Biophysical Journal, Volume *112*

Supplemental Information

Reduction of Confinement Error in Single-Molecule Tracking in Live Bacterial Cells Using SPICER

Christopher H. Bohrer, Kelsey Bettridge, and Jie Xiao

Contents

1 Methods **1**

2 Applying SPICER to 2d tracking data. **3**

3 Varying dimensions within the trajectory: **5**

4 Calculating the Likelihood of Multiple State Trajectories: **6**

5 SI figures referenced in the main text **7**

1 Methods

Simulations of SMT trajectories with two states

All simulated SMT trajectories used in this work were generated by the software provided in vbSPT using a rod-shaped cell like geometry (unless stated specifically, cell radius = 500 nm and cell length = 2.5 μm) and a single molecule localization error of 20nm (1). The diffusion coefficients defined in this work take into account this localization error with a time step of 5 ms. The length of individual trajectories follows an exponential distribution with a mean value of 6 steps (each step is 5 ms). The effect of confinement is reflected in the simulation through reflective boundaries at the cell membrane. In the two state model, we assume that State 1 and 2 are defined by two diffusion coefficients D_1 and D_2, and the transition probabilities between them are P_{12} and P_{21}, respectively. The reaction scheme for the maximum likelihood analysis is:

$$D_1 \underset{P_{21}}{\overset{P_{12}}{\rightleftharpoons}} D_2. \tag{1}$$

Parameters used in each simulated system in this paper are listed in each corresponding figure.

Maximal Likelihood method to identify parameters

We first convert each SMT trajectory to a SPICER trajectory using a fixed R-value, as defined in Figure 1. The R-value defines the confinement zone (red), and is the distance from the membrane boundary of the cell to the edge of the midcell region where the molecule diffuses freely and does not experience confinement (green). We then take all the converted trajectories

and scan the parameter space of the diffusion coefficients D_1, D_2 and transition probabilities P_{12}, P_{21} to obtain the best fit parameters for the system by maximizing the likelihood using a Markov Chain Monte Carlo (MCMC) approach with a preset number of search steps (2). The MCMC approach begins by selecting a random set of parameters D_1, D_2, P_{12} and P_{21}, and calculates the corresponding summed log likelihood value from all trajectories. A detailed description of the calculation can be found in the section Calculating the Likelihood of Multiple State Trajectories. The process is then iterated by systematically adjusting one of the parameters, chosen at random, by a small amount and then comparing the log likelihood at the new parameter value to the previous log likelihood. If the log likelihood is greater at the new value, the algorithm stays at the new position in parameter space. If the log likelihood is less than the old value, the algorithm takes the difference of the log likelihoods, and two outcomes can happen: 1. If the difference is less than the log of a uniform random number it accepts the new position 2. If the difference is more than the log of a uniform random number, the algorithm stays at the old position. The process repeats by adjusting a new randomly chosen parameter until it reaches a preset number of steps. In all analyses used in this work the number of steps was set at a number large enough so that all the parameters converge well before the end of step numbers.

The stochasticity in the parameter search allows the algorithm to fluctuate around parameters, defining a degree of uncertainty and avoiding local minimums in the parameter search (2). (An example of a parameter scan on a system is shown in Figure S2.) We used the log of the likelihood and summed up the log likelihood of each of the individual trajectories to incorporate the information from multiple trajectories, see Das *et al.* for the specific algorithms used in this work (2). The parameters that give the maximum log likelihood are identified as the best-suited parameters for the system. The percent error in this work is defined as $|X_{cal} - X_{true}|/X_{true} \times 100$.

Single molecule tracking data collection and analysis

Single molecule tracking was performed on live MG1655 *E .coli* cells using a photoactivatable fluorescent protein PAmCherry labeled RNA polymerase (RNAP). The PAmCherry gene was C-terminally fused to the *rpoC* gene, which encodes for the β' subunit of RNAP. This fusion gene replaces the endogenous copy in the chromosome, making it the sole source of β' subunit in the cell. Control experiments were performed to ensure that the fusion protein was not subject to proteolytic cleavage, as had been shown previously

(3), and that the cells grew otherwise normally as compared to wild-type cells, indicating the functionality of the RNAP fusion (4, 5).

