Biophysical Journal
Computational Tool

Biophysical Society

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

Christopher H. Bohrer,[1] Kelsey Bettridge,[1] and Jie Xiao[1,*]
[1]Department of Biophysics and Biophysical Chemistry, Johns Hopkins School of Medicine, Baltimore, Maryland

ABSTRACT Single-molecule tracking can extract quantitative kinetic information and identify possible state transitions of diffusing molecules (such as switching between binding and unbinding) in the in vivo environment of living cells. Confined diffusion, caused by the encompassing membrane boundary of the cell, results in increased uncertainties in identifying state-associated diffusion coefficients and transition probabilities. This problem is particularly acute in bacterial cells because of their small sizes. A standard approach to eliminating confinement errors in bacterial cells is to analyze molecule displacements only along the long axis of the cell, where molecules experience the least confinement, and hence turn three-dimensional tracking into a one-dimensional problem. However, this approach dramatically decreases the amount of data usable for statistical analysis and leads to increased uncertainties in identifying different states. Here, we present a simple algorithm, termed single-particle tracking improvement with confinement error reduction (SPICER), which significantly decreases confinement errors by selectively incorporating data not only from the long axis but also from the short axes of the cell. We validate SPICER using both reaction-diffusion simulations and experimental single-molecule tracking (SMT) data of RNA polymerase in live *Escherichia coli* cells. SPICER is easy to implement with existing SMT analysis routines and should find broad applications in SMT analysis.

INTRODUCTION

Single-molecule tracking (SMT) is a powerful technique for probing possible functional states of biomolecules in living cells (1–3). In a typical SMT experiment, a molecule's cellular positions are recorded by acquiring its fluorescent images consecutively at defined time intervals. From these images, an SMT trajectory, a time series of corresponding spatial coordinates of the molecule in reference to the cell, is extracted. From the statistical analysis of these SMT trajectories, different diffusive states of the molecule, each characterized by a different diffusion coefficient, D, can be obtained. These diffusive states and the associated population percentages can provide valuable information regarding possible functional states of the molecule. Recent SMT in bacterial cells have indeed shed light on the working mechanisms of transcription factors (4,5), RNA polymerase (6–8), DNA polymerase (9), ribosomes (10), cytoskeletal proteins (11,12), and more (13,14).

In addition to measuring a molecule's diffusion coefficients, SMT experiments offer another significant advantage, which is to obtain transition probabilities of molecules between different diffusive states. These transition probabilities provide crucial information regarding the kinetics of state switching, such as the binding and unbinding rates of a protein molecule to its target site and the lifetime of a particular functional state of the molecule (15,16). Such information is often difficult to obtain in live cells by other means.

Various algorithms based on the statistical analyses of a large number of SMT trajectories have been developed to obtain the transition probabilities and associated diffusive states from SMT experiments. Among them, the vbSPT algorithm developed by Persson et al. (16) has proven robust. vbSPT assumes a hidden Markov model (HMM) in which diffusing molecules make a memoryless jump in states defined by different diffusion coefficients, and uses a variational Bayesian approach to identify individual states and their associated kinetics (16).

Successful application of analysis methods like vbSPT requires the correct identification of different diffusive states, which are characterized by unique diffusion coefficients. However, in bacterial cells, the small cell size (1–2 μm) spatially confines a molecule's diffusion such that the measured apparent diffusion coefficient, D_{app}, of a molecule appears smaller than the actual value, leading to difficulties in identifying the correct diffusive state. A common practice for minimizing confinement effects is to use displacements measured only along the long axis of the cell, where

Submitted July 1, 2016, and accepted for publication January 9, 2017.
*Correspondence: xiao@jhmi.edu
Editor: Thomas Perkins.
http://dx.doi.org/10.1016/j.bpj.2017.01.004

molecules would experience the least confinement; we refer to this method as "1d" analysis throughout the work (7,16,17). However, using information only along one dimension rather than using all available dimensions leads to a less accurate determination of diffusion parameters. The reduced amount of data limits the available number of well-defined trajectories, which is particularly important for calculating transition probabilities (18,19).

Here, we present, to our knowledge, a new, simple algorithm, termed single-particle tracking improvement with confinement error reduction (SPICER). SPICER maintains the full length of an SMT trajectory and maximizes the amount of data used by calculating displacements in all dimensions available (2d or 3d) and only selectively switches to 1d (along the cell's long axis) when a molecule is likely to experience confinement. As such, the accuracy in determining both the diffusion states and transition probabilities is dramatically improved. We demonstrate the use of SPICER on both simulated and experimental SMT data of *Escherichia coli* RNA polymerase (RNAP). The simple implementation of the SPICER algorithm in SMT analysis should allow its wide application in probing in vivo dynamics of molecular events in small bacterial cells.

RESULTS

Operational principle of SPICER

To illustrate the operational principle of SPICER, we show in Fig. 1 a schematic 3d SMT trajectory of a diffusing molecule in a typical rod-shaped bacterial cell with two cross sections along the long, x (*top*), and short, y (*bottom*) axes of the cell. The parameter R 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*).

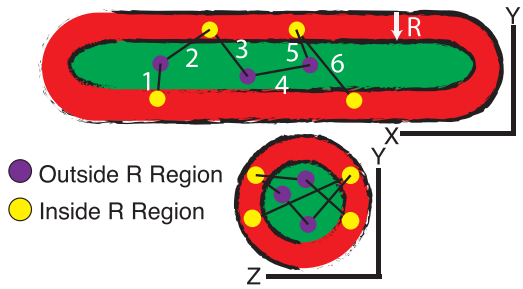

FIGURE 1 An example 3d SMT trajectory of a molecule in a rod-shaped bacterial cell. The purple solid circles are localizations of the molecule inside the confinement-free region (*green*). Displacements using these localizations as initial positions, i.e., displacements 2, 4, and 5, are calculated using their full 3d coordinates. Yellow solid circles are localizations of the molecule inside the R-region, where the molecule experiences confinement (*red*). Displacements using these localizations as initial positions, e.g., displacements 1, 3, and 6, are calculated using only 1d coordinates along the x (long) axis of the cell.

