Stochastic approach to the molecular counting problem in superresolution microscopy

Geoffrey C. Rollins, Jae Yen Shin, Carlos Bustamante, and Steve Press

Superresolution imaging methods—now widely used to characterize biological structures below the diffraction limit—are poised to reveal in quantitative detail the stoichiometry of protein complexes in living cells. In practice, the photophysical properties of the fluorophores used as tags in superresolution methods have posed a severe theoretical challenge toward achieving this goal. Here we develop a stochastic approach to enumerate fluorophores in a diffraction-limited area measured by superresolution microscopy. The method is a generalization of aggregated Markov methods developed in the ion channel literature for studying gating dynamics. We show that the method accurately and precisely enumerates fluorophores in simulated data while simultaneously determining the kinetic rates that govern the stochastic photophysics of the fluorophores to improve the prediction’s accuracy. This stochastic method overcomes several critical limitations of temporal thresholding methods.

Significance

Large and complex macromolecular assemblies—like RNA and DNA polymerases, the kinetochore, the ribosome, ATP synthase, and many others—are critical components of the cell. To fully characterize these molecular machines, it is not sufficient to rely solely on traditional structural methods, like cryo-EM and X-ray crystallography that provide great structural detail in vitro; we need experimental and theoretical tools that can describe the organization of these machines in their native cellular environment so as to better understand their function. Here we provide a strategy for extracting precisely this information directly from superresolution imaging data, a state-of-the-art technique for probing biological structures in living cells below the diffraction limit.


The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

1To whom correspondence should be addressed. Email: stevenpresse@gmail.com.

This article contains supporting information online at www.pnas.org (lookup/suppl/doi:10.1073/pnas.1408071112/-/DCSupplemental).
together and assigned to a single PA-FP. Bursts separated by a time longer than \( t_{\text{on}} \) are considered to be from separate PA-FPs. Our method overcomes several important limitations of temporal thresholding methods because it treats the properties of PA-FPs stochastically. First, by contrast to temporal thresholding methods that necessitate advance knowledge of kinetic rates to determine the optimal value of \( t_{\text{on}} \), our method does not require knowledge of kinetic rates beforehand. Kinetic rates are an output of the method, not an input. Second, the method is not tied to a specific kinetic model. We can easily explore alternative kinetic models, such as those with multiple blinking states or with time-varying rates. Third, we can correct for missed transitions which occur if typical PA-FP fluorescence times are as fast as or faster than the frame integration time. Finally, we can use fluorescence intensities (which can vary if multiple PA-FPs fluoresce at once) to inform our estimate for the complex stoichiometry.

In what follows we develop the theoretical methodology and benchmark the method on sample data as well as on synthetic data. We will describe how it is possible to derive from SR data not only the number of PA-FPs, as well as their photophysics, but also determine error bars around all of the estimates.

Results

Analysis of Simulated Data. Consider the model shown in Fig. 1 for the photophysics of a single PA-FP. There are four possible states: inactive (\( I \)), active (\( A \)), dark (\( D \)), or photobleached (\( B \)). Once active, the PA-FP has two options: (i) it can blink possibly transitioning multiple times between the active and the dark state or (ii) it can irreversibly photobleach. Fluorescence is only detected from the active state, not in the other three states. Now consider a collection of \( N \) identical PA-FPs, each of which is governed by the model of Fig. 1.

Thresholding methods used in enumerating PA-FPs do not treat the stochastic nature of the fluorescence bursts explicitly. Thus, these methods cannot separate overlapping PA-FP signals where one FP activates before the last FP has photobleached. These methods also cannot treat superposing signals, which occur when two or more FPs fluoresce at the same time. Overlapping and superposing signals arise when multiple PA-FP spikes occur over a short period. This can happen when (i) there are a large number of PA-FPs; (ii) the PA-FP blinking rate is fast and the PA-FP photobleaching rate is slow; and (iii) the PA-FP photoactivation rate is fast and the PA-FP photobleaching rate is slow.

