of 23% for RoSE (density= 9 mol./$\mu\mathrm{m}^2$) (Fig. 2D). Remarkably, these improvements come with a negligible loss in recall (13% in the worst case) across all densities for both algorithms (Fig. 2E). Overall, these simulation studies show that $\mathrm{WIF}_{\mathrm{avg}}$ is a reliable means of quantifying localization accuracy without having access to ground-truth molecular parameters.

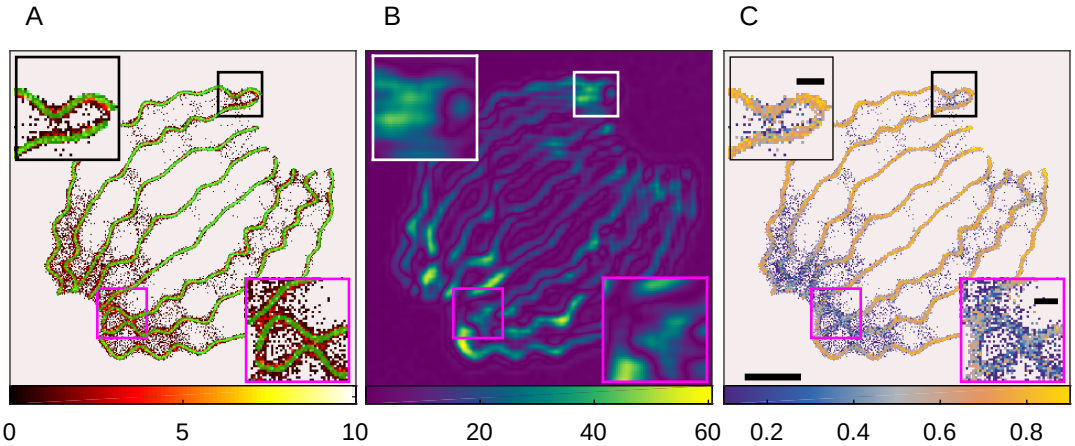

Fig. 3. Confidence map reveals artifacts in recovering a tubulin network from high-density SMLM data. (A) Recovered structure (red) using FALCON overlaid with the ground truth (green). (B) Error map recovered by SQUIRREL (brighter colors correspond to larger errors). (C) Confidence map (brighter colors indicate higher confidence) obtained by averaging localization confidences in each pixel. Colorbars: (A) number of localizations, (B) error, and (C) confidence per 20×20 nm^2. Scalebars: (C) $1\ \mu$m, insets: 200 nm.

Quantifying and revealing artifacts in high-density datasets. A key difficulty encountered in high-density (HD) localization, when images of molecules overlap on the camera, is image artifacts that distort SMLM reconstructions in a structured, or vectorial, manner (28). Constructing an SMLM error map using a reference image has been proposed to certify the reliability of an SMLM reconstruction, but such a map does not quantify the reliability of *each individual* localization within the image. Here, we illustrate the power of measuring SM confidence in quantifying and revealing artifacts in a challenging HD localization experiment.

We use FALCON (29), an HD localization algorithm, to reconstruct a simulated benchmark SMLM dataset (12) consisting of 360 HD frames of a tubulin network (Fig. 3A). In regions where the tubules coalesce, corresponding to higher blinking densities, we see numerous inaccurate localizations (Fig. 3A, insets). In particular, we see fused and broadened tubules instead of thin and separate structures. A reliable error map should assign low confidence or high error to such regions while discriminating fine but accurate details of the structure. Interestingly, we notice significant differences between an error map (Fig. 3B, obtained via SQUIRREL (11)) and the proposed confidence map (Fig. 3C). First, the error map appears to overestimate errors in regions with accurate localizations, while our confidence map exhibits low confidence for inaccurate localizations and assigns high confidence to neighboring, well-resolved parallel tubules (Fig. 3B,C, top insets). Second, the error map underestimates the error in the regions where tubules are apparently fused, whereas the confidence map assigns an overall low confidence to this region, suggesting potential artifacts (Fig. 3B,C, bottom insets). Overall, our confidence map enables scientists to discriminate *specific* SM localizations that are trustworthy, while also assigning low confidence values to those that are not, thereby maximizing the utility of SMLM datasets without throwing away useful localizations.