In previous studies, to avoid confinement, only displacements along the cell's long axis were used for HMM analysis. We refer to this method as "1d" analysis in this work (7,16,17). Consequently, in the 1d analysis, the available data in the trajectory ω shown in Fig. 1, represented by a series of single-step displacements, $\omega = (\Delta r_1^{3d}, \Delta r_2^{3d}, \Delta r_3^{3d}, \Delta r_4^{3d}, \Delta r_5^{3d}, r_6^{3d})$ and containing information in all three dimensions (3d), $(r_j^{3d})^2 = \Delta x_j^2 + \Delta y_j^2 + \Delta z_j^2$, is reduced to a third of the original amount of data, $(r_j^{1d})^2 = \Delta x_j^2$. SPICER increases the amount of data available and reduces the proportion of data that experiences confinement by analyzing a modified trajectory, $w' = (\Delta r_1^{1d}, \Delta r_2^{3d}, \Delta r_3^{1d}, \Delta r_4^{3d}, \Delta r_5^{3d}, \Delta r_6^{1d})$, where displacements $\Delta r_{1,3,6}^2$ are calculated in 1d along the long axis of the cell to avoid confinement in the R region, whereas $\Delta r_{2,4,5}^2$, are calculated in 3d, as the initial positions of these displacements are in the midcell and outside of the R region. As such, the full length of the trajectory is maintained, there is an increase of data being utilized with coordinates of all available dimensions, and there is a decrease in the proportion of data experiencing confinement. Note here that the same principle can be applied to 2d tracking experiments because of the symmetry of rod-shaped bacterial cells along the short axis. An example and associated discussion are provided in the Supporting Material.

Next, we demonstrate that switching dimensions within an SMT trajectory as described above does not modify the ability of SPICER to identify a set of most suitable parameters (diffusion coefficients, D, and transition probabilities, P) describing the trajectory using the maximum likelihood method (Supporting Material) (15).

The likelihood of having a diffusion coefficient, D, given a single displacement in d dimensions, $L(D \mid \Delta r_j)$, is proportional to the probability of having that displacement given the diffusion coefficient, $P(\Delta r_j \mid D)$, and is defined by the equation

$$L(D \mid \Delta r_j) \propto P(\Delta r_j \mid D) = \frac{e^{-\frac{\Delta r_j^2}{4D\tau}}}{(4\pi D\tau)^{d/2}}, \qquad (1)$$

where $\Delta r_j^2 = \Delta x_j^2$ for $d = 1$, $\Delta r_j^2 = \Delta x_j^2 + \Delta y_j^2$ for $d = 2$, and $\Delta r_j^2 = \Delta x_j^2 + \Delta y_j^2 + \Delta z_j^2$ for $d = 3$; D is the corresponding diffusion coefficient and τ is the time interval for each displacement. Equation 2 is the direct result of solving the diffusion equation with no barriers. If a molecule stays in one state, as defined by a single diffusion coefficient, D, the likelihood of having a particular trajectory specified by a series of experimentally measured displacements, w, will be

$$L(D \mid w) = \frac{e^{-\frac{\Delta r_1^2}{4D\tau}}}{(4\pi D\tau)^{d/2}} \times \frac{e^{-\frac{\Delta r_2^2}{4D\tau}}}{(4\pi D\tau)^{d/2}} \times \ldots \times \frac{e^{-\frac{\Delta r_j^2}{4D\tau}}}{(4\pi D\tau)^{d/2}}. \qquad (2)$$

Maximizing the likelihood, L, with respect to D results in the well-known relation $\langle r^2 \rangle = 2D\tau$ for $d = 1$, $\langle r^2 \rangle = 4D\tau$ for $d = 2$, and so on (15). In previous studies, the value d is set constant for all displacements in a trajectory. However, note here that the true diffusion coefficient, D, is independent of the d value used, that the probability of each displacement is independent of the other displacements at each time point, and that the likelihood, L, is an arbitrary multiplicative constant with no significance in its absolute value in isolation. Therefore, varying d values along a trajectory does not prevent maximizing the likelihood to find the best-fit parameter D. In the Supporting Material, we provide a further validation of this concept.

Selection of an optimal *R* value

Before one can analyze SMT data with SPICER, an optimal R-value for a given experimental system must be identified. The R-value defines the size of the confinement zone, within which the displacements of a trajectory are calculated using only $1d$ coordinates along the long axis. Displacements outside of the R region, toward the center of the cell, are computed using the full coordinates available in $2d$ or $3d$, depending on the experimental setup.

Intuitively, the size of the confinement zone, or the R-value, is primarily dependent on how fast the molecule diffuses. Molecules diffusing quickly require a large R-value to avoid confinement, whereas molecules diffusing slowly do not. Therefore, for a mixed population of molecules, the optimal R-value, R_{opt}, should be set for molecules that diffuse the fastest. Consequently, it follows that at a given imaging speed, one can construct a lookup table so that each estimated D_{max} value of a system corresponds to an optimal R-value.

