Supplementary Notes

1. Implementation of the compressed sensing algorithm

A major consideration of our implementation of the compressed sensing algorithm is the computation time. Therefore, we subdivided the original image into smaller patches, which not only reduces to total time to analyze the entire image, but also opens the possibility for parallel computing to further improve the computation speed. On the other hand, merging results from different patches becomes important. Our implementation took a simple approach that is facilitated by a small overlap between adjacent patches:

Given: camera image B, super-resolution image x, patch size u-by-v pixels, oversampling factor R, reduced chi-squared target ε, the imaging matrix A and the vector c for a small u-by-v image.

Dividing the camera image, B, into a set of small u-by-v image patches, b, with two pixels shared by adjacent patches.

For each small image, b, do:

1. Create an oversample grid, x, with a size of $R \times (u + 4)$ by $R \times (v + 4)$. The margin of extra 2 pixels on each edge of the patch accounts for the contribution from molecules outside.

2. Minimize $c^T x'$
   subject to $x_i \geq 0$
   and $||Ax' - b'||_2 \leq \varepsilon \times \mathrm{sqrt}(\mathrm{sum}(b))$

For each patch of the optimization result, x, set the outmost 3-pixel boundary to zero to avoid overlap (2 pixel extra marging + one pixel overlap between patches).

Add all x together to form the result of the full image.

The computation time of our compressed sensing algorithm depends on the number of molecules in an image frame. Analyzing a 32×32 pixel^2 image typically takes about 30 seconds on our Intel Xeon X5560 (2.8 GHz) workstation using a single computation thread. The slow computation speed is inherent for such a large optimization problem, and can also be partially attributed to our choice of MATLAB programming language instead of the faster C/C++. On the other hand, the fact that our algorithm performs independent optimizations on small, uniformly

sized image patches makes it easily adaptable to massive parallel computing, especially GPU computation[1,2], for much enhanced speed.

2. Evaluation of the compressed sensing method

We first analyzed the density of molecules identifiable by different methods. In the high photon number simulation which corresponds to bright organic fluorophores (Figure 1b), compressed sensing is much more efficient in identifying densely located molecules. The number of identified molecules using the single-molecule fitting algorithm peaks at 0.58 per μm^2 when the original image has approximately 2 molecules per μm^2. On the contrary, compressed sensing can identify as many as 8.8 molecules per μm^2 out of 12.5 molecules per μm^2 in the original image.

Next, we examined the localization precisions (Figure 1c). At low molecule density, the precision of compressed sensing is just slightly worse than that of single-molecule fitting, presumably due to the finite size of the grid and the unweighted least square constraints used. In simulations with only one molecule in the view field, both methods approach the theoretical limit (9.2 nm for fitting, 11.1 nm for compressed sensing, whereas the Cremer-Rao lower bound (CRLB)[2,3,4] is 8.8 nm; all numbers are FWHM). As expected, the localization precision starts to degrade with high molecule densities. At a density higher than 2 molecules per μm^2, compressed sensing gives slightly better precision than the fitting method.

We also analyzed the same set of simulated data with the previous reported DAOSTORM method using the Python code provided by its authors[5]. Supplementary Figure 4 compares the performance in molecule identification and localization precision. The number of molecules identified by DAOSTORM plateaus at 2.4 to 2.7 per μm^2 when the molecule density exceeds 4 per μm^2, At the same time, the localization error of DAOSTORM in this range of molecule density is about the same as the single-molecule fitting method. At relatively low molecule density (< 2 per μm^2), all three methods show similar localization precision. We note that our DAOSTORM results are consistent with those previously published[5] considering that our simulation uses a smaller average photon number per molecule (3,000 versus 8,333 (5,000 effective)), includes a higher background (70 photons per pixel versus 10), and considers the variation of photon numbers among molecules.

Finally, for the overall temporal resolution, Figure 1d and Supplementary Figure 9 summarizes the minimum number of image frames to attain a given overall spatial resolution. At a resolution of 40 nm (the best allowed by our grid size) to 110 nm (close to what can be achieved by the faster Structured Illumination Microscopy[6]), compressed sensing allows 1/6 to 1/15 of the number of camera frames to be used compared to the single-molecule fitting method. That means a factor of 6 to 15 of improvement in time resolution depending on the desired spatial resolution. For example, only 45 camera frames are needed to support a 100 nm spatial resolution, corresponding to 0.75 sec integration time if the camera operates at 60 frames / sec.

In the two additional sets of simulations with photon statistics corresponding to fluorescent protein mEos2 and a very low photon case (see Supplementary Figs. 6 and 8), we have also seen

the substantial improvement in molecule identification efficiency by compressed sensing. At 750 photons per molecule, compressed sensing identifies 6.7 molecules per μm^2 with a precision of 113 nm. At 200 photons per molecule, compressed sensing identifies 3.8 molecules per μm^2 with a precision of 126 nm. These values suggest an imaging speed that is approximately 75% and 40%, respectively, compared to the 3,000 photons per molecule case if one targets ~ 100 nm resolution.