Calibrating and validating WIF using SMLM of microtubules. A super-resolution dataset often contains well-isolated images of molecules, e.g., after a significant portion of them are bleached. These images can therefore serve as a useful internal control, taken under realistic conditions, to assess the performance of a PSF model as well as SMLM algorithms themselves on a particular dataset. As a practical example, we examine an SMLM dataset of blinking AlexaFluor 647-labeled microtubules (see Methods). We randomly selected 600 images of bright molecules sampled over the entire field of view (Fig. 4A). We used an ideal PSF model to localize these molecules using RoSE, but found that the mean confidence of these localizations is notably small ($\mathrm{WIF}_{\mathrm{avg}} = -0.36$), implying the presence of significant aberrations and PSF model mismatch (Fig. S7). We therefore calibrated our physics-based PSF model and re-analyzed the data (see Methods). After calibration, the estimated confidences of RoSE's localizations show a notable average increase of 0.79 ($\mathrm{WIF}_{\mathrm{avg}} = 0.43$). We also observe a rather broad distribution of confidences, suggesting that optical aberrations, such as defocus, vary throughout the structure (Fig. S7). RoSE's use of this calibrated PSF produces localizations with higher confidence values ($\mathrm{WIF}_{\mathrm{avg}} = 0.43$) compared to TS's use of an elliptical Gaussian PSF ($\mathrm{WIF}_{\mathrm{avg}} = 0.15$) (Fig. 4A). The higher average confidence score for RoSE suggests that it should recover the underlying structure with greater accuracy compared to TS.

We confirm the consistency of localization confidences, in the absence of the ground truth, through the perceived quality of the super-resolution reconstructions (Fig. 4B). We expect more confident localizations result in an image with greater resolution, whereas localizations with poor confidence fail to resolve fine details and potentially distort the underlying structure. Within a region containing a few parallel and well separated microtubules, we see similar confidences for both algorithms (Fig. 4H) resulting in images of similar quality (Fig. 4F,G). Conversely, for a region with intersecting microtubules, we observe marked qualitative and quantitative differences between the two reconstructions (Fig. 4C,D). RoSE is able to resolve structural details near the intersections, while the TS image contains missing and blurred localizations near the crossing points. Moreover, RoSE recovers the curved micro-

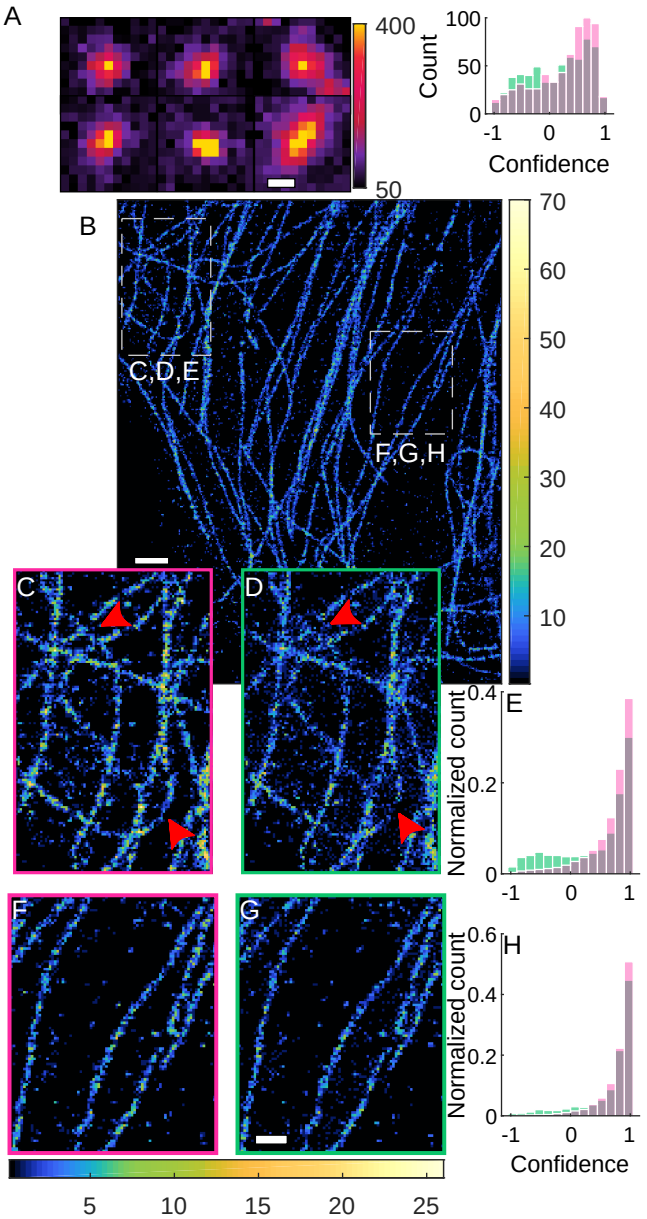

Fig. 4. Comparison of SMLM algorithms on experimental images of labeled microtubules. (A) Left: isolated images of Alexa Fluor 647 molecules. Right: localization confidences for 600 isolated molecules using RoSE (red) and TS (green). (B) Super-resolution image of Alexa Fluor 647-labeled microtubules recovered by RoSE. (C,D) Enlarged top-left region in (B) for RoSE and TS, respectively. (E) Histogram of confidences corresponding to localizations in (C) and (D) for RoSE (red) and TS (green), respectively. (F,G) Similar to (C,D) but for the middle-right region in (B). (H) Similar to (E) but for localizations in (F) and (G). Colorbars: (A) photons detected per 160×160 nm^2, (B) number of localizations per 40×40 nm^2. Scalebars: (A) 500 nm, (B) 1 μm, and (G) 500 nm.

tubule faithfully, whereas TS fails to reconstruct its central part (lower red arrow in Fig. 4C,D). Quantitatively, RoSE exhibits significantly greater confidence in its localizations compared to TS, which shows negative confidences for an appreciable number of localizations (Fig. 4E). This confidence gap, in part, can be caused by hidden parameters such as high blinking density. These SMLM reconstructions demonstrate that localization confidences obtained from both images of isolated molecules as well as HD datasets are a consistent and quantitative measure of algorithmic performance.