The RNAP fusion strain was inoculated from a freshly streaked LB plate into 2 mL of minimal M9 media and grown overnight at room temperature, shaking at 250 rpm. After 16 hours of growth, cells were diluted 1:200 into fresh minimal M9 and were shaken at room temperature until they were in mid-log phase growth (OD_{600} of ∼0.4). Cells were harvested by taking 1 mL of the cells and spinning them down at 8 rcf for two minutes. Next, 900 μL of the supernatant was removed from the tube and cells were resuspended in the remaining 100 μL of media, to obtain an OD_{600} of ∼4. A small amount of these dense cells, approximately 0.3 to 0.5 μL, was pipetted onto a freshly-prepared 3 % agarose gel pad. Cells were immobilized onto the gel pad by letting the cells dry in air for two minutes. After drying, the gel pad was covered with a clean coverslip to assemble the Bioptechs imaging chamber (Bioptechs Inc.).

Once immobilized on the agarose gel pad, we stochastically activated RNAP-PAmCherry molecules using 0.1 mW of 405 nm light, which converts the PAmCherry molecule from a dark state to a red-emitting state, used 50 mW of 568 nm light to excite individual RNAP-PAmCherry molecules and tracked their cellular positions at a frame rate of approximately 150 Hz (5 ms exposure, 6.74 ms per frame). At this imaging speed, we were able to capture RNAP-PAmCherry molecules up to a diffusion coefficient of 3 $\mu m^2/s$ with accuracy in the cellular position of the molecule of approximately 30 nm. Cellular positions and lengths were determined through the software U-Track (6) and screened based on their intensities and position within the field of view. Trajectories were re-cut into multiple subtrajectories consisting of only consecutive frames of molecular localizations (gaps in localizations are due to the inherent blinking properties of all fluorescent proteins). These subtrajectories are used in the analyses detailed below.

2 Applying SPICER to 2d tracking data.

An example of a 2d SMT trajectory modified by SPICER is shown in Figure S1, with the confinement zone shown in red and the freely diffusing region in green. Intuitively, the operational principle of SPICER is still justified for 2d tracking data as displacements in the center of the cell (green) will have a higher probability to belong to true localizations outside the confined R-region. This is because a rod-shaped bacterial cell is isotropic along the short axis of the cell. By having a large number of trajectories sampling all

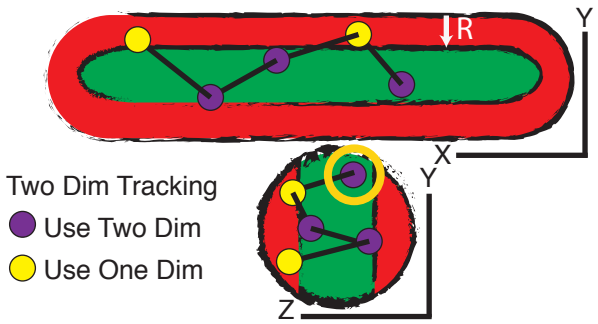

Figure S1: An example $2d$ SMT trajectory of a molecule in a bacterial cell. The purple circles are localizations inside the confinement-free region (green), and displacements calculated using these localizations as initial positions utilize their full $2d$ coordinates. Yellow circles are localizations inside the R-region and experience confinement (red). Displacements calculated using these localizations as initial positions only utilize coordinates along the x (long) axis of the cell. Both purple and yellow localizations are $2d$ projections of molecule positions in $3d$, and hence it is possible that a localization that appears to be outside the R-region is actually inside the R-region and experiences confinement (yellow hollow circle), but its full coordinates are used.

possible positions, localizations in the periphery and center of the cell will still have high probabilities to be correctly identified as inside or outside of the R-region. The use of SPICER on $2d$ tracking data is confirmed by applying SPICER to a variety of different systems of $2d$ tracking data; these results are illustrated in Figure S3, S5 and S6.

