Supplementary materials

Details of algorithmic comparisons

All parameters of the methods compared here are tuned for best performance.

FALCON [5]: We set the sparsity parameter (the ℓ_1 weights) κ to 3. The threshold above which the support is defined is set as 10% of the maximum intensity.

SPIDER [17]: We set the sparsity parameter, the weights of ℓ_0 regulation, $\kappa = 250$.

deconSTORM [7]: In our simulations accuracy continued to improve even after 5000 iterations, so we used 5000 iterations in our comparisons. Note computation time is proportional to the number of iterations, so this method could be sped up at the cost of some accuracy.

3B [8]: As mentioned in the Discussion, we were not able to obtain reasonable results using this method, even after > 1 day of computation, for data sets with e.g. 2000 observed frames. In personal communications with the developers of this method, it was emphasized that this approach is better suited for smaller images and smaller values of N, since the speed of this method decreases more or less with the square of the size of the PSF (in pixels), the size of the image area being analyzed, and the number of frames observed. Therefore we did not pursue further quantitative comparisons against this method.

Evaluation: Identified fluorophores are defined as estimates within some fixed distance from the true fluorophores. The cutoff radius we used was 50 nm. FALCON returns a list of fluorophore locations, whereas vEM, deconSTORM, and SPIDER return images of the estimated fluorophore density. To quantify fluorophore estimate accuracy for these methods we thresholded these images and took local weighted averages to obtain the estimated fluorophore locations.

Experimental details

Simulation: To validate our analysis method, we simulated grid data on a 32-by-32 pixel map with an image pixel size of 100 nm. The final resolved image sits on a 3x finer grid with super resolution pixel size of 33 nm. Unless stated otherwise, each frame has an emission rate of 0.04, corresponding to an average molecule density of 6.8 μm^{-2}. The average photon number is 1,000 per fluorophore with PSF width 150 nm in standard deviation or 353 nm in FWHM. To test the performance of our method under different practical situations, we varied critical parameters, including the number of frames, molecule density, and PSF width. In those simulations, we replace the above parameters with a range stated in the corresponding figure caption and keep other parameters unchanged.

Real data: We reconstruct a patch of tubulins on a 32-by-32 pixel image map with 5000 real experimental frames. The final resolved image sits on a 4x finer grid. We modeled the PSF as a Gaussian blur with width of 183 nm in standard deviation; this parameter was estimated by fitting observations of non-overlapping fluorophores to a 2D Gaussian function, following [19]. The dataset is provided as Tubulin ConjAL647 on the Single Molecule Localization Microscope website [4].

Markov model

In the main text, we focus on a model in which fluorophores become active according to a Poisson process with rate λ. The active fluorophores in one frame are conditionally independent from those in other frames, given λ. In this section, we incorporate the phenomenon that some fluorophores do not quench immediately, i.e. there is a probability that an active fluorophore will remain active in the following frames. Thus the active fluorophores in one frame consist of two groups: "newborn" fluorophores that activate from the dark state with rate λ (the same as before), plus fluorophores remaining active from the previous frame, each with probability α. Under this assumption, the active fluorophores in one frame are dependent on those in the frame before and after, so we denote the new model as the "Markov model" and the original model as the "non-Markov model."

The Markov model incorporates the positions of active fluorophores in neighboring frames $i+1$ and $i-1$, which can potentially be useful to help pinpoint the positions of active fluorophores in each frame i. Thus we would expect the Markov model to outperform the non-Markov in localizing individual fluorophores. (Similar neighboring-frame effects are incorporated in [7] and [8].) In this section, we will first introduce the Markov model, and then show results comparing the effectiveness of the Markov model versus the non-Markov model.