Quantifying algorithmic robustness and molecular heterogeneity. Next, we used WIF to characterize algorithmic performance on a Transient Amyloid Binding (TAB) (22) dataset imaging amyloid fibrils. Here, the relatively large shot noise in images of Nile red (<1000 photons per frame) tests the robustness of three distinct algorithms: TS with weighted-least squares (WLS) using a weighted Gaussian noise model; TS with maximum likelihood estimation (MLE) using a Poisson noise model; and RoSE, which uses a Poisson noise model but also is robust to image overlap. Qualitative and quantitative differences are readily noticeable between reconstructed images, particularly where the fibrillar bundle unwinds (Fig. 5A-C, insets). We attribute the poor localization of WLS, exemplified by broadening of the fibrils, to its lack of robustness to shot noise. By using instead a Poisson noise model, MLE recovers thinner and better resolved fibrils, but struggles to resolve fibrils at the top end of the structure (Fig. 5B,E). This inefficiency is probably due to algorithmic failure on images containing overlapping molecules. In contrast, RoSE localizations have greater precision and accuracy, thereby enabling the parallel unbundled filaments to be resolved (Fig. 5C,F). These perceived image qualities can be reliably quantified via WIF. Indeed, RoSE localizations show the greatest confidence of the three algorithms with $\mathrm{WIF}_{\mathrm{avg}} = 0.78$ while WLS shows a low $\mathrm{WIF}_{\mathrm{avg}}$ of 0.18 attesting to their excellent and poor recovery, respectively (Fig. 5G-I). Interestingly, we found that, in terms of FRC, RoSE has only 3% better resolution compared to MLE.

To further prove that WIF is a reliable measure of accuracy at the single-molecule level, we filtered out all localizations with confidence smaller than 0.5. Remarkably, filtered reconstructions from all three algorithms appear to resolve unbundled fibrils (Fig. 5J-L). In contrast, filtering based on estimated PSF width produces sub-optimal results. Notably, retaining MLE localizations within a strict width range $W_1 \in [90, 110$ nm$]$, improves filament resolvability at the cost of compromising sampling continuity (Fig. S8A). For a slightly larger range, $W_2 \in [70, 130$ nm$]$, the filtering is ineffective and the fibrils are not well resolved (Fig. S8B). In contrast, filtered localizations based on WIF, qualitatively and quantitatively, resolve fine fibrillar features (Fig. S8C).

A powerful feature of WIF is its ability to quantify an arbitrary discrepancy between a computational imaging model and SMLM measurements. This property is particularly useful since hidden physical parameters, which may be difficult to model accurately, can induce perturbations in the observed PSF. Therefore, we can use WIF to interrogate variations in the interactions of Nile red with amyloid fibrils that are encoded as subtle features within SMLM images. To demonstrate this capability, we analyzed TAB fibrillar datasets using RoSE and calculated the WIFs of localizations with greater than 400 detected photons (Fig. 6). Interestingly, WIF density plots reveal heterogeneous regions along both fibrils. Specifically, for segments of fibrils that are oriented away from the vertical axis, we see a larger fraction of localizations that have low confidence (<0.5) compared to regions that are vertically oriented (Fig. 6A,B). Quantitatively, the upper regions of two fibrils have 17% (Fig. 6C) and 37% (Fig. 6D) more localizations with confidence greater than 0.8.