It is important to note that while applicable to $2d$ SMT, the use of SPICER on $2d$ tracking data is at a disadvantage when compared to $3d$ tracking, due to the lack of information along the third dimension. The uncertainty in the third dimension creates a chance that a small percentage of the displacements selected by SPICER as having no confinement will possess some confinement error, as indicated by the circled spot in Figure S1. Hence, the application of SPICER to $2d$ data results in a less significant improvement in the calculation of the different parameters when compared to the $3d$ tracking data.

3 Varying dimensions within the trajectory:

To illustrate that d can be varied throughout a trajectory, we assume that a molecule has a trajectory w of N displacements, and spends v displacements within the R region and k displacements outside of the R region, with $v+k = N$. In the simplest scenario where the molecule exists only in one state, the likelihood of the molecule having a D value given the trajectory is:

$$L(D|w,R) = \frac{1}{((4\pi D\tau)^{1/2})^v} e^{\sum_i^v -\frac{\Delta x_i^2}{4D\tau}} \times \frac{1}{((4\pi D\tau)^{2/2})^k} e^{\sum_j^k -\frac{\Delta x_j^2+\Delta y_j^2}{4D\tau}} \quad (2)$$

The log of the likelihood can be expressed as:

$$l(D|w,R) = -(k+\frac{v}{2})\log(D4\pi\tau) - \sum_i^v \frac{\Delta x_i^2}{4D\tau} - \sum_j^k \frac{\Delta x_j^2+\Delta y_j^2}{4D\tau} \quad (3)$$

Simplifying by substituting with v = N-k results in:

$$l(D|w,R) = -(\frac{k+N}{2})\log(D4\pi\tau) - \sum_i^N \frac{\Delta x_i^2}{4D\tau} - \sum_j^k \frac{\Delta y_j^2}{4D\tau} \quad (4)$$

Maximize the log of the likelihood L with respect to D by taking the derivative results in:

$$D = \frac{\sum_i^N \Delta x_i^2 + \sum_j^k \Delta y_j^2}{2\tau(N+k)} \quad (5)$$

Which can be further converted to the mean squared displacement of each dimension by

$$D = \frac{\langle \Delta x^2 \rangle + \langle \Delta y^2 \rangle \times k/N}{2\tau(1+k/N)} \quad (6)$$

Equation 6 holds true irrespective of the value of k, be k=0 or N. For $0 < k < N$, the diffusion coefficient D that best fits the system is the proportioned combination of the two mean squared displacements. Equation 5 further emphasizes that changing the value of d in a trajectory has no effect on the parameters obtained by maximizing the likelihood as long as the R-value, and hence the number of v or k displacements, is kept constant.

Equation 6 demonstrates that including a proportion of non-confined displacements along the short axis in the likelihood calculation will increase the calculated diffusion coefficient if there is confinement experienced by $\langle \Delta x^2 \rangle$. This is why SPICER is able to outperform even the $1d$ analysis. For example if there was no confinement $\langle \Delta x^2 \rangle$ and $\langle \Delta y^2 \rangle$ would be equal, but because the $1d$ analysis still experiences confinement in the cell poles, $\langle \Delta x^2 \rangle$ is smaller than expected. By including data along the short axis with no confinement, outside the R region ($\langle \Delta y^2 \rangle > \langle \Delta x^2 \rangle$), the calculated diffusion coefficient rises when the likelihood is maximized, see Eq. 6.