We must emphasize that the resolution calculation above is for a 2D continuous sample. In real biological samples where the molecules are clustered into one-dimensional (e.g. microtubule) or "zero-dimensional" (e.g. clathrin-coated pits) structures, the number of frames needed to achieve a given molecule density can be even smaller[7].

3. Fundamental limit in molecule density

Compressed sensing is able to recover sparse or compressed signals using very few measurements from a linear system, without knowing in advance the support of the signal (i.e. where the signals are exactly zeros). An important concept in the compressed sensing theory is the restricted isometry principle (RIP)[8], which determines the accuracy and the stability of the signal recovery. A linear system $\boldsymbol{A}$ is said to obey RIP with a sparsity s (i.e., s non-zero elements in x) and an error tolerance δ_s if one has

$$(1 - \delta_s) \, ||\boldsymbol{x}||_2^2 \leq ||\boldsymbol{A}\boldsymbol{x}||_2^2 \leq (1 + \delta_s) \, ||\boldsymbol{x}||_2^2 \quad (1)$$

for all s-sparse vector $\boldsymbol{x}$, where δ_s is the smallest value such that the inequation holds and s-sparse means that $\boldsymbol{x}$ has at most s non-zero elements. To recover s-sparse vectors, it is important to consider δ_{2s}, since the difference between two s-sparse vectors must be a $2s$-sparse vector. If δ_{2s} is sufficiently smaller than 1 (practical implementations require $\delta_{2s} \leq 0.4$)[8], then any pair of s-sparse vectors will have a degree of distinguishability in the data space. The compressed sensing theory shows that s-sparse vectors can be exactly recovered via L1 norm minimization on the estimated signals[9].

Our problem of localizing sparse molecules fits the application scenario of compressed sensing. The condition on the molecule sparsity for exact signal recovery, however, requires a complete RIP analysis. Although such an analysis is complicated in general, here we try to use a simplified case to provide insights into the limit of molecular density that compressed sensing can detect. Specifically, we investigate what is the smallest distance of two molecules ($s = 2$) for them to be "perfectly" detected using compressed sensing. For this purpose, we calculate δ_4 as shown in Supplementary Figure 3 for different minimum distances between molecules. The system PSF is assumed to be a 2D symmetric Gaussian function, and the molecule distances are plotted in unit of PSF standard deviation σ. The δ_4 values are below 0.4 for molecule distances larger than 2.5σ (approximately the FWHM of the PSF), which indicates that the compressed sensing algorithm is able to achieve exact molecule localization for a molecule density up to approximately $0.16/\sigma^2$. Given that our observed PSF has $\sigma = 143$ nm, this limit corresponds to a molecular density of 7.8 μm^{-2}, consistent with fact that in the high photon number case, our molecule identification starts to lose efficiency at a molecule density around 9 μm^{-2}. It is also

consistent with the maximum density of recovered molecules (around 8.8 μm^{-2}) in our simulations despite our overly simplified derivations.

The conventional fitting-based algorithms may allow a similar minimum molecule distance for accurate detection if only a few molecules are present in the imaging field and the total number is known. Compressed sensing is able to automatically detect all the molecule locations without prior knowledge of their total number, as long as their mutual distance (therefore density) is not higher than the value suggested by the δ_4 plot. The advantage is more obvious when one image frame contains a large number of molecules. Finally, we would like to emphasize that a rigorous RIP analysis should calculate δ_s of all s values for the given linear system $\boldsymbol{A}$. The above description should be considered as only a rough estimation of the maximum density of detected molecules in the STORM imaging using compressed sensing, which will hopefully be replaced by future, more solid work.

It is also an interesting topic to compare the compressed sensing approach and the statistical approach. The goal of STORM imaging in general is to determine the molecule locations with some prior knowledge. The available prior knowledge in this problem includes the sparsity of the true signals, the probability distribution of the detected photons and the detector PSF. The advantage of compressed sensing is efficient recovery of sparse signals even if the measured data are heavily corrupted by noise or errors. On the other hand, statistics-based approaches improve the imaging performance by accurate modeling of the probability distribution of the detected photons. These two categories of methods use different prior knowledge in the same application, with different pros and cons. In the study presented in this paper, we adopt a classic form of compressed sensing reconstruction, using very limited prior knowledge of the statistics. As shown in the general framework (Eq. 1), only an estimated noise standard deviation of the measured data is required in our algorithm. However, compressed sensing is compatible with statistics-based approaches. We can further improve our algorithm performance, especially in the case of high noise, by modeling the noise statistics in our optimization framework. This improved framework will achieve the theoretical limit as shown in Ref. 10. It will be one of our future research directions.

References

1. Quan, T.W. *et al.*, "Ultra-fast, high-precision image analysis for localization-based super resolution microscopy", *Opt. Express*, 18, 11867-11876 (2010).
2. Smith, C. S., Joseph, N., Rieger B., & Lidke, K. A, "Fast, single-molecule localization that achieves theoretically minimum uncertainty", *Nat. Methods*, 7, 373-375 (2010).
3. Mortensen, K. I., Churchman, L. S., Spudich, J. A., & Flyvbjerg, H., "Optimized localization analysis for single-molecule tracking and super-resolution microscopy", *Nat. Methods*, 7, 377-381 (2010).
4. Abraham, A.V., Ram, S., Chao, J., Ward, E.S. & Ober, R.J., "Quantitative study of single molecule location estimation techniques", *Opt. Express*, 17, 23352-23373 (2009).