To create the lookup table, we simulated five sets of $3d$ SMT experiments in a rod-shaped cell with radius $r = 500$ nm and length $l = 2$ μm; each set contains 10,000 single-state SMT trajectories with a fixed D_{true} ranging from 0.4 to 4 $\mu m^2/s$, tracked with an imaging speed of 200 f/s. For each data set, we varied the R-value systematically from 50 to 500 nm at 50-nm intervals and used SPICER to identify the corresponding apparent D value (D_{app}) at each R-value. We then plotted the approximation percentage (D_{app}/D_{true}) of the data set at different R values (Fig. 2 *A*). The optimal R-value was then identified as the one at which the approximation percentage reaches a maximum. The reasoning is that if an R-value is correctly selected, D_{app} should contain the least confinement error in SPICER, hence approaching maximally the D_{true} value.

As shown in Fig. 2 *A*, data sets with small diffusion coefficients reach their maximal apparent diffusion coefficients, D_{app}, at small R values, consistent with the notion that slowly diffusing molecules experience less confinement and hence the R-region would be small. However, for the data set that has a $D = 4$ $\mu m^2/s$, $R_{opt} = 450$ nm, indicating

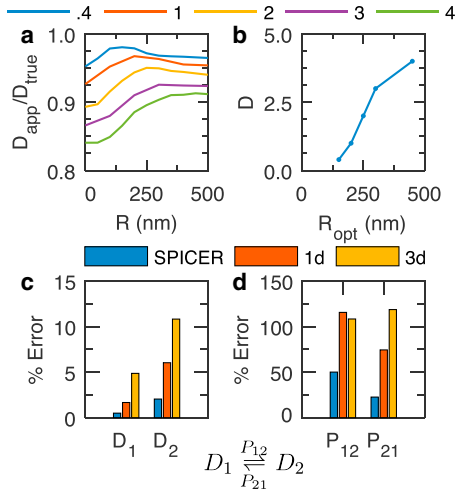

FIGURE 2 (*a* and *b*) Construction of a lookup table for finding the optimal R-value in a $3d$ tracking system. (*a*) Approximation percentage (D_{app}/D_{true}) of five simulated systems at different R-values, with D_{true} varying from 0.4 to 4 $\mu m^2/s$. (*b*) Optimal R-values at different diffusion coefficients are identified from (*a*) as the R-value at which the maximal D_{app}/D_{true} is reached. (*c* and *d*) Comparison of the performance of SPICER and conventional $1d$ and $3d$ analyses in identifying the diffusion coefficients (*c*) and transition probabilities (*d*) in a two-state system with $D_1 = 1$ $\mu m^2/s$, $D_2 = 0.7$ $\mu m^2/s$, and $P_{12} = P_{21} = 0.0244$. The percentage error is defined as $((|X - X_{true}|)/X_{true}) \times 100$.

that for molecules diffusing faster than 4 $\mu m^2/s$, the small size of a bacterial cell itself confines diffusion, regardless of how far away from the membrane the molecule is. In Fig. 2 *B*, we plotted the optimal R-value for each simulated D_{true} value. It can be seen that R_{opt} monotonically increases with D. Note here that although this lookup table is coarse-grained, with the R-value changing in 50-nm increments and the D-value changing in ~1-$\mu m^2/s$ increments, finer grains on the order of 5 nm and 0.1 $\mu m^2/s$ are not necessary. The typical spatial resolution in an SMT experiment in live bacterial cells is in the range 30–50 nm, and a change of 0.1 $\mu m^2/s$ in D does not lead to a significant change in the corresponding R-value within the 50-nm increment. Therefore, to use this lookup table, one can first estimate the largest diffusion coefficient of a given system using the $1d$ analysis, which approximates the true D value by eliminating confinement error along the cell long axis, and then use Fig. 2 *B* to estimate the R_{opt} value. A similar simulation and lookup table using $2d$ SMT data are shown in Fig. S3. A particular note here is that the lookup table is also related to the imaging speed, so that a fast-diffusing molecule imaged at a slow speed (long time intervals between subsequent acquisitions) will naturally require a larger R-value to accommodate the longer distance it travels during the time. Therefore, it is important to construct the lookup table based on the actual imaging condition, as we described above.

Next, we verified whether the utilization of an optimal R-value in SPICER indeed improves SMT analysis when molecules exist in two different diffusive states. We

simulated 25,000 trajectories of a two-state system. The two diffusion coefficients are $D_1 = 1\ \mu m^2/s$ and $D_2 = 0.7\ \mu m^2/s$, and the transition probabilities $P_{12} = P_{21} = 0.0244$, which corresponds to a transition rate of ~5 s^{-1}. We then analyzed this data set using 1*d* (only displacements along the cell long axis), 3*d* (using all displacements in three dimensions), or SPICER, in which the R_{opt} was chosen to be 200 nm (the larger D_1 at 1 $\mu m^2/s$) according to Fig. 2 *B*.

In Fig. 2, *C* and *D*, we plotted the percent error in each of the four parameters analyzed using the three methods. Clearly, 3*d* analysis led to the highest amount of error for all the four parameters, consistent with the presence of significant confinement errors when displacements in all three dimensions were used under this condition. The 1*d* analysis showed improvement compared to the 3*d* analysis, especially in the identification of *D*, but it was significantly outperformed by SPICER, in which the percentage errors in all four parameters were the smallest. Note that localizations in the *R*-region of cell poles are still confined in SPICER, even though only their displacements along the cell's long axis are used, just as in 1*d* analysis. Nevertheless, SPICER outperforms the 1*d* analysis, because in SPICER, the full coordinates of localizations outside of the *R*-region are given more weight than their corresponding 1*d* coordinates in the search for optimal parameters. We further verified that the same trend holds for a variety of systems with different diffusive parameters (Fig. S4 Table S1). These results demonstrate that by increasing the proportion of data containing full coordinates, SPICER, with an optimal *R*-value identified from the lookup table (Fig. 2 *B*), indeed improves the accuracy in determining both diffusion coefficients and transition probabilities of a diffusive system.