4 Calculating the Likelihood of Multiple State Trajectories:

In this section we describe the methodology created by Das *et al.* to calculate the likelihood of a single particle trajectory with multiple states (2). For a two state system there are four parameters, $\sigma = [D_1, D_2, P_{12}, P_{21}]$. The likelihood of having a particular single particle trajectory, ω=$(\Delta r_1, \Delta r_2, ...r_N)$, is

$$L(\sigma|\omega) \propto P(\omega|\sigma) = \sum_{All(S)} P(\omega|S,\sigma) \times P(S|\sigma) \tag{7}$$

where S is the state sequence of the particle throughout the trajectory, and $All(S)$ is the sum over all of the possible state sequences. The term $P(S|\sigma)$ is the probability of having a particular state sequence S given the two transition probabilities, creating a dependence upon the transition probabilities. The term $P(\omega|S,\sigma)$ is only dependent upon the diffusion coefficients with the particular diffusion coefficient defined by the state sequence S.

Because the summation is over all possible state sequences, we utilize the forward-backward algorithm to calculate the likelihood of a trajectory (2). The forward-backward algorithm determines the likelihood of a trajectory up to the displacement Δr_j, recursively, using the following equation

$$\alpha_j^i = P[\Delta r_1, \Delta r_2..\Delta r_j, s_j = i|\sigma] = [\alpha_{j-1}^1 * P_{1i} + \alpha_{j-1}^2 * P_{2i}] * P(\Delta r_j|s_j = i, \sigma) \tag{8}$$

with

$$P(\Delta r_j|s_j = i, \sigma) = \frac{e^{-\frac{\Delta r_j^2}{4D_i\tau}}}{(4\pi D_i\tau)^{d/2}} \tag{9}$$

where α_j^i is the forward variable, which gives the probability of observing the trajectory and being in state i, $s_j = i$ at displacement j. The initial forward variable is calculated from the overall probability of being in either state 1 or 2, see Das *et. al* for details.

Given that the total length of the trajectory is N, the probability of having the trajectory ω given the four parameters is

$$l(\sigma|\omega) \propto P(\omega|\sigma) = \alpha_N^{i=1} + \alpha_N^{i=2} \tag{10}$$

To account for all trajectories, we calculate the log of the likelihood for each of the trajectories and then maximize the sum of the log of the likelihoods with respect to the four parameters using the MCMC approach as described in the main text.

$$L(\sigma|\omega_k) = log[l(\sigma|\omega_k)] \tag{11}$$

$$L(\sigma|\omega_{All(k)}) = \sum_{k=1}^{M} log[l(\sigma|\omega_k)] \tag{12}$$

5 SI figures referenced in the main text

Figure S2: Parameter scan using MCMC approach
Figure S3: Determining optimal R-values for $2d$ tracking
Figure S4: Application of SPICER to a variety of different systems $3d$
Figure S5: Application of SPICER to a variety of different systems $2d$
Figure S6: Application of SPICER to systems with varying diffusion coefficients for $2d$ tracking data.
Table S1: Parameters of systems analyzed in Figures S4 and S5

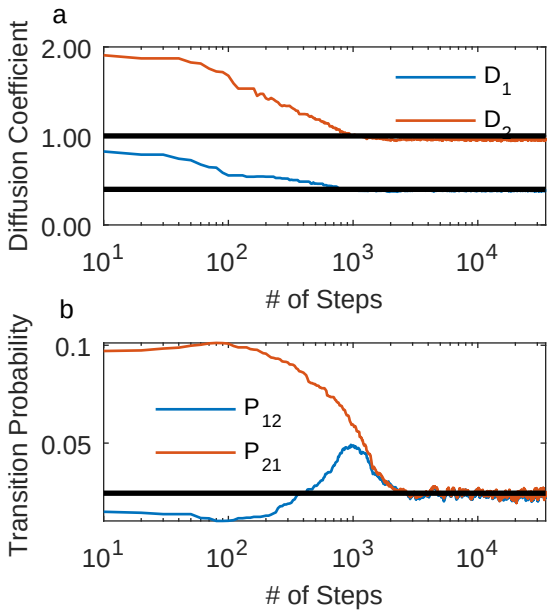

Figure S2: An example of a parameter scan using the MCMC approach. The black lines in the two graphs represent the true values, $D_1 = 1\mu m^2/s$, $D_2 = .4\mu m^2/s$, $P_{12} = P_{21} = .0244$ ($k = 5/sec$), of the two state simulation with 50,000 trajectories.