Without a priori knowing the PA-FP number, \( N \), or the PA-FP’s properties, it is not possible to ascertain whether overlapping and superposing signals are improbable. Thus, in these conditions, if the problem at hand is specifically to determine \( N \), the use of thresholding methods may not be appropriate. The method developed here, by contrast, is applicable for any \( N \) and PA-FP, regardless of its photophysical properties (activation, blinking, and photobleaching) because it explicitly treats them as the stochastic events they are.

To benchmark our method for finding kinetic rates and \( N \) from the data, we first analyzed 200 simulated trajectories, each one representing a collection of \( N = 5 \) PA-FPs.

These simulated trajectories (a sample of which is shown in Fig. S1) were generated using the Gillespie stochastic simulation algorithm (43). That is, the PA-FP photophysical dynamics were simulated stochastically according to the model given in Fig. 1 and the time trace recorded a fluorescence signal only during times in which any of the PA-FPs were in the \( A \) state.

There are several assumptions that these simulations make:

\( i \) We assume that we have a unique dark and active state. If the number of dark and active states is greater than one, we will find that our estimates for the rates will come with large error bars (we will later see this for the case of Dendra2) and, in this case, additional information must be provided to specify the number of states needed.

\( ii \) We assume that the breadth of the distribution over \( N \) for our simulated data arises only from finite data. In reality, finite data provides a lower bound on the breadth of the distribution over \( N \). The remainder of the distribution’s breadth arises from the intrinsic variability in the stoichiometry of the complex which we will discuss later.

\( iii \) We assume all PA-FPs photoactivate. Recent work suggests that the photoactivation efficiency ranges from only 40–60% depending on the PA-FP (44); to correct for this, we would amend our PA-FP topology to add another state to which state \( I \) can irreversibly transition without ever entering state \( A \).

\( iv \) We assume that maturation of all PA-FPs (mEos2 and Dendra2) used in the in vitro experiments is complete; we assume this because the elapsed time—from protein expression to the point where purified protein are obtained—is longer than the maturation half-time (90 and 120 min for Dendra2 and mEos2, respectively) (45, 46).

\( v \) We assume that protein complexes are well separated in space, which is often the case in experiments. For instance, the bacterial flagellar motor complex—with ~22 MotB monomers forming part of each complex—are spatially well separated (20), as are individual kinetochore-associated with six or more proteins linking kinetochore microtubules to centromeric DNA (47). For the complicated case where complexes are closely packed (i.e., packed within region spatially nonresolvable by PALM), our analysis should yield a distribution over \( N \) peaked at multiples of some number, e.g., \( N = 5 \); this, in itself, is important information that would then motivate future experiments in an effort to establish whether complexes aggregate or coincidentally colocalize within an ROI.

We simulated three different blinking scenarios: moderate blinking (set 1), fast blinking (set 2), and slow blinking (set 3). We generated 200 synthetic traces for each scenario. The parameters for the three sets are summarized in Table 1. We tuned the ratio between the blinking rate and the photobleaching rate (\( k_{d}/k_{b} \)) to control the blinking behavior. If \( k_{d} \) is large, relative to \( k_{b} \), then many blinking events will occur before photobleaching. If \( k_{d} \) is much smaller than \( k_{b} \), the PA-FP is more likely to photobleach without blinking.

To find the model parameters (\( k_{d}, k_{b}, k_{a}, k_{r}, N \)) from the synthetic data, we use a maximum-likelihood procedure detailed in Materials and Methods. The likelihood function is based on the likelihood of observing a particular on-off fluorescence time trace in the data, which is then maximized with respect to the model parameters.

In the analysis of these data sets, we used a bootstrapping approach (resampling data with replacement) to determine the precision of our parameter estimates (48, 49). We randomly selected...
molecules were immobilized on a streptavidin-coated glass cover-
slip, in a manner such that oligomerization was negligible and
molecules were spatially well separated enough to be detected
individually \((N = 1)\). The sample was simultaneously illuminated
with 405-nm and 561-nm lasers to activate (from a precursor
state to \(I\)) and then excite (from \(I\) to \(A\)) the Dendra2 molecules,
respectively, until all of the molecules were photobleached (see
Lee et al. (40) for further details). Individual emission bursts
from the EMCCD output were processed into single-molecule
time traces for analysis as described in ref. 40.