5. Holden, S.J., Uphoff, S. & Kapanidis, A.N. “DAOSTORM: an algorithm for high-density super-resolution microscopy”, *Nat. Methods* **6**, 279-280 (2011).

6. Kner, P., Chhun, B. B., Griffis, E. R., Winoto, L., & Gustafsson, M. G. L., “Super-resolution video microscopy of live cells by structured illumination”, *Nat. Methods*, 6, 339-342 (2009).

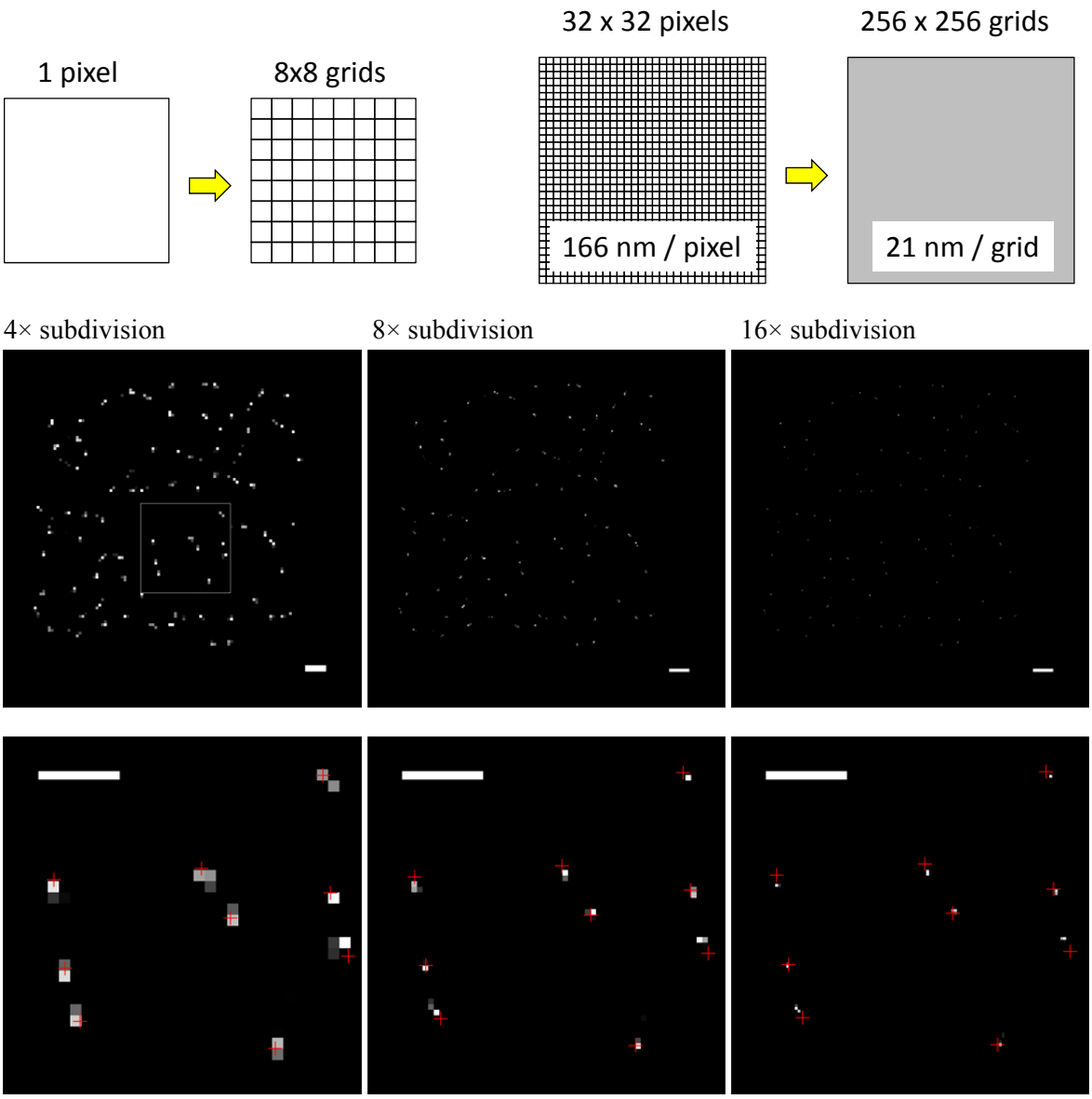
7. Jones, S. A., Shim, S.-H., He, J., & Zhuang, X., “Fast, three-dimensional super-resolution imaging of live cells”, *Nat. Methods*, 8, 499-505 (2011).

8. Candès, E.J., “The restricted isometry property and its implications for compressed sensing”, *Comptes Rendus Mathematique*, **346**, 589-592 (2008).