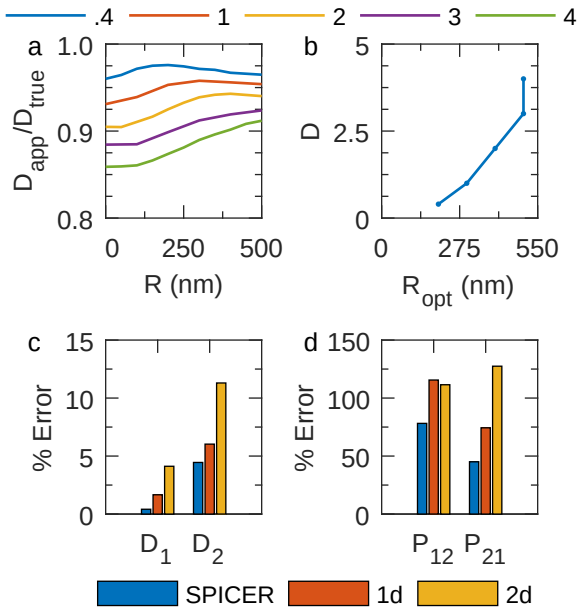

Figure S3: (a and b):Finding optimal R-values for $2d$ tracking systems (a) Approximation percentage (D_{app}/D_{true}) of five simulated systems at different R-values with D_{true} varying from 0.4 to $4\mu m^2/s$, tracking at an imaging speed of 200 f/s. (b) Optimal R-value lookup identified at different diffusion coefficients from (a). (c and d): Comparison of the performance of SPICER and conventional $1d$ and $2d$ analyses in identifying the diffusion coefficients (c) and transition probabilities (d) in a two-state system with $D_1 = 1\mu m^2/s$, $D_2 = .7\mu m^2/s$, and $P_{12} = P_{21} = .0244$. The percentage error is defined as $\frac{|X - X_{true}|}{X_{true}} \times 100$.