We simultaneously extracted the four kinetic rates \((k_a, k_d, k_b, k_r)\)
and found that Dendra2 blinking probability is low
\((k_d/k_b = 0.3)\); its behavior most closely resembles that of set 3 of
our simulated data. Dendra2 molecules are more likely to pho-
tobleach upon activation than blink. Our \(N\) distribution is sharply
peaked at 1 (Fig. 6), as expected for this data set, because the
experiments were designed to separate and isolate Dendra2
molecules on the coverslip. Our rate estimates (drawn from 300
bootstrap iterations) compare well with those of Lee et al. (40),
who found a similar blinking rate (Table 2). Our results differ
from their results most significantly for \(k_r\), the rate of recovery
from \(D\) to \(A\). Lee et al. (40) fit the distribution of fluorescence-off
times to determine \(k_r\) and found that the distribution fit poorly to
a single exponential, but that it was well fit by a double exponential
\((k_{d1} e^{-k_d t} + ak_{r2} e^{-k_r t})\). The poor fit to a single exponential agrees
with the fact that our histogram of \(k_r\) is heavy tailed toward slower
rates (Fig. 6). Their findings and ours suggest that the kinetic
model of Dendra2 (Fig. 1) should be amended to two dark states
rather than one. We will explore the possibility of including more
dark states, and the implication of uniquely identifying a model
when using more dark states, in future work.

**Missed Transitions May Be Necessary in Tackling Larger Data Sets.** To
tackle more challenging data sets, we need to correct for missed
transitions when PA-FP residence times in the dark or active
state are on the order of the temporal resolution or data ac-
quisition time \((t_d, t_a)\). The mathematical details are relegated to Materials and Methods. Typically, the lower bound
for this time resolution is limited by the camera’s acquisition
rate, i.e., the frame integration time.

In Fig. 7, **Left**, we show a distribution over \(N\) obtained by
combining 10 traces \((N = 10)\) of the \(N = 1\) from the in vitro data

---

### Table 1. Transition rates for simulated data

<table>
<thead>
<tr>
<th>Rate constant</th>
<th>Set 1</th>
<th>Set 2</th>
<th>Set 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>(k_a)</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>(k_d)</td>
<td>3.0</td>
<td>10.0</td>
<td>0.1</td>
</tr>
<tr>
<td>(k_r)</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>(k_b)</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>(k_d/k_b)</td>
<td>3.0</td>
<td>10.0</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Kinetic rates used to generate simulated data sets. Rates in units of Hz.

---

Fig. 2. Histogram of bootstrapping results from simulated data set 1. Histogram of bootstrapping results from simulated data set 1 with 200 traces (200 bootstrap iterations). The theoretically expected results are shown in the dotted line.
to generate more challenging yet controlled data sets. We generated 1,000 such traces. When \( t_d = 0 \), we underestimate \( N \) by a factor of \( \sim 2 \) (even after 200 bootstrap iterations). This underestimation is most likely due to the missed transitions of the active state rather than of the dark state. Because the mean dwell time in the active state (83 ms \( \sim 1/(k_d + k_b) \)) from Table 2 is about as long as the data integration time (50 ms), we expect to miss about half the transitions to the active state, which agrees well with our result (\( N = 5 \) using \( t_d = 0 \) instead of the expected \( N = 10 \); Fig. 7). By contrast, the dwell in the dark (1,100 ms) is much larger (\( \sim 20 \) times) than this integration time. We did not undercount when we considered the \( N = 1 \) (Fig. 6) case because the probability of observing exactly \( N = 0 \) is zero because the first fluorescent peak is automatically counted as one emission burst, which obviously corresponds to the activation of a new molecule. We did not overcount because our theory correctly differentiates between blinks and new photoactivations (as benchmarked for our simulated data sets).