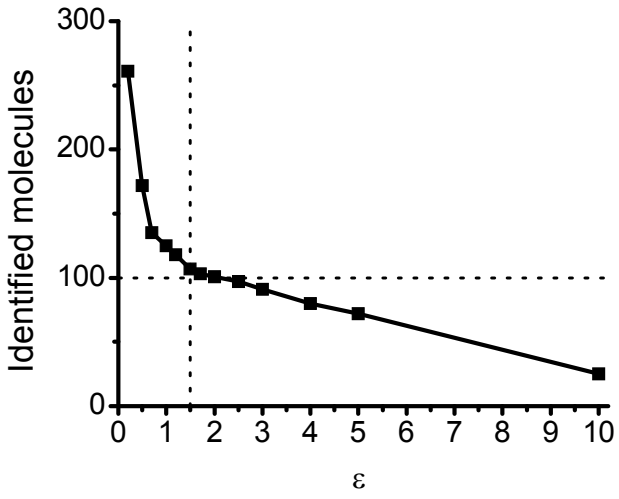
9. Candès, E.J., Romberg, J. & Tao, T., “Stable signal recovery from incomplete andinaccurate measurements”, *Comm. Pure Appl. Math.*, 59 1207-1223 (2005).

10. Willett, R. M. & Raginsky, M., “Performance bounds on compressed sensing with Poisson noise”, *Proceedings of the 2009 IEEE international conference on Symposium on Information Theory*, 1, 174-178 (2009).

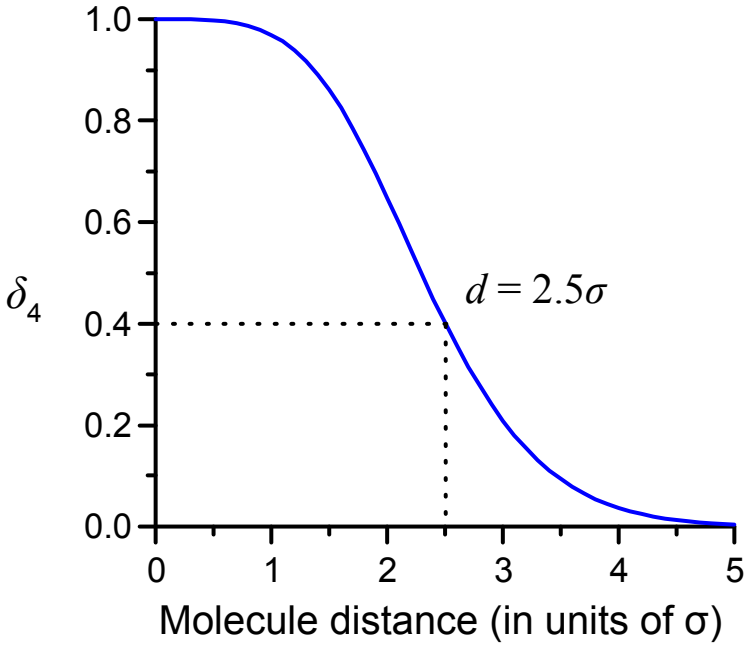
Supplementary Figures



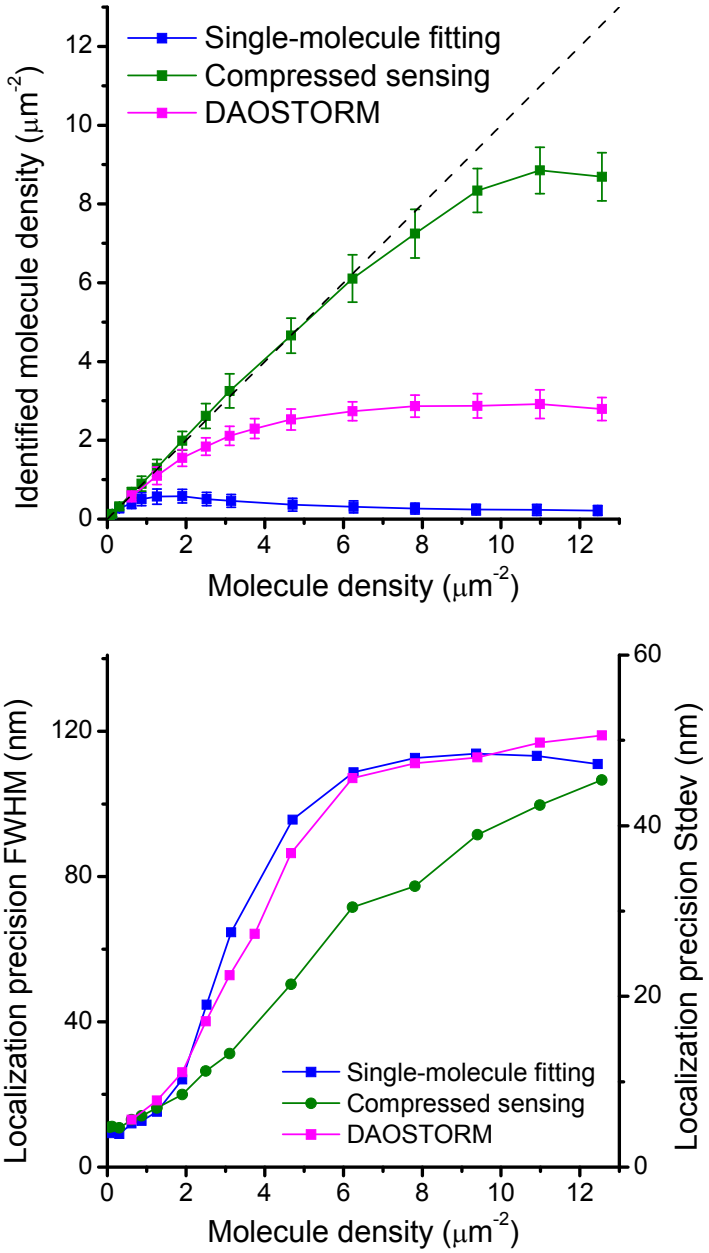
Supplementary Figure 1. Analyzing single molecule images by compressed sensing. (**Upper row**) Scheme of dividing image pixels into oversampled grids. (**Middle row**) Comparing the analysis of the image in Figure 1a using different oversampling factors. (**Lower row**) Zoomed-in displays of the compressed sensing results for the boxed regions. Note that the difference in molecular identification efficiency is relatively small. The computation times for 4×, 8×, and 16× oversampling are 15 sec, 28 sec and 134 sec, respectively. Scale bars: 300 nm.