SPICER improves accuracy in identifying states with close diffusion coefficients

One important criterion used by the HMM to identify different diffusive states is the difference between diffusion coefficients associated with each state. If the diffusion coefficients of the two states are close to each other, the considerable overlap of the displacement distributions will lead to difficulties in determining the associated state of a displacement, and consequently to large errors in identifying corresponding transition probabilities. SPICER should be especially useful in improving data analysis under this circumstance, as it can effectively eliminate confinement error without significantly reducing the available data.

To compare the performance of SPICER with traditional 1*d* and 3*d* analyses under these scenarios, we simulated eight different systems with 50,000 trajectories each, with P_{12} and P_{21} set to 0.0224 ($k = 5\ s^{-1}$), D_1 to 1 $\mu m^2/s$, and D_2 varied between 0.8 and 0.2 $\mu m^2/s$. We analyzed these systems as described in the Supporting Material and plotted the average percentage errors in *D* and *P* for the three methods (Fig. 3).

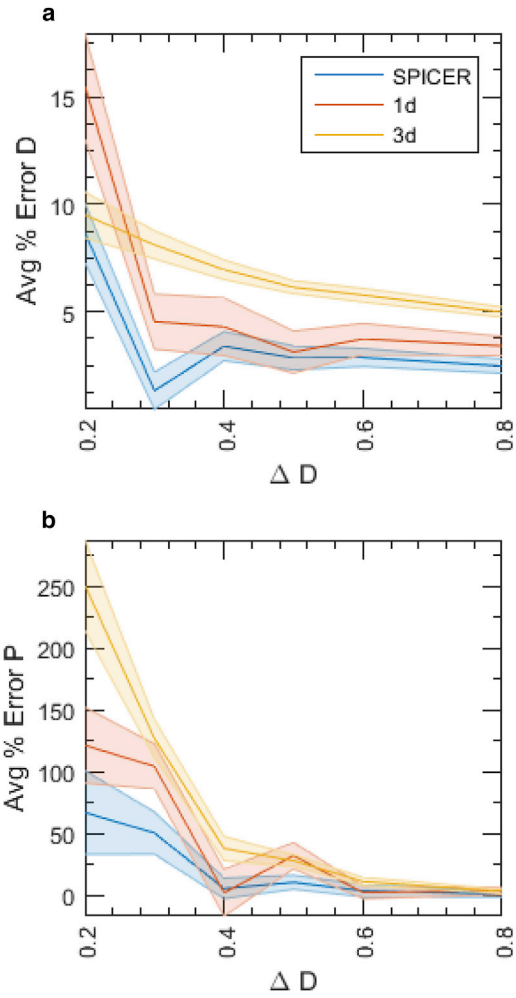

FIGURE 3 Comparison of the averaged percent errors in identifying diffusion coefficients (*a*) and transition probabilities (*b*) of systems with varying separations (ΔD) between the diffusion coefficients of the two states using SPICER, 1*d*, or 3*d* analysis. The larger *D* is fixed at 1 $\mu m^2/s$, with the smaller *D* varying between 0.8 and 0.2 $\mu m^2/s$. The average percent error is calculated as $((|D_1 - D_1^{true}|/D_1^{true}) + (|D_2 - D_2^{true}|/D_2^{true})) \times 50$ or $((|P_{12} - P_{12}^{true}|/P_{12}^{true}) + (|P_{21} - P_{21}^{true}|/P_{21}^{true})) \times 50$. The shaded region indicates the uncertainty in the parameter and is defined as the standard deviation of the parameter during the Markov chain Monte Carlo approach.

Consistent with what we expected, when ΔD decreases, the percentage errors in *D* and *P* increase for all three methods, but SPICER consistently outperforms the 1*d* and 3*d* analyses, in particular with smaller ΔD values. Only at larger values, $\Delta D > 0.6\ \mu m^2/s$, is the improvement less dramatic. These results thus demonstrate SPICER's unique advantage in systems where the diffusion coefficients of two states are closely spaced relative to each other.

SPICER requires fewer trajectories compared to 1*d* or 3*d* analysis to achieve the same level of error reduction

SMT analysis usually requires a large number of trajectories (on the order of 10^4 if the average length of trajectories is

short (16)), so that diffusion coefficients and state transitions can be determined with statistical significance. However, experimentally, it is time-consuming to collect tens of thousands of SMT trajectories. To investigate whether SPICER helps in lowering this requirement, we used the same two-state system analyzed in Fig. 2 and varied the number of trajectories used in the analysis. In Fig. 4, we show that for all three methods (1*d*, 3*d*, and SPICER) the averaged percent error in *D* plateaus when the number of trajectories is >5000; the averaged percent error in *P* plateaus when the number is greater than 10,000, as accurate determination of *P* requires a higher number of trajectories. However, even at a low number of trajectories (~3000), averaged percent errors of *D* and *P* in SPICER are substantially lower than those in 1*d* and 3*d* analyses, approaching the level that would be achieved by 10,000 trajectories with the 1*d* analyses. Note here that the decreases in the total error are mainly brought about by the minimization of confinement error in SPICER, which compensates for errors caused by an insufficient number of trajectories, as that occurs in 1*d* and 3*d* analyses. These results demonstrate that increasing the number of trajectories used will not improve the error in the calculated parameters when confinement error is present in the commonly used 1*d* and 3*d* analyses. The application of SPICER raises the proportion of data without confinement and leads to the least amount of error in determining the diffusion coefficients and transition probabilities in these systems.