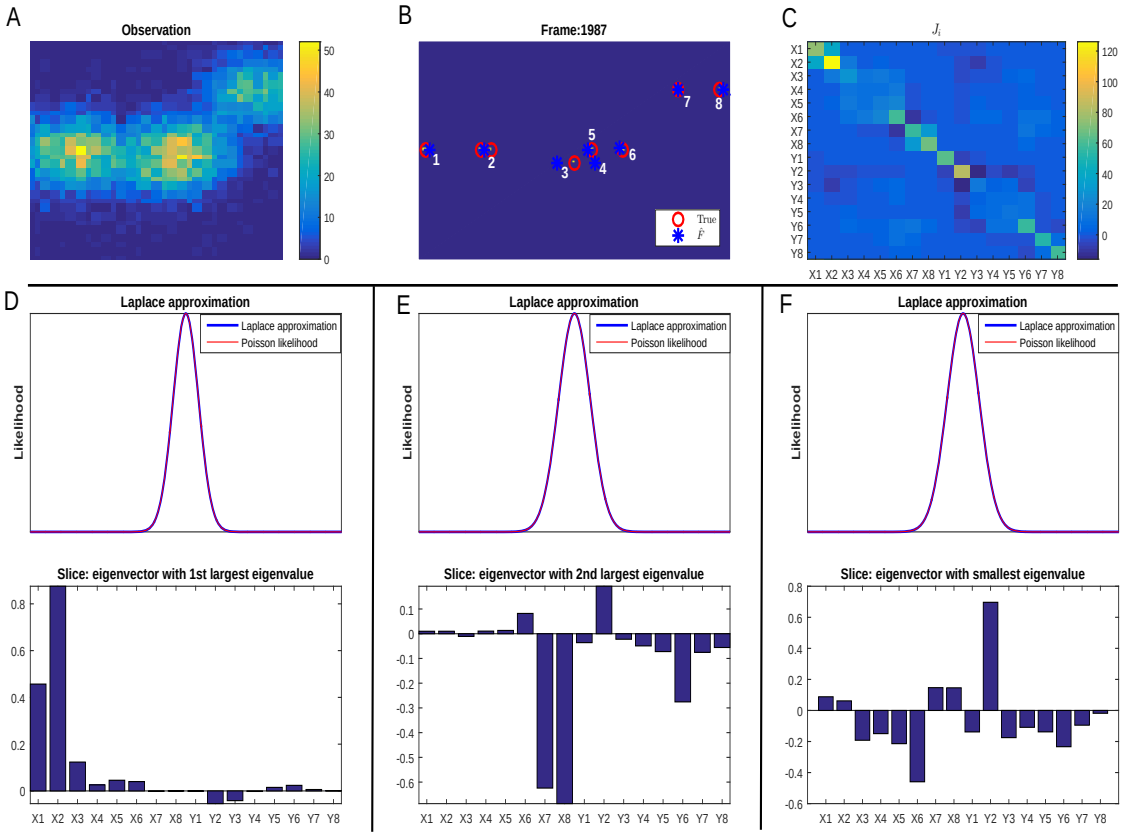

Figure 8: **Accuracy of the Laplace approximation**. We use the same frame as in Figure 2 in the main text as an illustration. **(A)** Y_i. **(B)** Red circles are true positions; blue stars are $\hat{F}_i^j$. **(C)** Fisher information matrix J_i. Xj indicates x coordinates of fluorophore j, and Yj the y coordinates. FALCON inferred 8 fluorophores in this frame, so we have 16 total coordinates. Note that the blocks of J_i corresponding to fluorophores 3,4,5,6 have smaller values, indicating reduced estimation accuracy due to overlapping PSF bumps. Panels **(D), (E), (F)** indicate the accuracy of the Laplace approximation for the Poisson likelihood. Since the likelihood w.r.t. F_i is a 16-dimensional function, we can only display slices of this function. We choose slices in the direction of eigenvectors of J_i — the principal components of the Laplace approximation. (Directions appear in bottom panels.) The red (Poisson likelihood) and blue (Laplace approximation) curves align very closely in each of the three directional slices shown here.

The model

We begin by writing down the Markov model for the time series of activations of fluorophores $I_{:,xy}$ at location xy across all N frames:

$$p(I_{:,xy}|\lambda_{xy},\alpha) = p(I_{1,xy}) \prod_{i>1}^{N} p(I_{i,xy}|I_{i-1,xy},\lambda_{xy},\alpha) \tag{22}$$