To approximately correct for the undercounting when \( N \) is larger than 1, we increased \( t_d \) to 50 ms, which considerably improves our \( N \) estimate (Fig. 7).

Next, if we set the dead time closer to the active-state dwell time, \( t_d = 100 \) ms the distribution shifts still further toward the correct value (Fig. 7).

To avoid using this approximate treatment, a goal should be to develop PA-FPs that blink with kinetics slower than the data’s acquisition rate.

**Making Use of Intensity Measurements.** Previously, we showed how to correct for missed transitions in dealing with real data sets. Here we describe how to use fluorescence intensity data by separating bright states into several distinct aggregates also relevant to analyzing real data sets (Fig. 8).

Fig. 3. Histogram of bootstrapping results from simulated data set 2. Histogram of bootstrapping results from simulated data set 2 with 200 traces (200 bootstrap iterations). An overall bias toward slow \( k_d \) is observed in the \( k_d \) distribution. We will discuss how to correct for missed transitions, which gives rise to this bias, later. The theoretically expected results are shown in the dotted line.

Fig. 4. Histogram of bootstrapping results from simulated data set 2 with 5-ms time bins. We followed the same procedure as in Fig. 3 except we now used 5-ms—as opposed to 50-ms—time bins.
to extract all rates as well as \( N \) (Fig. 9). The estimate for \( N \) sharpens dramatically, even when trying to estimate larger \( N \), when information on the rates is provided and the data are instead entirely focused on estimating \( N \) more precisely (Fig. 5). All code used to generate the results can be obtained by emailing the corresponding author.

**Discussion**

Common biochemical/biophysical methods used to establish the stoichiometry of protein complexes often lack single molecule resolution. For example, in the ion-channel literature the following methods are common: densitometry of silver-stained and Western-blotted protein complexes (in quantifying atrial inward rectifying potassium channel complex stoichiometry) (50); voltage-clamp electrical recordings [combined with channel inhibitors release to count potassium voltage gated channel (KCNE) subunits] (7); LC-MS (in counting subunit e used in bovine heart mitochondria ATP synthase self-association) (51); coimmunoprecipitation/immunocytochemistry (to assess colocalization of potassium voltage-gated channels such as KCNE2 to KCNQ1 and KCNE1 in cardiomyocytes) (17).

Fluorescence-based methods are also common in enumerating proteins but have severe limitations. For instance, flow cytometry (52, 53) yields estimates for intracellular protein number, but provides no spatial resolution and is less accurate for proteins numbering less than a few hundred (53). By contrast, quantitative fluorescence microscopy (54–56), which is diffraction limited, provides a maximum spatial resolution of a few hundred nanomoles and is commonly used in protein counting, but still provides an order of magnitude less resolution than PALM or STORM (stochastic optical reconstruction microscopy), another SR technique. Using quantitative fluorescence, Joglekar et al. (54) counted proteins in the kinetochore of budding yeast using the fluorescence intensity from tagged Cse4p (57–59) as a fluorescence intensity standard. This indirect method can lead to counting errors because standards themselves can be poorly characterized (18, 19).

Stepwise photobleaching methods, in turn, count proteins by monitoring discrete fluorescence intensity drops as individual proteins photobleach over one ROI. This method was used to count KCNE1s in the cardiac channel \( I_{Ks} \) (5). Photobleaching does not require fluorescence standards but is still diffraction limited (20–23). Furthermore, the noise inherent to photobleaching data limits the method to complexes with less than 5–6 proteins (21).

By comparison, the method presented here provides a principled recipe for enumerating PA-FPs while gathering critical photophysical properties of the PA-FPs directly and self-consistently from the same SR data set used to determine \( N \).