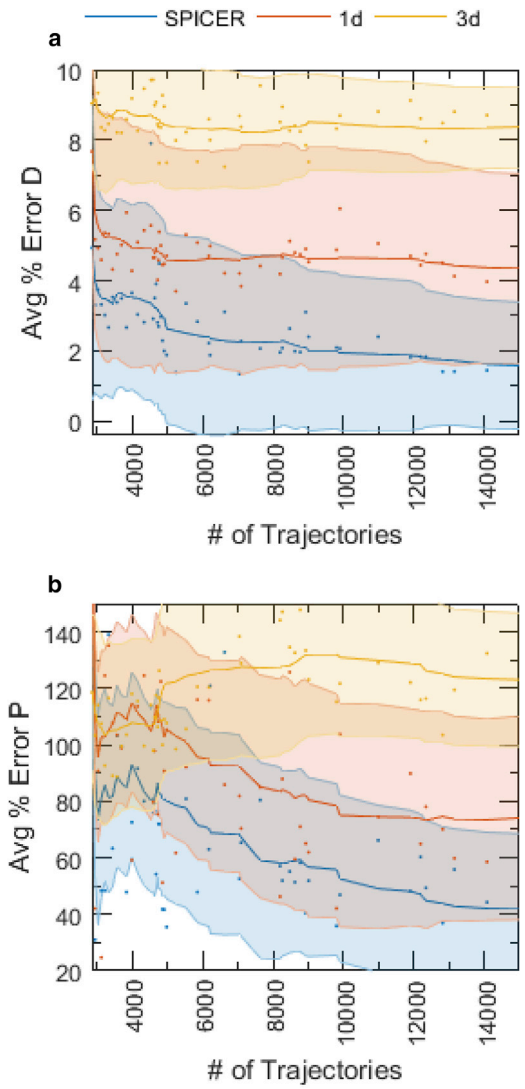

FIGURE 4 Comparison of the averaged percent errors in identifying diffusion coefficients (*a*) and transition probabilities (*b*) of the two-state system with a varying number of trajectories, shown in Fig. 2, *c* and *d*, using SPICER, 1*d*, and 3*d* analyses. The averaged percent error and shaded region are calculated the same way as for Fig. 3. The solid lines are the 10-point moving averages of raw data (*scattered dots*), and the shaded areas are the moving averages of the standard deviations of the parameters during the Markov chain Monte Carlo approach.

Validating SPICER using experimental RNAP tracking data

To further validate SPICER with experimentally obtained data, we performed 2*d* SMT on RNA polymerase (RNAP) in live *E. coli* cells. RNAP is primarily found within the nucleoid; because of its frequent interactions with chromosomal DNA, it has relatively small diffusion coefficients (6). Thus, RNAP would experience less confinement from the membrane than would other freely diffusing protein molecules in cells, and it can serve as a control system with negligible confinement to validate the SPICER algorithm.

We used a functional RNAP-PAmCherry fusion (gift from Dr. D. J. Jin of the National Cancer Institute) that is integrated into the *E. coli* chromosome, replacing the endogenous *rpoC* gene, which encodes the β' subunit of RNAP (Supporting Material). Under our imaging conditions, we collected a total of ~25,000 trajectories, with the average trajectory length at ~3, in RNAP-PAmCherry-expressing cells grown in minimal M9 medium with a 5-ms exposure time. We first used conventional 1*d* and 2*d* analyses to determine that under this condition, the best model describing the diffusive behaviors of RNAP is a two-state model. The two *D* values from 1*d* and 2*d* analyses are similar to each other ($D_1 = 0.38\ \mu m^2/s$ and $D_2 = 0.1\ \mu m^2/s$; Fig. 5 *A*) and are consistent with values from previous SMT studies of RNAP (6). However, transition probabilities obtained from the 1*d* analysis are significantly lower than those obtained from the 2*d* analysis (Fig. 5 *B*). The lower transition probabilities of the 1*d* analysis are most likely due to short trajectory lengths (with around three displacements) combined with the reduced amount of data in the 1*d* analysis, which makes it difficult to observe rare transitions between states. Using simulations, we further verified that indeed at this slow-diffusion condition, 2*d* analysis describes the system more accurately than 1*d* analysis (Fig. 5, *C* and *D*).

Next, we applied SPICER using an *R*-value of 200 nm, identified as optimal using the procedure described in the

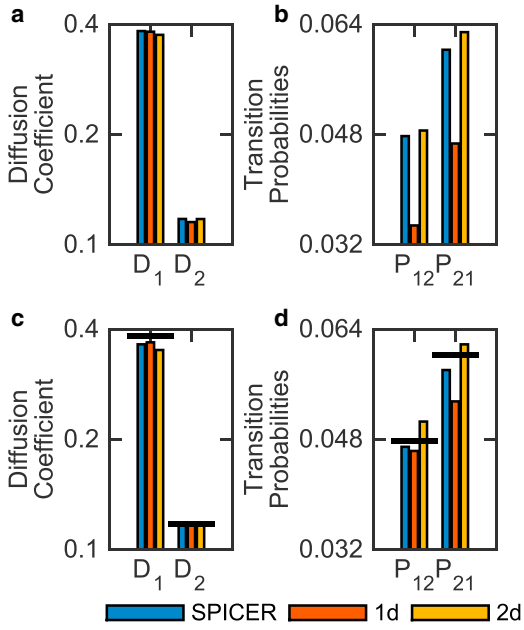

FIGURE 5 Validation of SPICER using experimentally acquired 2*d* SMT data of RNAP in live *E. coli* cells. (*a* and *b*) Comparison of the identified D_1 and D_2 (*a*) and P_{12} and P_{21} (*b*) values using SPICER, 1*d*, and 2*d* analyses. (*c* and *d*) Simulation of a similar system showing the same trend, that 2*d* and SPICER analyses are significantly more accurate than the 1*d* analysis, with SPICER reflecting the true values most closely. The true values for the simulation are shown as horizontal black lines.

previous section (Fig. S3), and obtained a new set of D_1, D_2, P_{12}, and P_{21}. As shown in Fig. 5 *A*, diffusion coefficients obtained using the three methods are similar to each other, suggesting that at this slow diffusing rate the confinement error is low, and that all the methods are capable of identifying *D* sufficiently well with the acquired number of trajectories. However, transition probabilities from SPICER and the 2*d* analysis are similar to each other and are both significantly higher than those obtained from the 1*d* analysis (Fig. 5 *B*). These results further demonstrate that SPICER can be used to analyze experimental SMT data with high accuracy.