(As usual, fluorophores in different locations xy activate conditionally independently given λ.) The transition matrix is given by

$$\begin{cases} P(I_{i,xy}=1|I_{i-1,xy}=0,\lambda,\alpha) = \lambda \\ P(I_{i,xy}=0|I_{i-1,xy}=0,\lambda,\alpha) = 1-\lambda \\ P(I_{i,xy}=1|I_{i-1,xy}=1,\lambda,\alpha) = \alpha+\lambda \\ P(I_{i,xy}=0|I_{i-1,xy}=1,\lambda,\alpha) = 1-\alpha-\lambda, \end{cases} \tag{23}$$

where α is the probability that an active fluorophore remains active in the next frame, and λ is defined in eq. 2, 3; note that the probability of a new fluorophore activating in any frame is typically fairly low (to guarantee that each image I_i is sparse), so $\lambda \ll 1$, while the probability of remaining active may be non-negligible. We are most interested here in the case that α is significantly greater than 0, implying $\lambda \ll \alpha$. Finally, note

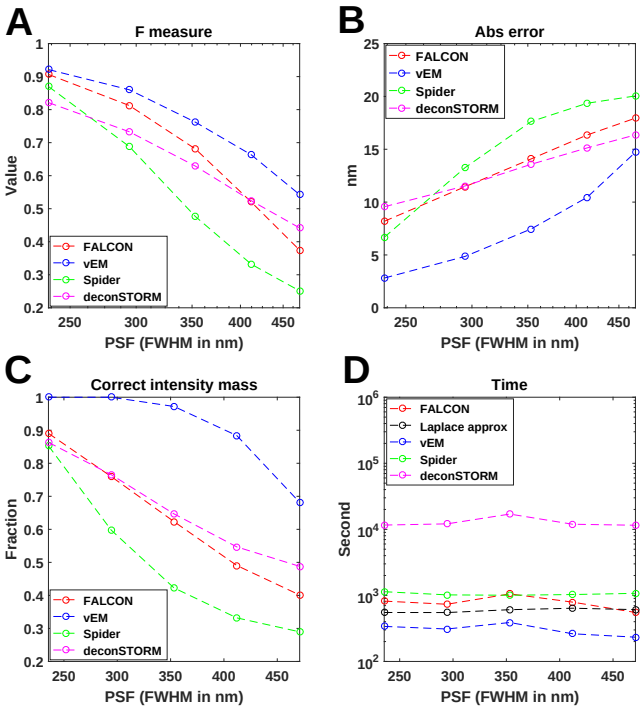

Figure 9: **Evaluation of vEM in comparison with FALCON [5], deconSTORM [7], and SPIDER [17] as a function of PSF width**. Panel layout as in Fig. 5. 2000 frames were used here.

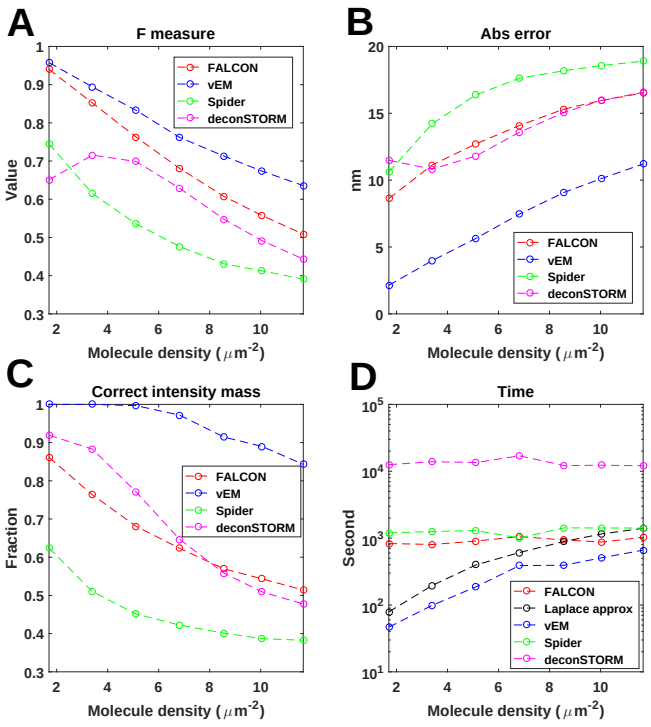

Figure 10: **Evaluation of vEM in comparison with FALCON [5], deconSTORM [7], and SPIDER [17] as a function of fluorophore density** p. Panel layout as in Figures 5 and 9. We found that deconSTORM tends to have an inflated false positive rate (i.e., lower precision) for small values of p in panel (A). Also note that vEM is relatively cheaper computationally for small p, where there are fewer fluorophores to iterate over in the Laplace approximation and vEM steps. 2000 frames were used here.

that we use a Bernoulli emission model here instead of the Poisson emission model used in the main text (eq. 5); this simplifies the derivations below. Of course the Poisson and Bernoulli models are identical in the limit of small λ.