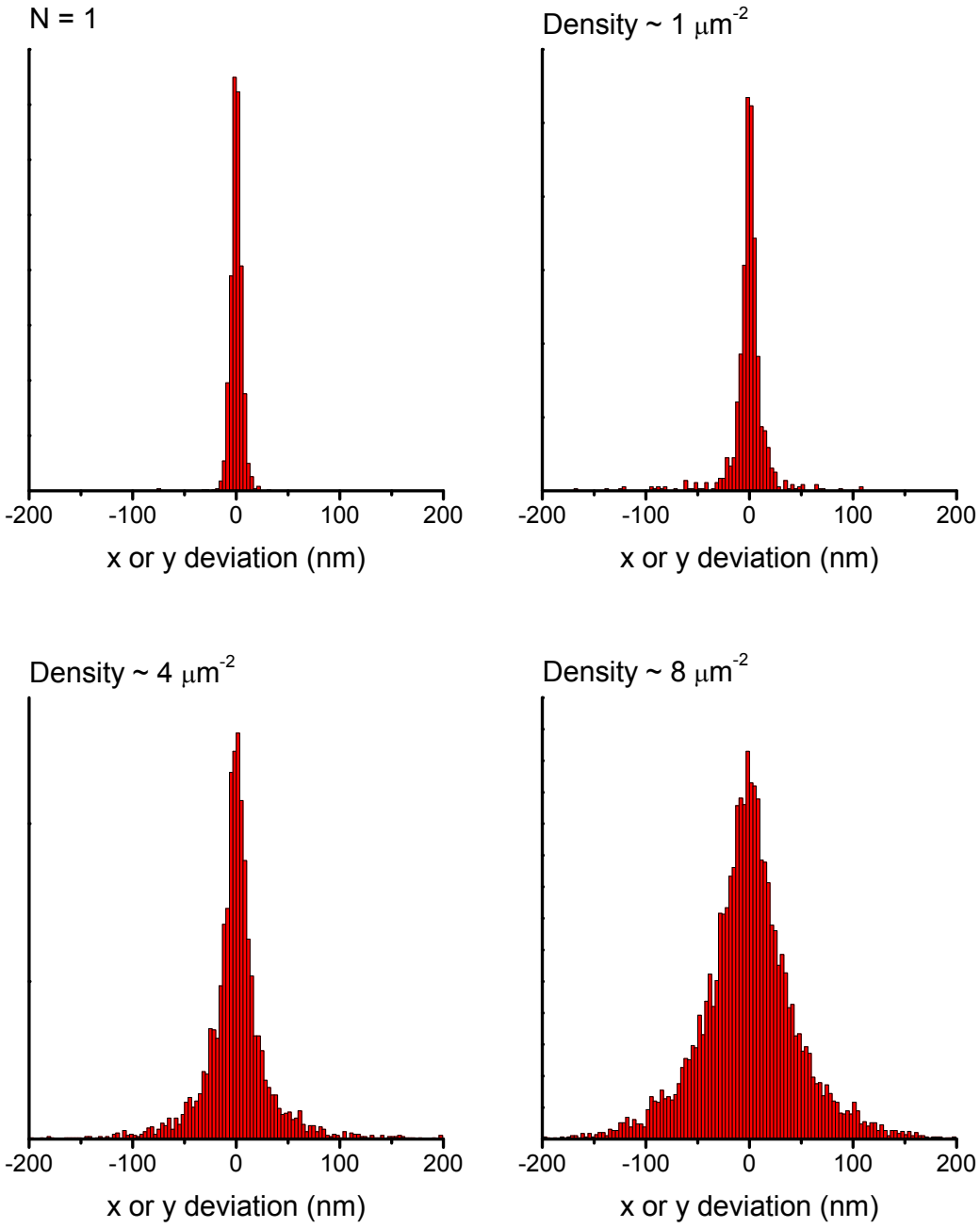
Supplementary Figure 2. Effect of ε on compressed sensing. Showing the number of molecules identified in the simulated image of Figure 1a (100 molecules in the simulation). Because ε corresponds to the reduced χ^2, an ε value of 1 corresponds to perfect modeling of the image if the noise estimation is exact. With an ε value smaller than 1, the image is over fitted which results in over-identification, whereas an ε value larger than 2 enforces the sparsity condition too strongly. In practice, we choose an ε value of 1.5 for all of our analyses (with an extra factor of √2 in analyzing experimental images from EMCCD to account for its extra noise).