To examine the origin of this heterogeneity, we directly

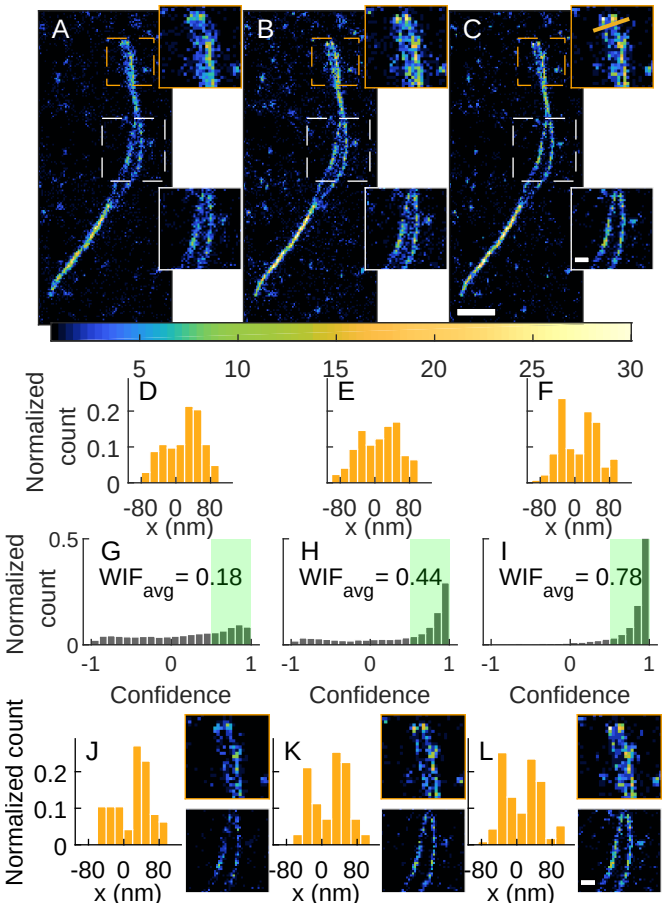

Fig. 5. Quantifying algorithmic robustness and enhancing reconstruction accuracy in SMLM of amyloid fibrils. Super-resolution image of twisted fibrils recovered by (A) TS weighted least-squares (WLS), (B) TS maximum-liklihood estimation (MLE), and (C) RoSE. (D-F) Histogram of localizations projected onto the gold line in (C, top inset) from (D) WLS, (E) MLE, and (F) RoSE. (G) WIFs for all WLS localizations in (A). (H) WIFs for all MLE localizations in (B). (I) WIFs for all RoSE localizations in (C). In (G-I), green regions denote localizations with confidence greater than 0.5. (J-L) Histograms of localizations with confidence greater than 0.5 projected onto the golden line in (C, top inset) and corresponding filtered inset images for (J) WLS, (K) MLE, and (L) RoSE. Colorbar: number of localizations per 58.8×58.5 nm^2. Scalebars: (C) 500 nm, (C inset) 150 nm, (L inset) 150 nm.

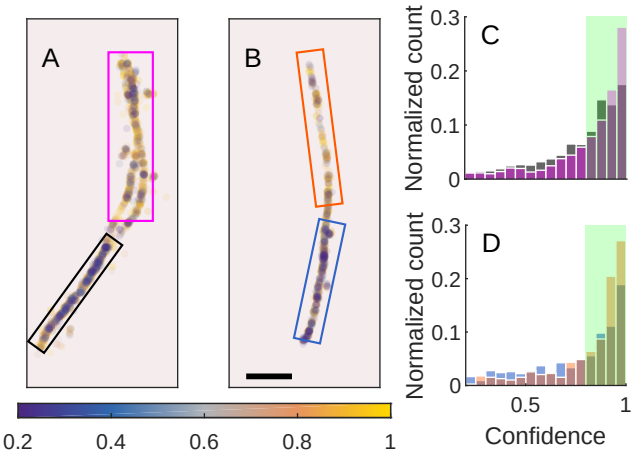

Fig. 6. Heterogeneity in Nile red interactions with amyloid fibrils. (A, B) WIFs of bright localizations (>400 detected photons) detected by RoSE on two fibrils. (C, D) WIFs of localizations within corresponding boxed regions in (A, upper magenta and lower black) and (B, upper orange and lower blue), respectively. In (C) and (D), green regions indicate localizations with $\mathrm{WIF} > 0.8$, corresponding to (A, magenta) 63%, (A, black) 54%, (B, orange) 62%, and (B, blue) 45% of the localizations. Colorbar: confidence. Scalebar: 500 nm.

compare observed PSFs from high- and low-confidence regions. Curiously, PSFs in the bottom regions are slightly elongated along an axis parallel the fibril itself, whereas PSFs from the top regions better match our model (Fig. S9). These features may be attributed to the orientation of Nile red molecules upon binding to fibrils (30–32). We stress that the influence of molecular orientation on these PSFs is detected and quantified by WIF and cannot otherwise be distinguished by estimates of PSF width (Fig. S9).

Discussion

WIF is a computational tool that utilizes mathematical models of the imaging system and measurement noise to measure the statistical confidence of each localization within an SMLM image. We used WIF to benchmark the accuracy of SMLM algorithms on a variety of simulated and experimental datasets. We also demonstrated WIF for analyzing how sample properties (e.g., defocus, dipole emission, molecular density) affect reconstruction accuracy. WIF quantifies robustly *deterministic* model mismatch between experimental images and a computational imaging model, which affects accuracy, in the presence of *stochastic* measurement noise, which affects precision. Intuitively, low SNRs make the detection of minor model mismatches, such as defocus, comparatively difficult (Fig. S4). While WIF has excellent sensitivity for detecting overlapping molecules (Fig. 1D) and dipole-like emission patterns (Figs. S5-S6), WIF cannot *explain* the source of low confidence values that cause localization inaccuracies or heterogeneities; rather, it detects and quantifies these effects.