After the Laplace approximation our full approximate loglikelihood is

$$\ln p(I, Y|\lambda, \alpha) = \sum_{i=1}^{N} \Bigg\{ \sum_{x}^{D} \sum_{y}^{D} \ln p(I_{i,xy}|I_{i-1,xy}, \lambda_{xy}, \alpha) + \ln \mathcal{N}(F_i|\hat{F}_i, \hat{\Sigma}_i) \Bigg\} \tag{24}$$

where in eq. 24 we define I_0 as all zeros.

Variational EM Algorithm

Now we proceed as before and maximize the ELBO (eq. 9) to obtain the E and M steps. We will assume that α is known. In reality, of course, α is unknown and needs to be estimated along with λ. It is straightforward to derive EM iterations for α as well, but we do not pursue this here. Instead, in the Results section we will simulate data from the Markov model and estimate λ using a known value of α, to give the Markov approach the best possible chance of improving over the results of the non-Markov model.

M step

$$\hat{\lambda} = \arg\max_{\lambda} \mathcal{L}(\lambda, q(F)) \tag{25}$$

$$= \arg\max_{\lambda} \mathbb{E}_q[P(I, Y|m, \lambda, \alpha)] \tag{26}$$

$$= \arg\max_{\lambda} \mathbb{E}_q[\sum_{i}^{N} \ln p(I_i|I_{i-1}, \lambda, \alpha)] \tag{27}$$

$$= \arg\max_{\lambda} \mathbb{E}_q[\sum_{i}^{N} (1 - I_{i-1}) \ln \left[\lambda^{I_i}(1-\lambda)^{1-I_i}\right] + (I_{i-1}) \ln \left[(\alpha+\lambda)^{I_i}(1-\alpha-\lambda)^{1-I_i}\right]]; \tag{28}$$

eq. 28 combines the cases in eq. 23 and uses the property that I_i can only take values of 0 or 1. We have dropped the pixel subscript xy to simplify notation; the maximization problem is separable over locations xy and therefore we can optimize for each pixel independently in parallel.

Setting the derivative w.r.t. λ to zero, we obtain

$$\frac{Q_{1,N}}{\hat{\lambda}} - \frac{Q'_{1,N}}{\hat{\lambda}} - \frac{N - Q_{1,N}}{1-\hat{\lambda}} + \frac{Q_{0,N-1} - Q'_{1,N}}{1-\hat{\lambda}} + \frac{Q'_{1,N}}{\alpha+\hat{\lambda}} - \frac{Q_{0,N-1} - Q'_{1,N}}{1-\alpha-\hat{\lambda}} = 0, \tag{29}$$

where we define $Q \in R_+^{D\times D}$ and $Q' \in R_+^{D\times D}$:

$$Q_{S,N} = \sum_{i=S}^{N} E(I_i) = \sum_{i=S}^{N}\sum_{j}^{m_i} q_{ij}(F_{ij}) \tag{30}$$

$$Q'_{S,N} = \sum_{i=S}^{N} E(I_i) \odot E(I_{(i-1)}) \tag{31}$$

$$= \sum_{i=S}^{N} \left\{ \sum_{j=1}^{m_i} q_{ij}(F_{ij}) \right\} \odot \left\{ \sum_{k=1}^{m_{(i-1)}} q_{ik}(F_{ik}) \right\}. \tag{32}$$

Now set

$$k_1 = Q_{1,N} - Q'_{1,N} \tag{33}$$

$$k_2 = N - Q_{1,N} - Q_{0,N-1} + Q'_{1,N} \tag{34}$$

$$k_3 = \frac{Q'_{1,N}}{\alpha+\hat{\lambda}} - \frac{Q_{0,N-1} - Q'_{1,N}}{1-\alpha-\hat{\lambda}} \tag{35}$$

$$\approx \frac{Q'_{1,N}}{\alpha} - \frac{Q_{0,N-1} - Q'_{1,N}}{1-\alpha} \tag{36}$$

where in eq. 36, we used $\lambda << \alpha$. Therefore, eq. 29 becomes

$$\frac{k_1}{\hat{\lambda}} - \frac{k_2}{1-\hat{\lambda}} + k_3 = 0, \tag{37}$$

which reduces to a simple quadratic equation in $\hat{\lambda}$. We select the valid solution $\hat{\lambda} \in (0, 1)$.