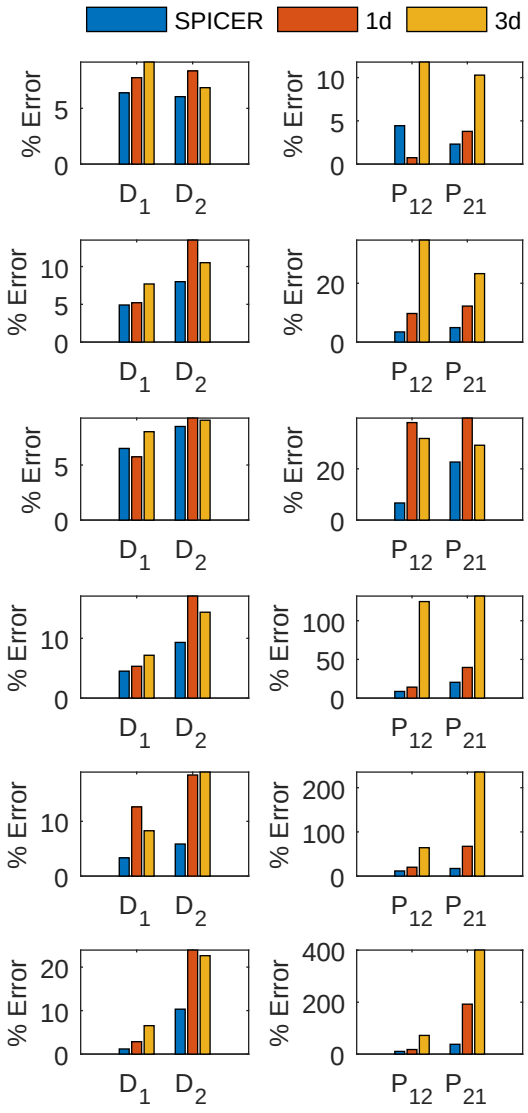

Figure S4: Percent errors in D_1, D_2 (left column) and P_{12}, P_{21} (right column) identified using SPICER, $1d$ and $3d$ analyses for different $3d$-tracking systems listed in Table S1. Each row in the figure corresponds to the same row in Table S1. In all the systems tested, SPICER outperforms the $1d$ and $3d$ analyses.

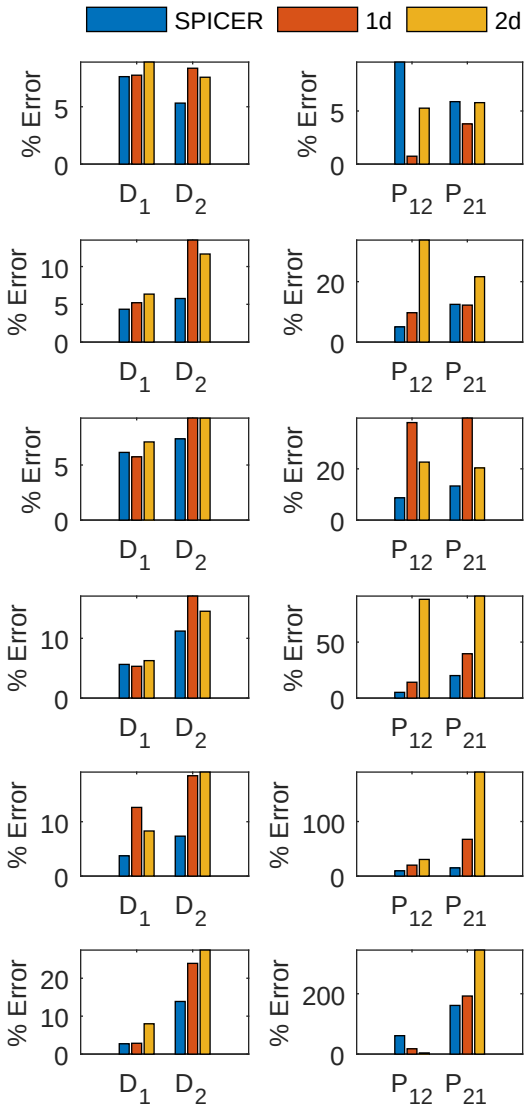

Figure S5: Percent errors in D_1, D_2 (left column) and P_{12}, P_{21} (right column) identified using SPICER, $1d$ and $2d$ analyses for different 2d-tracking systems listed in Table S1. Each row in the figure corresponds to the same row in Table S1. In all the systems tested, SPICER outperforms the $1d$ and $2d$ analyses.