CONCLUSIONS

In this work, we present a simple algorithm, SPICER, to reduce the confinement error in SMT analysis in small bacterial cells. SPICER calculates displacements in all dimensions available (2*d* or 3*d*) and only selectively switches to 1*d* (along the cell's long axis) when a molecule is within a pre-defined *R*-region where it likely experiences confinement. We provided lookup tables and experimental guidelines for finding an optimal *R*-value. The complete package of SPICER is available for download at https://github.com/XiaoLabJHU/SPICER.git. Using simulations, we compared SPICER with commonly used SMT analyses and show that SPICER consistently improves the accuracy in determining diffusion coefficients and state-transition probabilities in SMT analyses. Even when compared to the 1*d* analysis, the traditional method used to relieve confinement in multistate systems, the confinement in the poles of the cells allows SPICER to outperform the 1*d* analysis. This improvement is achieved by increasing the overall proportion of molecules experiencing free diffusion during the maximization of the likelihood. Furthermore, SPICER performs significantly better than previous methods when the separation between diffusion coefficients of two different states is small, and when the acquired number of SMT trajectories is low (< 3000). We further validated SPICER using experimentally obtained SMT trajectories of RNAP in live *E. coli* cells. SPICER should be particularly useful for comparing SMT results in bacterial cell size mutants, as the influence of confinement in cells of different sizes can be easily accounted for with SPICER. Furthermore, the central concept of SPICER can be generalized to other cell geometries as long as localizations in the *R*-region can be used along a particular dimension in which the molecule experiences the least confinement.

SUPPORTING MATERIAL

Supporting Materials and Methods, six figures, and one table are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(17)30043-7.

AUTHOR CONTRIBUTIONS

C.H.B. designed research, performed research, developed ideas, analyzed data, and wrote the article. K.B. performed experimental research and analyzed data. J.X. designed research and wrote the article.

REFERENCES

1. Sauer, M. 2013. Localization microscopy coming of age: from concepts to biological impact. *J. Cell Sci.* 126:3505–3513.
2. Gahlmann, A., and W. E. Moerner. 2014. Exploring bacterial cell biology with single-molecule tracking and super-resolution imaging. *Nat. Rev. Microbiol.* 12:9–22.
3. Vrljic, M., S. Y. Nishimura, and W. E. Moerner. 2007. Single-molecule tracking. *Methods Mol. Biol.* 398:193–219.
4. Izeddin, I., V. Récamier, …, X. Darzacq. 2014. Single-molecule tracking in live cells reveals distinct target-search strategies of transcription factors in the nucleus. *eLife.* 3:e02230.
5. Elf, J., G.-W. Li, and X. S. Xie. 2007. Probing transcription factor dynamics at the single-molecule level in a living cell. *Science.* 316:1191–1194.
6. Stracy, M., C. Lesterlin, …, A. N. Kapanidis. 2015. Live-cell superresolution microscopy reveals the organization of RNA polymerase in the bacterial nucleoid. *Proc. Natl. Acad. Sci. USA.* 112:E4390–E4399.
7. Bakshi, S., R. M. Dalrymple, …, J. C. Weisshaar. 2013. Partitioning of RNA polymerase activity in live *Escherichia coli* from analysis of single-molecule diffusive trajectories. *Biophys. J.* 105:2676–2686.
8. Bakshi, S., A. Siryaporn, …, J. C. Weisshaar. 2012. Superresolution imaging of ribosomes and RNA polymerase in live *Escherichia coli* cells. *Mol. Microbiol.* 85:21–38.

9. Uphoff, S., R. Reyes-Lamothe, …, A. N. Kapanidis. 2013. Single-molecule DNA repair in live bacteria. *Proc. Natl. Acad. Sci. USA.* 110:8063–8068.

10. Sanamrad, A., F. Persson, …, J. Elf. 2014. Single-particle tracking reveals that free ribosomal subunits are not excluded from the *Escherichia coli* nucleoid. *Proc. Natl. Acad. Sci. USA.* 111:11413–11418.

11. Niu, L., and J. Yu. 2008. Investigating intracellular dynamics of FtsZ cytoskeleton with photoactivation single-molecule tracking. *Biophys. J.* 95:2009–2016.

12. Kim, S. Y., Z. Gitai, …, W. E. Moerner. 2006. Single molecules of the bacterial actin MreB undergo directed treadmilling motion in *Caulobacter crescentus*. *Proc. Natl. Acad. Sci. USA.* 103:10929–10934.

13. Plochowietz, A., I. Farrell, …, A. N. Kapanidis. 2016. In vivo single-RNA tracking shows that most tRNA diffuses freely in live bacteria. *Nucleic Acids Res.* Published online September 12, 2016. http://dx.doi.org/10.1093/nar/gkw787.

14. Liao, Y., J. W. Schroeder, …, J. S. Biteen. 2015. Single-molecule motions and interactions in live cells reveal target search dynamics in mismatch repair. *Proc. Natl. Acad. Sci. USA.* 112:E6898–E6906.

15. 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 Comput. Biol.* 5:e1000556.

16. Persson, F., M. Lindén, …, J. Elf. 2013. Extracting intracellular diffusive states and transition rates from single-molecule tracking data. *Nat. Methods.* 10:265–269.