Similarly as in the non-Markov model, we can perform soft-thresholding on the obtained value to increase the sparsity of $\hat{\lambda}$.

Finally, note that if we set $\alpha = 0$, eq. 29 becomes

$$\frac{Q}{\hat{\lambda}} - \frac{N-Q}{1-\hat{\lambda}} = 0; \tag{38}$$

this leads to $\hat{\lambda} = Q/N$, which corresponds to the M step in the non-Markov model (description under eq. 16), as desired.

E step:

$$q_{ij}(F_i^j) = \arg\max_{q_{ij}} \mathcal{L}(\hat{\lambda}, q(F)) \tag{39}$$

$$\propto \exp\left\{ \mathbb{E}_{q_{\setminus ij}}\big[\ln P(I, Y|m, \hat{\lambda}, \alpha)\big] \right\} \tag{40}$$

$$\propto \exp\left\{ \mathbb{E}_{q_{\setminus ij}}\Big[\ln p(I_i|I_{i-1}, \hat{\lambda}, \alpha) + \ln p(I_{i+1}|I_i, \hat{\lambda}, \alpha) + \ln \mathcal{N}(F_i|\hat{F}_i, \hat{\Sigma}_i)\Big] \right\} \tag{41}$$

$$\propto \exp\Bigg\{$$

$$\underbrace{+\left(\mathbb{E}_{q_{\setminus ij}}\big[I_{i-1}\big] \ln\big[(\frac{1-\hat{\lambda}}{\hat{\lambda}})(\frac{\alpha+\hat{\lambda}}{1-\alpha-\hat{\lambda}}) \big] + \ln\frac{\hat{\lambda}}{1-\hat{\lambda}} \right)_{F_i^j}}_{\text{effect of } I_{i-1}} \tag{42}$$

$$\underbrace{+\left(\mathbb{E}_{q_{\setminus ij}}\big[I_{i+1}\big] \ln\big[(\frac{1-\hat{\lambda}}{\hat{\lambda}})(\frac{\alpha+\hat{\lambda}}{1-\alpha-\hat{\lambda}}) \big] + \ln\frac{1-\alpha-\hat{\lambda}}{1-\hat{\lambda}} \right)_{F_i^j}}_{\text{effect of } I_{i+1}} \tag{43}$$

$$\underbrace{-\frac{1}{2}(F_i^j - \hat{F}_i^j)^T J_i^{jj}(F_i^j - \hat{F}_i^j) - \sum_{k\neq j}^{m_i}(F_i^j - \hat{F}_i^j)^T J_i^{jk}(\mu_{ik} - \hat{F}_i^k)}_{\text{Laplace approx}} \tag{44}$$

$$\Bigg\},$$

where the operator $()_{F_i^j}$ is the value of the 2D function of the variable in the parentheses at location F_i^j.

Note that if $\alpha = 0$ and $\lambda \ll 1$, then

$$eq.42 + eq.43 \approx \ln\lambda;$$

therefore, we recover the E step in the non-Markov model (eq.20), as desired.

Results

To quantify the benefits of including the Markov terms in the model, we generate data from the Markov model with a range of parameters and perform inference with both the Markov and non-Markov models. Note that the non-Markov model is mis-specified in these simulations, while the Markov model is given the "unfair" advantage of knowing the true value of α. Nonetheless, somewhat surprisingly, our basic conclusion is that incorporating the Markov effects has only a small effect on inference performance. Fig. 11 shows that

the estimation accuracy on individual fluorophores is improved modestly if the Markov terms are included, once α is sufficiently large. However, in Fig. 12 we see that the Markov terms lead to negligible improvement in the overall estimate of λ, which is the main object of interest in many super-resolution imaging studies. Thus we conclude that the "local" information encoded by the Markov terms in the model is mostly redundant with the "global" information encoded by our estimate $\hat{\lambda}$, which is shared across frames to improve our estimate of each I_i.