WIF exhibits several advantages over existing methods for quantifying reconstruction accuracy in experimental SMLM. First, WIF does not require labeled training data to judge the trustworthiness of predetermined image features; a model of the imaging system, i.e., its PSF, suffices. Second, it does not need ground-truth knowledge of SM positions, which would be prohibitive in most SMLM applications. Third, it obviates the need for a secondary imaging modality for comparison and is therefore more robust than such methods; it does not require alignment between modalities. More fundamentally, WIF exploits a unique property of SMLM compared to other non-SM super-resolution optical methodologies (e.g., structured illumination, RESOLFT, and STED); imaging the entirety (peak and spatial decay) of each SM PSF synergistically creates well-behaved gradient flows along the likelihood surface that are used in computing WIF (SI Section 2). WIF quantifies errors by using knowledge from its PSF model to explore the *object* space of molecular positions and brightnesses; leveraging Wasserstein distance ensures that meaningful perturbations to SM positions are tested. In contrast, computing mismatches in *image* space (e.g., PSF width in Figs. 1, S4-S6) is insensitive to molecular overlap, defocus, and dipole emission artifacts without assuming strong statistical priors on the spatial distribution of molecules or a simplified PSF (33).

We believe that WIF will become a valuable tool for the SMLM community as it offers the unique capability of quanti-

fying localization accuracy and heterogeneity in experimental SMLM datasets. WIF can also be used for online tuning of parameters (e.g., activation density and imaging buffer conditions) during an experiment to maximize imaging accuracy and resolution. It also offers a reliable means to detect otherwise hidden properties $\boldsymbol{\beta}$ of the sample of interest, such as molecular orientation shown here, allowing for the discovery of new biophysical and biochemical phenomena at the nanoscale. While a majority of neural network training methods in SMLM utilize simulated data (34) or experimental data assuming a perfectly matched model (35), the discriminative power of WIF may enable these networks to be trained robustly on experimental data in the presence of mismatches stemming from hidden parameters. Finally, WIF can be readily extended to 3D (36) or molecular orientation (37–42) SMLM using calibrated models.

Inspired by measuring distribution model risk (43), i.e., quantifying the impact of using a "best-guess" model instead of the true model, WIF represents an advance in statistical quantification in imaging science (44–46), where the reliability of *each quantitative feature* within a scientific image can now be evaluated. The benefits of integrating WIF into downstream analysis (47, 48) (e.g., SM clustering, counting, and co-localization) and even in other imaging modalities (e.g., astronomical imaging, positron emission tomography, and computed tomography) are exciting opportunities yet to be explored.

Materials and Methods

Super-resolution imaging of labeled microtubules. The microtubules of BSC-1 cells were immunolabeled with Alexa Fluor 647 (Invitrogen) and imaged under blinking conditions (49) with glucose oxidase/catalase and mM concentrations of mercaptoethylamine (MEA) as in (50). The sample was imaged using an Olympus IX71 epifluorescence microscope equipped with a 100× 1.4 NA oil-immersion objective lens (Olympus UPlan-SApo 100×/1.40). Fluorophores were excited using a 641-nm laser source (Coherent Cube, peak intensity ~10 kW/cm^2). Fluorescence from the microscope was filtered using a dichroic beamsplitter (Semrock, Di01-R635) and bandpass filter (Omega, 3RD650-710) and separated into two orthogonally-polarized detection channels, but only one channel was used for this analysis. Fluorescence photons were captured using an electron-multiplying (EM) CCD camera (Andor iXon+ DU897-E) at an EM gain setting of 300 with a pixel size of 160×160 nm^2 in object space. For the 284k localizations shown in Fig. 4B, 2287 photons were detected on average with a background of 76 photons per pixel.

Transient Amyloid Binding imaging. The 42 amino-acid residue amyloid-beta peptide (Aβ42, ERI Amyloid Laboratory) was dissolved in hexafluoro-2-propanol (HFIP) and sonicated at room temperature for one hour. After flash freezing in liquid nitrogen, HFIP was removed by lyophilization and stored at -20 °C. To further purify the protein, the lyophilized Aβ42 was dissolved in 10 mM NaOH, sonicated for 25 min in a cold water bath and filtered first through a 0.22 μm and then through a 30 kD centrifugal membrane filter (Millipore Sigma, UFC30GV and UFC5030) as described previously (22). To prepare fibrils, we incubated 10 μM monomeric Aβ42 in phosphate-buffered saline (PBS, 150 mM NaCl, 50 mM Na_3PO_4, pH 7.4) at 37 °C with 200 rpm shaking for 20-42 hours. The aggregated structures were adsorbed to an ozone-cleaned cell culture chamber (Lab Tek, No. 1.5H, 170 ± 5 μm thickness) for 1 hour followed by a rinse using PBS. A PBS solution (200 μL) containing 50 μM Nile red (Fisher Scientific, AC415711000) was placed into the amyloid-adsorbed chambers for transient amyloid binding.