17. Bakshi, S., B. P. Bratton, and J. C. Weisshaar. 2011. Subdiffraction-limit study of Kaede diffusion and spatial distribution in live *Escherichia coli*. *Biophys. J.* 101:2535–2544.

18. Chung, I., R. Akita, …, I. Mellman. 2010. Spatial control of EGF receptor activation by reversible dimerization on living cells. *Nature.* 464:783–787.

19. Beausang, J. F., C. Zurla, …, P. C. Nelson. 2007. DNA looping kinetics analyzed using diffusive hidden Markov model. *Biophys. J.* 92:L64–L66.

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

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

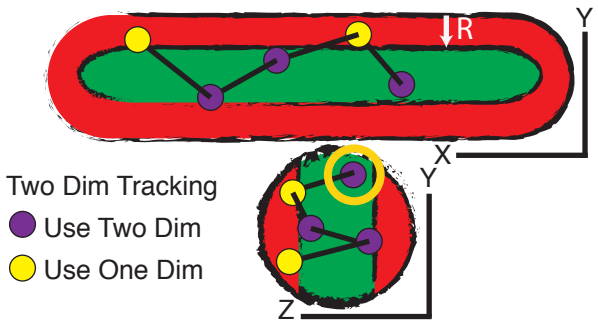

Figure S1: An example $2d$ SMT trajectory of a molecule in a bacterial cell. The purple circles are localizations inside the confinement-free region (green), and displacements calculated using these localizations as initial positions utilize their full $2d$ coordinates. Yellow circles are localizations 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

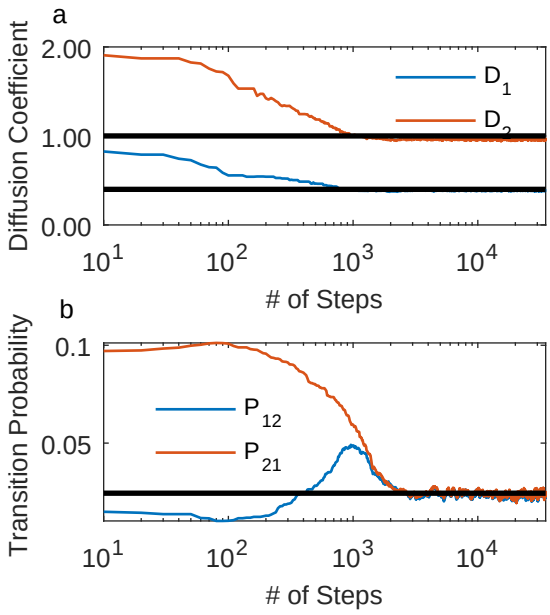

Figure S2: An example of a parameter scan using the MCMC approach. The black lines in the two graphs represent the true values, $D_1 = 1\mu m^2/s$, $D_2 = .4\mu m^2/s$, $P_{12} = P_{21} = .0244$ ($k = 5/sec$), of the two state simulation with 50,000 trajectories.

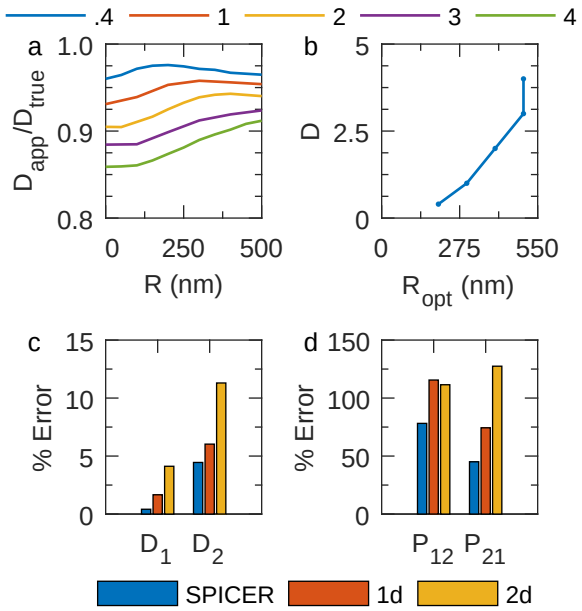

Figure S3: (a and b):Finding optimal R-values for $2d$ tracking systems (a) Approximation percentage (D_{app}/D_{true}) of five simulated systems at different R-values with D_{true} varying from 0.4 to $4\mu m^2/s$, tracking at an imaging speed of 200 f/s. (b) Optimal R-value lookup identified at different 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 $\frac{|X - X_{true}|}{X_{true}} \times 100$.