Acknowledgements

We thank Professor John P. Cunningham for helpful discussions. We thank Wenyu Gao, Jiangfan Zhang, Le Zheng, and Xuexin Wei for testing the code. Funding for this research was provided by Simons Foundation Global Brain Research Award 325398, ONR N00014-14-1-0243, ARO MURI W911NF-12-1-0594, and a Google Faculty Research award; in addition, this work was supported by the Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior/ Interior Business Center (DoI/IBC) contract number D16PC00008. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoI/IBC, or the U.S. Government.

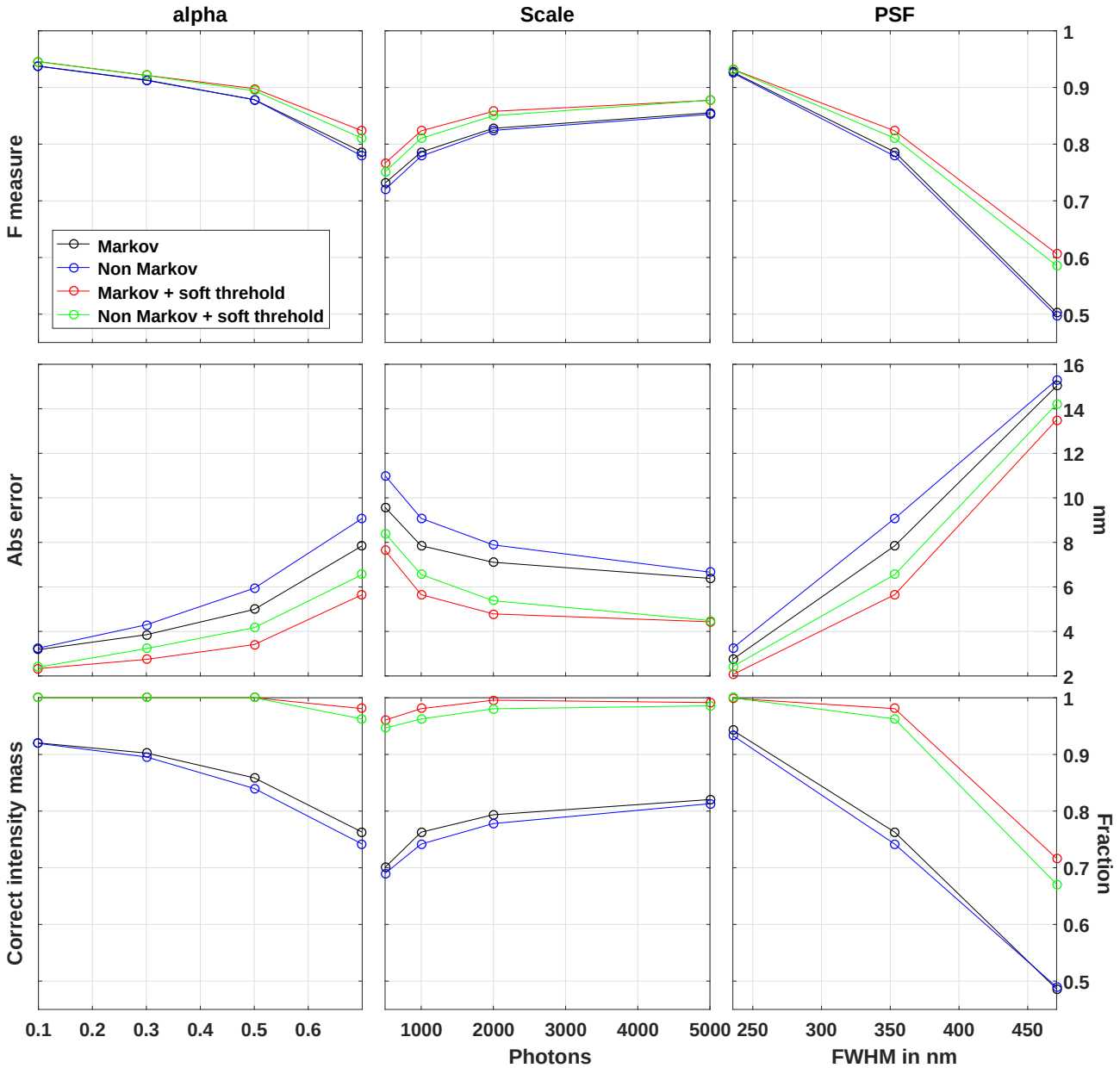

Figure 11: **Evaluation of Markov and non-Markov models as a function of α (First column), *Scale* (Second column), and *PSF* (Third column), with and without soft threholding.** The performance is quantified using the same measures as in the main text. *Scale* is the average number of photons per active fluorophore. We use $N = 2000$ frames. Emission rate p is 0.01. When not indicated otherwise, α, *Scale*, and *PSF* are set to be 0.7, 1000 photons, and 353.25 nm respectively. Note that this is rather large value of α; as seen in the left panel, smaller differences between the Markov and non-Markov models are seen when smaller values of α are used.