A qualitative reason explaining why likelihood functions peak at the correct value of \( N \) (Fig. S3) is illustrated with the following example. Suppose we start by guessing that all fluorescence spikes in a time trace are due to one PA-FP (i.e., \( N = 1 \)). Thus, all events, except the first photoactivation, are blinks. Assuming we only have a single \( D \) state, all waiting times between successive fluorescence spikes should therefore be sampled from this distribution: \( e^{-k_t t} \). However, if we have many PA-FPs, not all waiting times will be sampled from \( e^{-k_t t} \), and the likelihood of \( N = 1 \) will be small. We can repeat a similar reasoning for any \( N \) value.

![Fig. 5. Histogram of \( N \) obtained by assuming rates. We demonstrate that we can accurately determine \( N \) if the rates are known. We use parameters from set 1 from Table 1 and use the simulated data to determine \( N \) (and not the rates). (Left) Shown with 50-ms time resolution (and therefore underestimated \( N \)). (Right) Improvement afforded by smaller time resolution (5 ms).](image1)

![Fig. 6. Histogram of bootstrapping results from in vitro data. Histogram of bootstrapping results from in vitro data with 1,000 time traces (300 bootstrap iterations). The dashed lines show the parameter values from Lee et al. (40).](image2)
and shows how one can take advantage of intensity problem from SR data, it includes the treatment of the Materials and Methods measurements.

shows that PA-FPs with kinetics slower than the integration time should improve the estimate for number of macrostates.

Lee et al. (40). Rates in units of Hz.

Increasing the dead time, \( t_d \), helps approximately correct for missed transitions. We combine 10 \( N = 1 \) in vitro traces and use the combined traces to determine rates and \( N \) (whose theoretical value should be \( N = 10 \)). We generated 1,000 of such traces (with 200 bootstrap iterations). The distribution over \( N \) is shown for increasing values of \( t_d \). Dendra2 has a dwell time in the bright state, which is on average approximately as long as the time resolution of experiment (50 ms). Thus, when \( t_d \) is zero, approximately half the transitions to the bright state are missed. As \( t_d \) increases, our estimate for \( N \) improves. Fig. 4 shows that PA-FPs with kinetics slower than the integration time should improve the estimate for \( N \).

<table>
<thead>
<tr>
<th>Table 2. In vitro Dendra2 kinetic rates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate constant</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>( k_a )</td>
</tr>
<tr>
<td>( k_d )</td>
</tr>
<tr>
<td>( k_f )</td>
</tr>
<tr>
<td>( k_s )</td>
</tr>
<tr>
<td>( k_d/k_b )</td>
</tr>
</tbody>
</table>

In vitro Dendra2 kinetic rates from the current AMM analysis and from Lee et al. (40). Rates in units of Hz.

Another decisive advantage of our method is its ability to generate entire distributions of \( N \) for a given complex; this is particularly relevant given that protein complex stoichiometry may vary from complex to complex based on the dissociation constant of the various subunits in those complexes (5, 60), tissue development, and even its disease state (8, 61, 62). Developing therapeutic agents against some heart diseases, for instance, may require a quantitative assessment of protein complex composition in living tissues (7, 62–64) and finding drugs to restore native complex stoichiometries is one such strategy (8, 62).

Our method is a first step in the quantitative determination of the number of PA-FPs we allow to be simultaneously photoactive. We use this as a way to tune the size of the state space to save on computational time in situations where we expect photoactivation events to be well separated in time.

Model Kinetics. The transition rate from macrostate \( s_i \) to macrostate \( s_j \) is simply the transition rate for one PA-FP multiplied by a combinatoric factor (for the population of the appropriate microstate). The macrostate transition rates are summarized in Table S2.