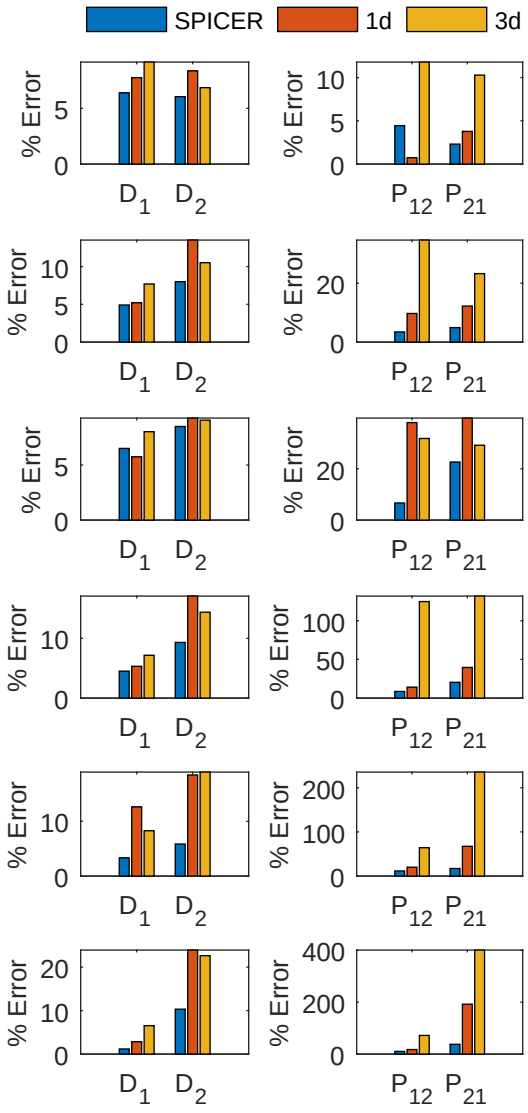

Figure S4: Percent errors in D_1, D_2 (left column) and P_{12}, P_{21} (right column) identified using SPICER, $1d$ and $3d$ analyses for different $3d$-tracking systems listed in Table S1. Each row in the figure corresponds to the same row in Table S1. In all the systems tested, SPICER outperforms the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

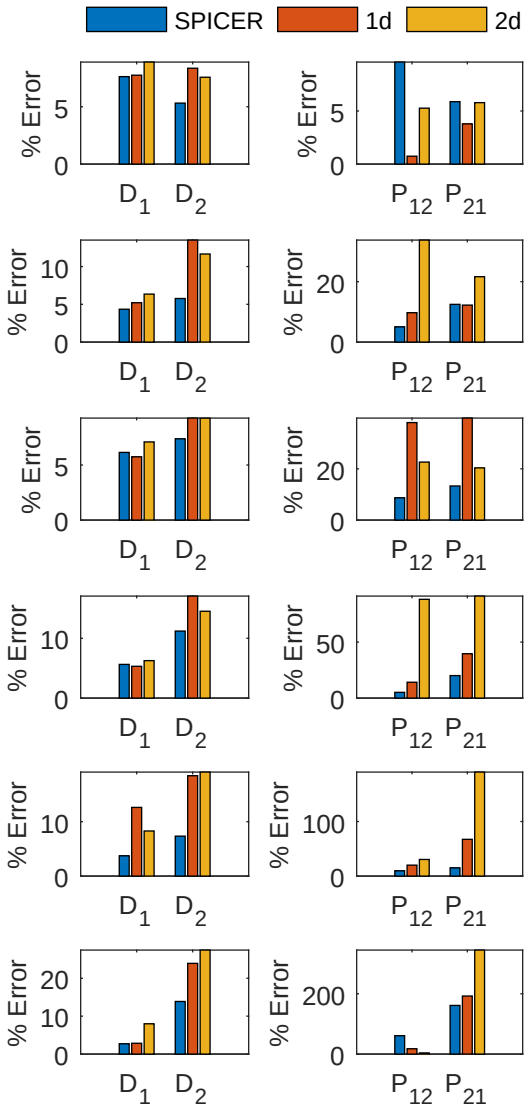

Figure S5: Percent errors in D_1, D_2 (left column) and P_{12}, P_{21} (right column) identified using SPICER, $1d$ and $2d$ analyses for different 2d-tracking systems listed in Table S1. Each row in the figure corresponds to the same row in Table S1. In all the systems tested, SPICER outperforms the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

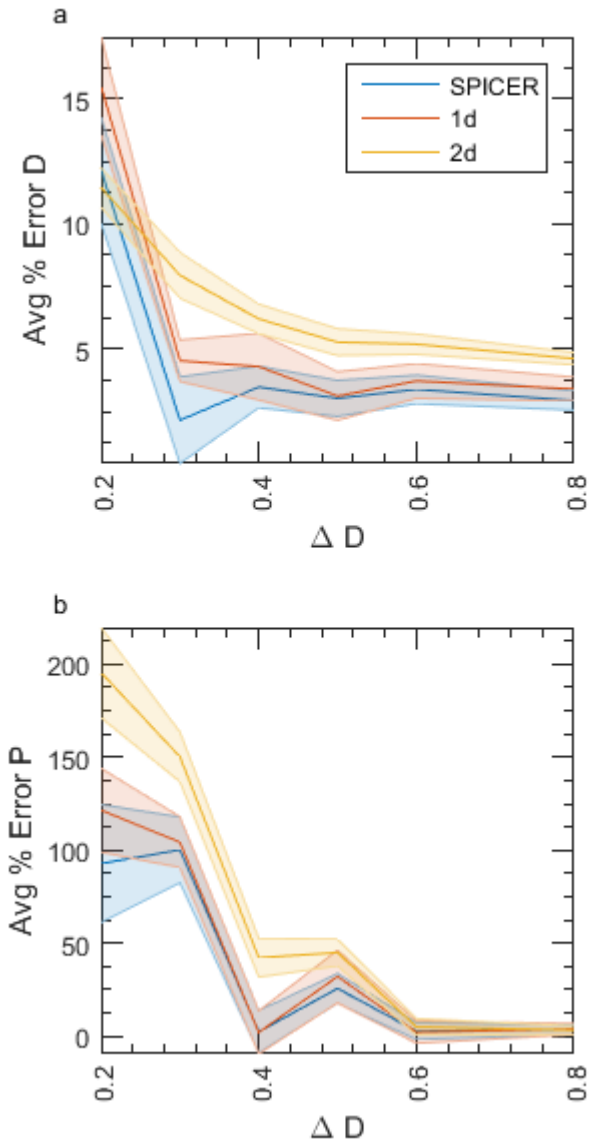

Figure S6: Comparison of averaged percent error in identifying diffusion coefficients (a) and transition probabilities (b) of systems with varying separations between the diffusion coefficients of the two states (ΔD) using SPICER, $1d$ or $2d$ analysis. The larger D is fixed at $1\mu m^2/s$ with the 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.