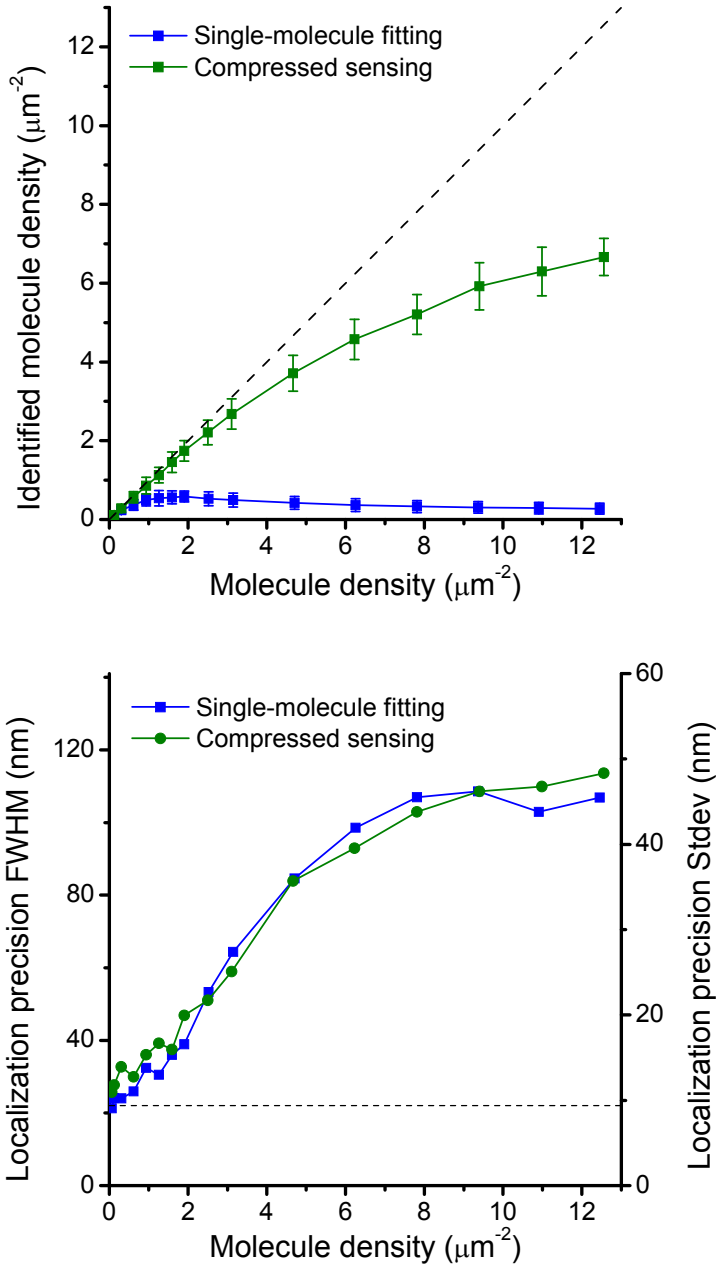
Supplementary Figure 3. Analysis of the limitation in inter-molecular distances for "perfect" recovery using compressed sensing. Plotting the value of δ_4 as a function of the separation between two molecules in units of the standard deviation of the PSF, σ. δ_4 reaches a value of 0.4 (the threshold for perfect recovery, see Candès, E.J., *Comptes Rendus Mathematique*, **346**, 589-592 (2008)) when the separation is larger than 2.5σ, or approximately the FWHM of the PSF.