The dynamics of the PA-FP AMM are governed by a rate matrix, \( Q \). Each off-diagonal matrix element \( q_{ij} \) equals the transition rate of \( s_i \rightarrow s_j \). The diagonal elements are set so that each row sums to zero: \( q_{ii} = -\sum_{j \neq i} q_{ij} \). The macrostate transition probabilities at any time \( t \) are given by the Kolmogorov equation (34):

\[
\frac{dP(t)}{dt} = P(t)Q
\]

whose solution is given by

\[
P(t) = P(0)e^{Qt}.
\]

In the case of dark and bright observation classes, we can partition the rate matrix \( Q \) into four submatrices, based on the dark (subscript \( d \)) and bright (subscript \( b \)) aggregated classes:

\[
Q = \begin{pmatrix}
Q_{dd} & Q_{db} \\
Q_{bd} & Q_{bb}
\end{pmatrix}
\]

The submatrix \( Q_{dd} \) contains the rates for transitions from states in class \( d \) to other states in class \( d \); \( Q_{bb} \) contains the rates for transitions from the states in class \( d \) to the states in class \( b \). The other two submatrices are similarly defined.

Another possibility would be that multiple levels of fluorescence are observable: bright1, bright2, bright3, etc. For example, in the case of two bright aggregated classes, macrostates with \( A = 1 \) would be grouped into the bright1 class, which we denote \( b_1 \). Macrostates with \( A > 2 \) would be grouped into the bright2 class, which we denote \( b_2 \). By extension, if three intensity
levels were resolvable, \( b_2 \) would contain macrostates with \( A > 1 \) and class \( b_3 \) would contain macrostates with \( A > 2 \). In the case of one dark class and two bright observation classes, our rate matrix would be partitioned into nine submatrices:

\[
\mathbf{Q} = \begin{pmatrix}
Q_{dd} & Q_{dh_1} & Q_{dh_2} \\
Q_{bd_1} & Q_{bh_1} & Q_{bh_2} \\
Q_{bd_2} & Q_{bd_1} & Q_{bh_2}
\end{pmatrix}
\]

Having multiple bright aggregates is advantageous because it shrinks the size of the largest submatrix in the likelihood calculation. Furthermore, if the time traces infrequently visit the higher brightness classes, then fewer exponentiations of the largest submatrix will be necessary.

**Likelihood Function.** The likelihood function that we describe in this section provides an answer to the following question: Given a kinetic model and a set of kinetic rates, what’s the likelihood of observing the data? Our goal is to determine the kinetic rates and \( N \) that maximize the likelihood function, i.e., maximize the likelihood of observing the data that was collected by PALM.

Suppose we have a trajectory of \( L \) dwells representing the dynamics of a collection of \( N \) PA-FPs. Associated with each dwell is an observed aggregated class and a dwell time. The set of observation classes is \( h = \{ h_1, \ldots, h_L \} \). The set of dwell times is \( t = \{ t_1, \ldots, t_L \} \). So, during dwell \( i \) we observe class \( h_i \in \{ d, b_1, b_2, b_3 \} \) for duration \( t_i \). See Fig. 10 for an illustration. The probability densities for dwelling in the dark class for time \( t \) and then transitioning to the bright class are given by the elements of the following matrix:

\[
G_{db}(t) = e^{Q_{dd}t}Q_{db}.
\]

The \((i, j)\)th element of \( G_{db} \) is the probability density of entering class \( d \) from its \( i \)th state, dwelling in class \( d \) for time \( t \), and then transitioning to the \( j \)th state of class \( b \).

When we have missed transitions, we implement renormalized transition matrices—instead of transition matrices like those given by Eq. 5—that account for missed transitions to the bright state, say, by resuming over all possible missed events \((37, 65)\):

\[
G_{db}(t) = \exp(Q_{dd}t)\exp\left( (Q_{dd} - Q_{db}[t - e^{Q_{bd}t}]Q_{bd})(t - t_d) \right)Q_{db}.
\]

Next, we wish to calculate the likelihood of the dwell trajectory \( h \), given the model parameters \( \theta \), where \( \theta \) is the set of parameters \( N \) and the transition

---

**Fig. 8.** Using intensity measurements can improve the estimate for \( N \). (A) When \( N = 5 \) we can observe, in principle, a fluorescence intensity ranging from one to five bright (if all PA-FPs are on simultaneously). Here we show how many states coincide with the different fluorescence intensities up to three simultaneously bright. (B) We show that we can take advantage of the intensity measurements to compute a correctly centered distribution for \( N \) (using 200 synthetic traces with 200 bootstrap iterations).