Blinking of Nile red molecules on fibrils were imaged using a home-built epifluorescence microscope equipped with a 100× 1.4 NA oil-immersion objective lens (Olympus, UPlan-SApo 100×/1.40). The samples were excited using a 561-nm laser source (Coherent Sapphire, peak intensity ~0.45 kW/cm^2). Fluorescence was filtered by a dichroic beamsplitter (Semrock, Di03-R488/561) and a bandpass filter (Semrock, FF01-523/610) and separated into two orthogonally-polarized detection channels by a polarizing beamsplitter cube (Meadowlark Optics). Both channels were captured by a scientific CMOS camera (Hamamatsu, C11440-22CU) with a pixel size of 58.5×58.5 nm^2 in object space. Only one of the channels was analyzed in this work. For the 12k localizations shown in Fig. 5C, 390 photons were detected on average with a background of 5 photons per pixel. For the 931 localizations shown in Fig. 6B, 785 photons were detected on average with a background of 2.4 photons per pixel.

Synthetic data. We generated images of molecules via a vectorial image-formation model (24), assuming unpolarized ideal PSFs. Briefly, a molecule is modeled as a dipole rotating uniformly within a cone with a half-angle α. A rotationally fixed dipole corresponds to $\alpha = 0$, while $\alpha = 90°$ represents an isotropic molecule. Molecular blinking trajectories were simulated using a two state Markov chain (28). We used a wavelength of 637 nm, NA = 1.4, and spatially uniform background. We simulated a camera with 58.5×58.5 nm^2 square pixels in object space.

Poisson log likelihood. Consider a set of N molecules

$$\mathcal{M} = \sum_{i=1}^{N} s_i \delta(\boldsymbol{r} - \boldsymbol{r}_i),$$

where $s_i \geq 0$ and $\boldsymbol{r}_i \in \mathbb{R}^3$ denote i^{th} molecules' brightness (in photons) and position, respectively. The resulting intensity μ_j, that is, the expected number of photons detected on camera, for each pixel j can be written as

$$\mu_j = \sum_{i=1}^{N} \{s_i q_j(\boldsymbol{r}_i) + b_j\}, \quad j \in \{1, \ldots, m\},$$

where $q_j(\boldsymbol{r}_i)$ represents the value of the PSF q (for i^{th} molecule) at j^{th} pixel; b_j denotes the expected number of background photons at j^{th} pixel.

If we denote $\boldsymbol{g} \in \mathbb{R}^m$ as m pixels of photon counts captured by a camera, the negative Poisson log likelihood is then given by

$$\mathcal{L}(q, \mathcal{M}, \boldsymbol{g}) = \sum_{j=1}^{m} \{\mu_j - g_j \log(\mu_j)\}.$$

Jaccard index. Following (8), given a set of ground-truth positions and corresponding localizations, we first match these points by solving a bipartite graph-matching problem of minimizing the sum of distances between the two elements of a pair. We say that a pairing is successful if the distance between the corresponding two elements is smaller than twice the full width at half maximum (FWHM) of the localization precision σ, which is calculated using the theoretical Cramér-Rao bound (51) ($\sigma = 3.4$ nm with 2000 photons detected). The elements that are paired with a ground-truth position are counted as true positive (TP) and those without a pair are counted as false positive (FP). Finally, the ground-truth molecules without a match are counted as false negative (FN).

PSF modeling for computing Wasserstein-induced flux. For simulation studies, we used an ideal, unpolarized standard PSF resulting from an isotropic emitter (Figs. 1, 2, 3, S3, S4, S5, S6), while for experimental data (Figs. 4, 5, 6, S7, S8, S9), we used a linearly-polarized PSF, also resulting from an isotropic emitter (SI Section 2).

In addition to the ideal PSFs modeled above, we needed to calibrate the aberrations present in the PSF used for microtubule imaging (Fig. 4). We modeled the microscope pupil function P as

$$P(u, v) = \exp\left(j \sum_{i=3}^{l} \{a_i Z_i(u, v)\}\right) \cdot P_0(u, v),$$

where (u, v) are microscope's pupil coordinates; Z_i and a_i represent the ith Zernike basis function and its corresponding coefficient; and P_0 denotes the pupil function of the uncalibrated model. We used 33 Zernike modes corresponding to $l = 35$.

Using RoSE, we localized well-isolated molecules over a large field-of-view (FOV) corresponding to Fig. 4. Next, for each localization, we obtained an image of size 11×11 pixels with the localized molecule being at its center. We excluded molecules with brightnesses less than 3000 photons or with positions away from the origin by more that a pixel. Next, we randomly selected 600 of these images to estimate the Zernike coefficients, i.e., $\{a_1, \ldots, a_l\}$, as described previously (52). The calibrated PSF (Fig. S7) is then computed based on recovered P.