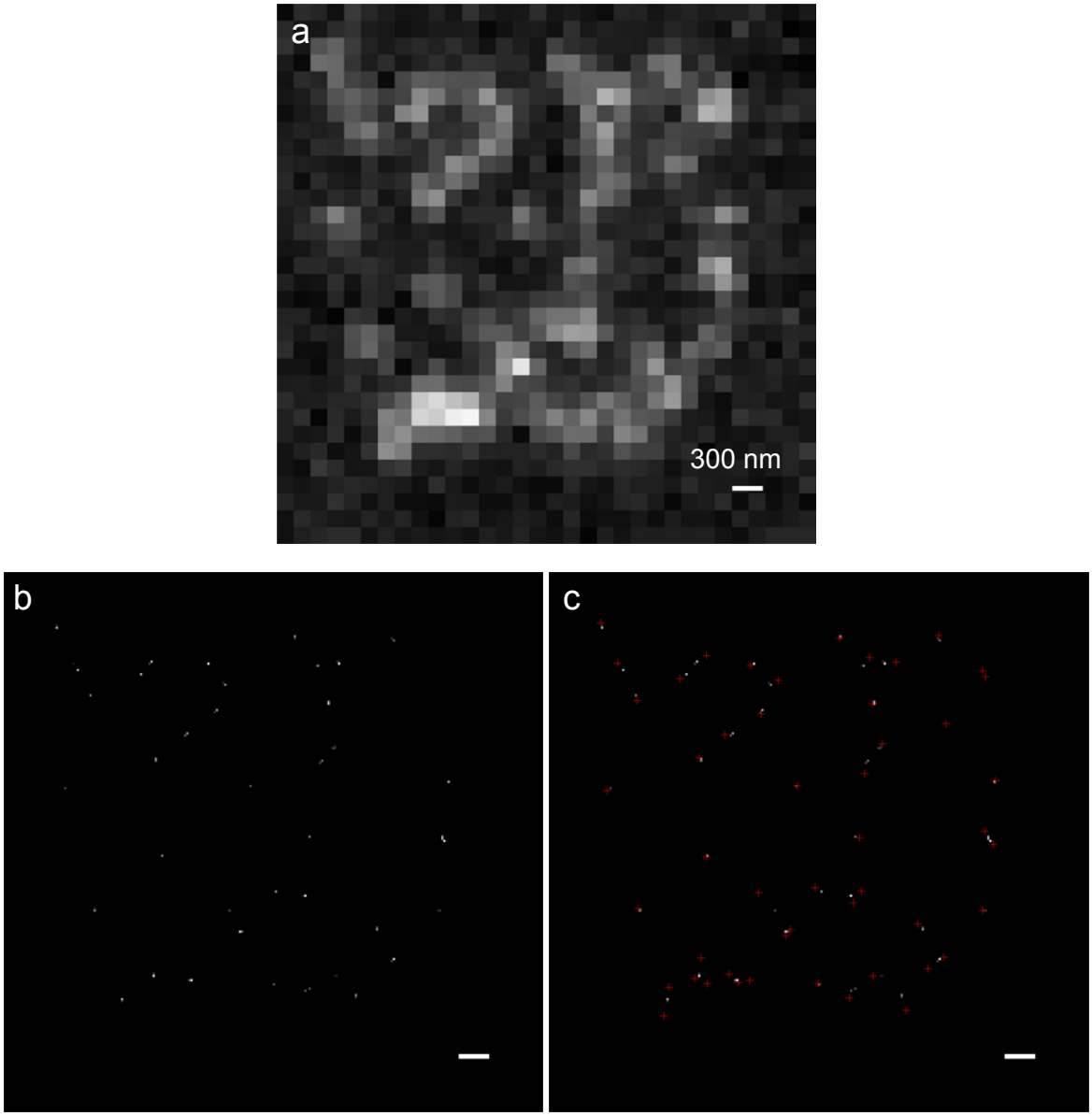
Supplementary Figure 4. Comparing compressed sensing to DAOSTORM in the high photon count case corresponding to the Alexa Fluor 647 dye (mean photon number 3,000 per molecule, background 70 photons per pixel). The same simulation data sets are analyzed by DAOSTORM using the Python code provided by Holden, Uphoff & Kapanidis, *Nat. Methods* **6**, 279-280 (2011). The number of molecules identified and the localization precisions are compared to those of the fitting method and compressed sensing in the same way as in Figures 1b and 1c.



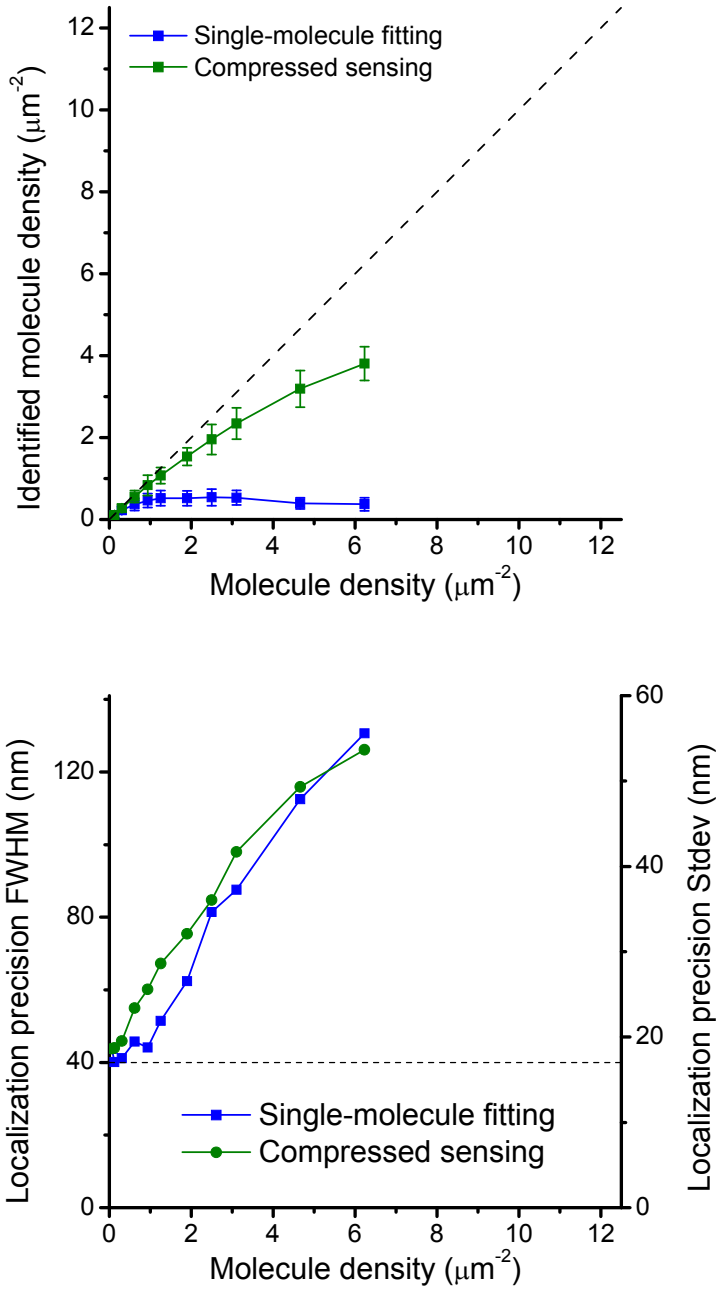
Supplementary Figure 5. Localization precision of the compressed sensing algorithm as a function of molecule density. The analysis is done to simulated data in the high photon count case corresponding to the Alexa Fluor 647 dye (mean photon number 3,000 per molecule, background 70 photons per pixel). Each molecule identified by compressed sensing is matched to the closest "true" molecule position. The two histograms of the *x* offsets and *y* offsets are summed and displayed in each panel. These histograms are fitted with Gaussian functions whose widths are presented in Figure 1c and Supplementary Figure 4.