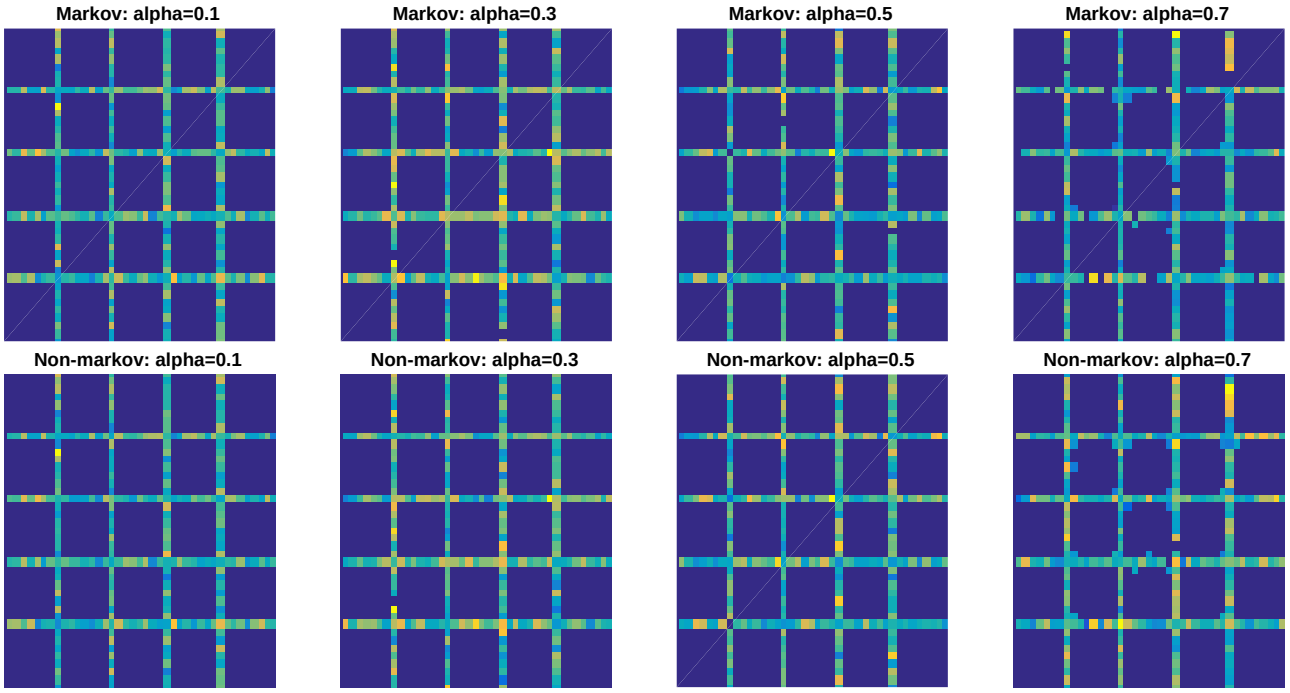

Figure 12: **Estimates of Markov and non-Markov model as a function of α with soft-thresholding.**. The final resolved images of Fig. 11 (First column). The test image was the same grid used in the main text. Recall that α influences the number of active fluorophores (with large α corresponding to more persistent fluorophores and therefore higher fluorophore density); the average fluorophore densities in the four columns shown here are 1.90, 2.55, 3.44, and $5.75\mu m^{-2}$ (left to right). We use $N = 2000$ frames. Emission rate p is 0.01, with an average of 1000 photons per fluorophore. PSF is 353.25 nm. Note that no major differences are seen between the Markov and non-Markov estimates. Similar results were obtained over a wide range of parameters (not shown).