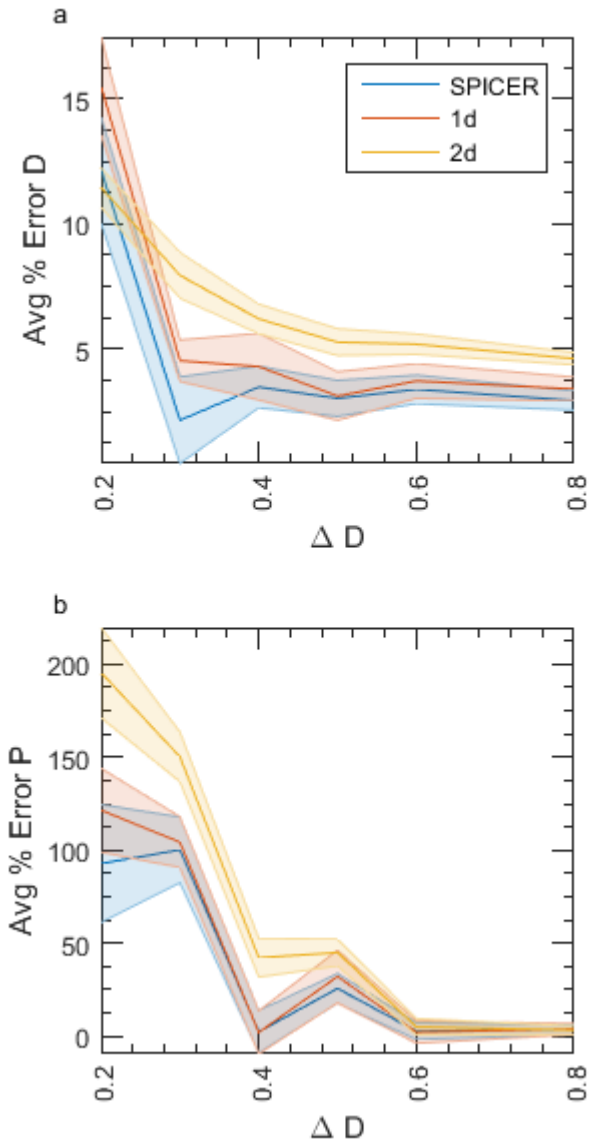

Figure S6: Comparison of averaged percent error in identifying diffusion coefficients (a) and transition probabilities (b) of systems with varying separations between the diffusion coefficients of the two states (ΔD) using SPICER, $1d$ or $2d$ analysis. The larger D is fixed at $1\mu m^2/s$ with the smaller D varying between 0.8 and 0.2 $\mu m^2/s$. The average percent error is calculated as $(\frac{|D_1-D_1^{true}|}{D_1^{true}}+\frac{|D_2-D_2^{true}|}{D_2^{true}})\times 50$ or $(\frac{|P_{12}-P_{12}^{true}|}{P_{12}^{true}}+\frac{|P_{21}-P_{21}^{true}|}{P_{21}^{true}})\times 50$. The shaded region indicates the uncertainty in the parameter and defined as the standard deviation of the parameter during the MCMC approach.

D_1	D_2	P_{12}	P_{21}	# of Traj
1	.4	.0476 (k=10/sec)	.0476	35000
1	.5	.0476 (k=10/sec)	.0476	35000
1	.6	.0696 (k=14/sec)	.0696	35000
1	.7	.0242 (k=5/sec)	.0387 (k=8/sec)	35000
1	.8	.0929 (k=20/sec)	.0464	35000
1	.9	.0714 (k=15/sec)	.0340	35000

Table 1: The parameters of the two state systems for the two SI figures S3 and S4.

References

[1] Persson, F., M. Lindén, C. Unoson, and J. Elf, 2013. Extracting intracellular diffusive states and transition rates from single-molecule tracking data. Nature Methods 10:265–269.

[2] Das, R., C. W. Cairo, and D. Coombs, 2009. A hidden Markov model for single particle tracks quantifies dynamic interactions between LFA-1 and the actin cytoskeleton. PLoS Computational Biology 5:e1000556.

[3] Endesfelder, U., K. Finan, S. J. Holden, P. R. Cook, A. N. Kapanidis, and M. Heilemann, 2013. Multiscale spatial organization of RNA polymerase in Escherichia coli. Biophysical journal 105:172–181.

[4] Bettridge, K., C. Bohrer, X. Weng, and X. Jie, 2016. Dynamics of RNA Polymerase in live E. coli cells. manuscript in preparation .

[5] Weng, X., C. Bohrer, A. Lagda, and X. Jie, 2016. Spatial organization of transcription revealed by functional superresolution imaging in E. coli. manuscript in preparation .

[6] Jaqaman, K., D. Loerke, M. Mettlen, H. Kuwata, S. Grinstein, S. L. Schmid, and G. Danuser, 2008. Robust single-particle tracking in live-cell time-lapse sequences. Nature Methods 5:695–702.