Data and code availability. The data and code that support the findings of this study is available from the corresponding author upon reasonable request.

Acknowledgements. We thank W.E. Moerner and Steffen J. Sahl for contributing the SMLM microtubule dataset. We also thank Jin Lu, Oumeng Zhang, and Tingting Wu for helpful discussion. Research reported in this publication was supported by the National Science Foundation under grant number ECCS-1653777 and by the National Institute of General Medical Sciences of the National Institutes of Health under grant number R35GM124858 to M.D.L.

References

1. Moerner WE, Shechtman Y, Wang Q (2015) Single-molecule spectroscopy and imaging over the decades. *Faraday discussions* 184:9–36.
2. Sauer M, Heilemann M (2017) Single-molecule localization microscopy in Eukaryotes. *Chemical Reviews* 117(11):7478–7509.
3. Sahl SJ, Hell SW, Jakobs S (2017) Fluorescence nanoscopy in cell biology. *Nature Reviews. Molecular Cell Biology* 18(11):685–701.
4. Baddeley D, Bewersdorf J (2018) Biological insight from super-resolution microscopy: What we can learn from localization-based images. *Annual Review of Biochemistry* 87(1):965–989.
5. Moringo NA, Shen H, Bishop LDC, Wang W, Landes CF (2018) Enhancing analytical separations using super-resolution microscopy. *Annual Review of Physical Chemistry* 69(1):353–375.
6. Chen T, et al. (2017) Optical super-resolution imaging of surface reactions. *Chemical Reviews* 117(11):7510–7537.
7. Willets KA, Wilson AJ, Sundaresan V, Joshi PB (2017) Super-resolution imaging and plasmonics. *Chemical Reviews* 117(11):7538–7582.
8. Sage D, et al. (2019) Super-resolution fight club: assessment of 2D and 3D single-molecule localization microscopy software. *Nature Methods* 16(5):387–395.
9. Cohen EA, Abraham AV, Ramakrishnan S, Ober RJ (2019) Resolution limit of image analysis algorithms. *Nature communications* 10(1):793.
10. Nieuwenhuizen RPJ, et al. (2013) Measuring image resolution in optical nanoscopy. *Nature Methods* 10(6):557–562.
11. Culley S, et al. (2018) Quantitative mapping and minimization of super-resolution optical imaging artifacts. *Nature Methods* 15(4):263–266.
12. Sage D, et al. (2015) Quantitative evaluation of software packages for single-molecule localization microscopy. *Nature Methods* 12(8):717–724.
13. Fox-Roberts P, et al. (2017) Local dimensionality determines imaging speed in localization microscopy. *Nature Communications* 8:13558.
14. Deschout H, et al. (2014) Precisely and accurately localizing single emitters in fluorescence microscopy. *Nature Methods* 11(3):253–266.
15. Enderlein J, Toprak E, Selvin PR (2006) Polarization effect on position accuracy of fluorophore localization. *Optics Express* 14(18):8111–8120.
16. Engelhardt J, et al. (2010) Molecular orientation affects localization accuracy in superresolution far-field fluorescence microscopy. *Nano Letters* 11(1):209–213.
17. Lew MD, Backlund MP, Moerner W (2013) Rotational mobility of single molecules affects localization accuracy in super-resolution fluorescence microscopy. *Nano Letters* 13(9):3967–3972.
18. Burke D, Patton B, Huang F, Bewersdorf J, Booth MJ (2015) Adaptive optics correction of specimen-induced aberrations in single-molecule switching microscopy. *Optica* 2(2):177–185.
19. von Diezmann A, Lee MY, Lew MD, Moerner W (2015) Correcting field-dependent aberrations with nanoscale accuracy in three-dimensional single-molecule localization microscopy. *Optica* 2(11):985–993.
20. Ji N (2017) Adaptive optical fluorescence microscopy. *Nature Methods* 14(4):374–380.
21. Villani C (2008) *Optimal transport: old and new.* (Springer Science & Business Media) Vol. 338.
22. Spehar K, et al. (2018) Super-resolution imaging of amyloid structures over extended times by using transient binding of single Thioflavin T molecules. *ChemBioChem* 19(18):1944–1948.
23. Lee JE, et al. (2018) Mapping surface hydrophobicity of α-synuclein oligomers at the nanoscale. *Nano Letters* 18(12):7494–7501.
24. Backer AS, Moerner WE (2014) Extending single-molecule microscopy using optical Fourier processing. *The Journal of Physical Chemistry. B* 118:8313–8329.
25. Ovesný M, Křížek P, Borkovec J, Svindrych Z, Hagen GM (2014) ThunderSTORM: a comprehensive ImageJ plug-in for palm and STORM data analysis and super-resolution imaging. *Bioinformatics* 30:2389–2390.
26. Dauphin YN, et al. (2014) Identifying and attacking the saddle point problem in high-dimensional non-convex optimization in *Advances in neural information processing systems*. pp. 2933–2941.