**Fig. 9.** Intensity measurements become particularly useful when estimating rates and \( N \) for larger \( N \). Here we used simulated set 1 with \( N = 10 \) (200 synthetic traces with 200 bootstrap iterations). If the data are instead focused on determining only \( N \), rather than \( N \) and the rates, the distribution over \( N \) becomes dramatically sharper even for larger \( N \) (Fig. 5).
Fig. 10. Likelihood function. An idealized time trace. Each dwell is color-coded with its corresponding term in the likelihood function.

\[ f(t|h | \theta) = P_{\text{init}} G_{\text{db}(t)} G_{\text{bf}(t)} G_{\text{db}(t)} \exp(-Q_{\text{dd \ final}}) P_{\text{final}} \]

where \( P_{\text{init}} \) is a probability vector with all probability density in the fully inactive macrostate. The parameters, \( \theta \), determine the elements of the rate matrix \( Q \). The final factor of \( e^{-Q_{dd \ final}} \) comes from the fact that all of the PA-FPs irreversibly photobleach by the end of the trajectory. After all photobleaching events occur, the system will dwell in the dark class indefinitely; we represent this with a long final dwell in the dark class for time \( t_{\text{final}} = 10^4 \) s. \( P_{\text{final}} \) is a probability vector with all probability density in the fully photobleached macrostate.

Numerical Evaluation of the Likelihood Function. Our goal is to find the set of parameters \( \theta \) that maximize the likelihood function given the data (a dwell trajectory represented by \( t \) and \( h \); in practice, this is accomplished by minimizing \(-\log(f(t|h | \theta))\) with respect to the rates for a fixed value of \( N \), and then repeating this process for other values of \( N \)). The optimization is performed in Python using the limited-memory Broyden-Fletcher-Goldfarb-Shanno (BFGS) minimizer, implemented in Scipy (66). BFGS is a quasi-Newton method that performs well on nonsmooth optimization problems. The computational time is plotted in Fig. 55. The scaling depends on \( A_{\text{max}} \), and it is advantageous to set \( A_{\text{max}} \) to a small value when possible (when activation events are well separated).

As an aside, we point out that \( A_{\text{max}} = 1 \) is still distinct from thresholding methods. Consider the following scenario: a PA-FP activates, and then blinks. Whereas the first PA-FP is in the dark state, a second PA-FP activates and photobleaches before the first molecule recovers from blinking. This scenario would still obey \( A_{\text{max}} = 1 \), but it would be forbidden in thresholding.

To determine whether likelihood maximization runs converge to the correct parameter values, as a preliminary test we computed 1D slices in parameter space around the true value of each parameter. Each panel of Fig. 53 was obtained by holding four of the five parameters \( (k_{\text{on}}, k_{\text{off}}, k_{\text{b}}, k_{\text{d}}) \) constant at their true values and then varying the remaining parameter over a range encompassing its true value. We see that the four kinetic rates are peaked in the correct location at their true values. In the 1D slice for \( N \), we see a maximum (as determined by the value of the likelihood function) that spans \( N = 4 \) and \( N = 5 \).

Next, we assessed the ability of the numerical maximization procedure to converge to the correct parameter values. The 1D slices discussed above suggest that the likelihood function maximum is in the correct region of parameter space, but a separate question is, Can we simultaneously determine all five parameters? We found that the likelihood-maximization procedure converges to the correct kinetic rates within 100 cycles. Fig. 56 shows the convergence of the rate estimates from one of the maximization runs. For simulated fast (set 2) and slow blinking (set 3), we found that convergence of the likelihood maximization also occurred within \(-100 \) cycles.