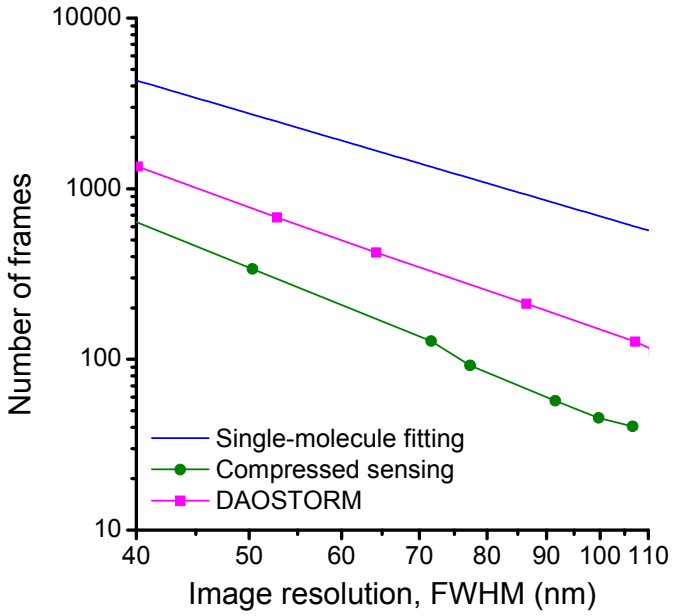
Supplementary Figure 6. Comparing compressed sensing and single-molecule fitting in analyzing simulated data corresponding to the fluorescent protein mEos2 images (mean photon number 750 per molecule, background 50 photons per pixel). The dashed line in the localization precision plot represents the CRLB for single-molecule localization (22 nm FWHM).



Supplementary Figure 7. Compressed sensing analysis in a low signal-to-noise case. The simulation is performed with 50 molecules with 200 photons per molecule and a background of 20 photons per pixel. Showing (**a**) the simulated camera image, (**b**) compressed sensing result, and (**c**) the comparison of the compressed sensing result (red crosses) with the molecule positions in the simulation. Scale bars: 300 nm. It is evident that the result from compressed sensing generally matches well with the simulation input, although closely located molecules are occasionally merged into one because the high noise can make them statistically indistinguishable.



Supplementary Figure 8. Comparing compressed sensing and single-molecule fitting in analyzing simulated low signal-to-noise images (mean photon number 200 per molecule, background 10 photons per pixel). The dashed line in the localization precision plot represents the CRLB for single-molecule localization (40 nm FWHM).



Supplementary Figure 9. Minimum number of frames to achieve a given overall image resolution for a continuous 2D sample based on the simulation results, displayed on a log-log plot with the image resolution set between 40 and 110 nm. The analysis is done to simulated data in the high photon count case corresponding to the Alexa Fluor 647 dye (mean photon number 3,000 per molecule, background 70 photons per pixel). The curve for the fitting method is calculated using a constant 0.58 μm^{-2} identified molecule density, whereas the curves for the compressed sensing and DAOSTORM are calculated using identified molecule densities that allow the corresponding localization precisions to match the desired image resolution.

Supplementary Video

STORM "movie" of microtubules in a living *Drosophila* S2 cell stably expressing mEos2-fused tubulin, with a time resolution of 3 sec. The movie is reconstructed from 4349 camera frames (77 sec) and plays 11 times as fast as real time. Three snapshots from the movie are shown in Figure 2b. Scale bar: 1 μm.