27. Santambrogio F (2017) {Euclidean, metric, and Wasserstein} gradient flows: an overview. *Bulletin of Mathematical Sciences* 7(1):87–154.
28. Mazidi H, Lu J, Nehorai A, Lew MD (2018) Minimizing structural bias in single-molecule super-resolution microscopy. *Scientific Reports* 8:13133.
29. Min J, et al. (2015) FALCON: fast and unbiased reconstruction of high-density super-resolution microscopy data. *Scientific Reports* 4:4577.
30. Duboisset J, et al. (2013) Thioflavine-T and congo red reveal the polymorphism of insulin amyloid fibrils when probed by polarization-resolved fluorescence microscopy. *The Journal of Physical Chemistry B* 117(3):784–788.
31. Shaban HA, Valades-Cruz CA, Savatier J, Brasselet S (2017) Polarized super-resolution structural imaging inside amyloid fibrils using Thioflavine t. *Scientific Reports* 7:12482.
32. Varela JA, et al. (2018) Optical structural analysis of individual α-synuclein oligomers. *Angewandte Chemie International Edition* 57(18):4886–4890.
33. Lindén M, Ćurić V, Amselem E, Elf J (2017) Pointwise error estimates in localization microscopy. *Nature Communications* 8(1):15115.
34. Hershko E, Weiss LE, Michaeli T, Shechtman Y (2019) Multicolor localization microscopy and point-spread-function engineering by deep learning. *Optics Express* 27(5):6158–6183.
35. Kim T, Moon S, Xu K (2019) Information-rich localization microscopy through machine learning. *Nature Communications* 10(1):1996.
36. von Diezmann A, Shechtman Y, Moerner W (2017) Three-dimensional localization of single molecules for super-resolution imaging and single-particle tracking. *Chemical Reviews* 117(11):7244–7275.
37. Aguet F, Geissbühler S, Märki I, Lasser T, Unser M (2009) Super-resolution orientation estimation and localization of fluorescent dipoles using 3-D steerable filters. *Optics Express* 17(8):6829–6848.
38. Stallinga S, Rieger B (2012) Position and orientation estimation of fixed dipole emitters using an effective Hermite point spread function model. *Optics Express* 20(6):5896–5921.
39. Backlund MP, et al. (2012) Simultaneous, accurate measurement of the 3D position and orientation of single molecules. *Proceedings of the National Academy of Sciences of the United States of America* 109(47):19087–19092.
40. Backlund MP, Lew MD, Backer AS, Sahl SJ, Moerner W (2014) The role of molecular dipole orientation in single-molecule fluorescence microscopy and implications for super-resolution imaging. *ChemPhysChem* 15(4):587–599.
41. Zhang O, Lu J, Ding T, Lew MD (2018) Imaging the three-dimensional orientation and rotational mobility of fluorescent emitters using the Tri-spot point spread function. *Applied Physics Letters* 113(3):031103.
42. Mazidi H, King ES, Zhang O, Nehorai A, Lew MD (2019) Dense super-resolution imaging of molecular orientation via joint sparse basis deconvolution and spatial pooling. *IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)* pp. 325–329.
43. Breuer T, Csiszár I (2016) Measuring distribution model risk. *Mathematical Finance* 26(2):395–411.
44. Beck RN (1993) Overview of imaging science. *Proceedings of the National Academy of Sciences* 90(21):9746–9750.
45. Barrett HH, Myers KJ (2013) *Foundations of image science*. (John Wiley & Sons).
46. Pereyra M (2017) Maximum-a-posteriori estimation with Bayesian confidence regions. *SIAM Journal on Imaging Sciences* 10(1):285–302.
47. Nicovich PR, Owen DM, Gaus K (2017) Turning single-molecule localization microscopy into a quantitative bioanalytical tool. *Nature Protocols* 12(3):453.
48. Schnitzbauer J, et al. (2018) Correlation analysis framework for localization-based super-resolution microscopy. *Proceedings of the National Academy of Sciences* 115(13):3219–3224.
49. Dempsey GT, Vaughan JC, Chen KH, Bates M, Zhuang X (2011) Evaluation of fluorophores for optimal performance in localization-based super-resolution imaging. *Nature Methods* 8(12):1027–1036.
50. Lee HlD, Sahl SJ, Lew MD, Moerner W (2012) The double-helix microscope super-resolves extended biological structures by localizing single blinking molecules in three dimensions with nanoscale precision. *Applied Physics Letters* 100(15):153701.
51. Snyder DL, Miller MI (2012) *Random point processes in time and space*. (Springer Science & Business Media).
52. Petrov PN, Shechtman Y, Moerner WE (2017) Measurement-based estimation of global pupil functions in 3D localization microscopy. *Optics Express* 25(7):7945–7959.