ACKNOWLEDGMENTS. We thank S. H. Lee and A. Lee for providing the Dendra2 in vitro data and for numerous helpful discussions about PALM and the blinking paradigm. Support for this work was provided by a National Science Foundation (NSF) Graduate Research Fellowship (to G.C.R.), the Burroughs Wellcome Fund and NSF Grant MCB-1412259 (to S.P.); and the Howard Hughes Medical Institute and NIH Grant R01-GM032543 (to C.B.).

Fig. S1. Simulated time traces for a collection of PA-FPs. Sample traces from Gillespie simulations with parameters: \( N = 5 \) and rates from data set 1.

Fig. S2. Histogram of bootstrapping results from simulated data set 3. Histogram of bootstrapping results from simulated data set 3 with 200 traces (200 bootstrap iterations). The theoretically expected results are shown in the dotted line.
Fig. S3. One-dimensional slices through the likelihood function of simulated data set 1. In each panel, one parameter is varied while the other parameters are held at their true values. We find the likelihood is maximized at the true parameter values.

Fig. S4. Growth of state space with number of PA-FPs. As the number of PA-FPs increases, the number of macrostates grows exponentially. We tune the growth rate by setting $A_{\text{max}}$, the number of PA-FPs that we allow to be simultaneously active.

Fig. S5. Algorithm time scaling. (Left) Computation time for a single matrix exponential calculation. (Right) Estimated time for full likelihood maximization.
Fig. S6. Convergence of likelihood maximization of data set 1. Convergence of likelihood maximization of data set 1. The maximization converges close to true the parameter values. The parameters are estimated from the results of many independent maximization runs, like the one depicted here. In these runs, $N = 5$. Similar plots can be generated for the other data sets.

Table S1. Macrostates for a collection with $N = 2$ PA-FPs

<table>
<thead>
<tr>
<th>Macrostate</th>
<th>${l,A,D,B}$</th>
<th>Aggregated class</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s_1$</td>
<td>${2,0,0,0}$</td>
<td>Dark</td>
</tr>
<tr>
<td>$s_2$</td>
<td>${0,0,2,0}$</td>
<td>Dark</td>
</tr>
<tr>
<td>$s_3$</td>
<td>${0,0,0,2}$</td>
<td>Dark</td>
</tr>
<tr>
<td>$s_4$</td>
<td>${1,0,1,0}$</td>
<td>Dark</td>
</tr>
<tr>
<td>$s_5$</td>
<td>${1,0,1,1}$</td>
<td>Dark</td>
</tr>
<tr>
<td>$s_6$</td>
<td>${0,0,1,1}$</td>
<td>Dark</td>
</tr>
<tr>
<td>$s_7$</td>
<td>${0,2,0,0}$</td>
<td>Bright</td>
</tr>
<tr>
<td>$s_8$</td>
<td>${1,1,0,0}$</td>
<td>Bright</td>
</tr>
<tr>
<td>$s_9$</td>
<td>${0,1,1,0}$</td>
<td>Bright</td>
</tr>
<tr>
<td>$s_{10}$</td>
<td>${0,1,0,1}$</td>
<td>Bright</td>
</tr>
</tbody>
</table>

Macrostates for a collection with $N = 2$ PA-FPs.

Table S2. Kinetic rates of a collection of PA-FPs

<table>
<thead>
<tr>
<th>Transition type</th>
<th>Change in population</th>
<th>Rate $s_i \rightarrow s_j$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activate</td>
<td>${-1, +1,0,0}$</td>
<td>$N_{ij}k_{sl}(t)$</td>
</tr>
<tr>
<td>Blink</td>
<td>${0, -1, +1,0}$</td>
<td>$N_{ij}k_{ld}$</td>
</tr>
<tr>
<td>Recover</td>
<td>${0, +1, -1,0}$</td>
<td>$N_{ij}k_{lr}$</td>
</tr>
<tr>
<td>Photobleach</td>
<td>${0, -1,0, +1}$</td>
<td>$N_{ij}k_{lb}$</td>
</tr>
</tbody>
</table>

Kinetic rates of a collection of PA-FPs.