Supplemental Material for: Analyzing Single Molecule Time Series Using Nonparametric Bayesian Inference

Keegan E. Hines[1], John R. Bankston[2], and Richard W. Aldrich[1]

[1]Center for Learning and Memory and Department of Neuroscience ,The University of Texas at Austin, Austin,TX,78712

[2]Department of Physiology and Biophysics, University of Washington School of Medicine, Seattle, WA, 98195

1 Idealization of Single Channel Records

We have gathered single BK channel recordings at multiple holding voltages and calcium concentrations (see Methods). Figure S1 shows a subset of the collected data, filtered at 10 kHz, at the indicated holding voltage and calcium concentration. In order to idealize single channel recordings, we treat a single channel time series as a two-state hidden Markov model. Here, the open and closed states each correspond to different levels of current obscured by noise, each with different variability. Notice that a threshold method would yield very similar results to any model with a symmetric noise distribution, but makes the assumption that the current variance is the same for both open and closed states. We prefer not to make this assumption and so model closed and open states corresponding to Normal distributions each with distinct mean and variance. Using the Gibbs sampling approach described in the Theory section, we utilize a latent indicator variable $s_1, ..., s_N$ to denote the hidden state from which each data point was likely to have been drawn. Thus, after Gibbs sampling, the indicator variables $s_1, ..., s_N$ yield the idealized trajectory through the hidden states. This Bayesian approach to idealization of ion channel records has been used previously and was thoroughly compared to previous methods (1, 2), so we omit such a discussion here. Figure S2 shows an example of this method. The data points are overlaid with colors corresponding to which conductance state (closed or open) each point was likely drawn from. With an idealized trace, we simply count how many consecutive samples are spent in a state before transitioning to the other state: this is a dwell time in one of the states. Decomposing the whole recording in this way yields a distribution of dwell-time events in the open state and in the closed state.

0006-3495/08/09/2624/12 $2.00

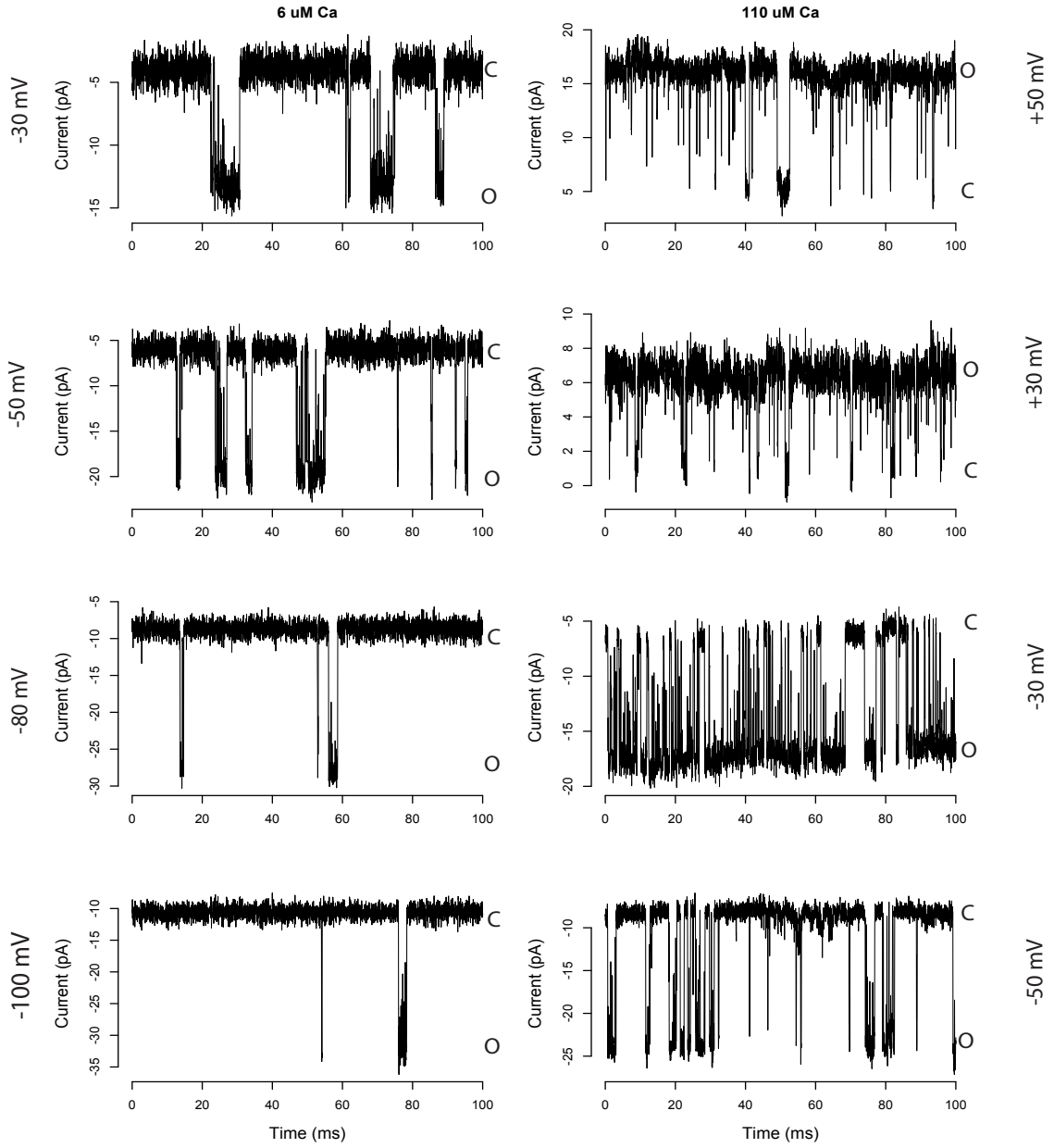

Figure S1: Example data from a single BK channel at multiple holding voltages and calcium concentrations, as indicated.

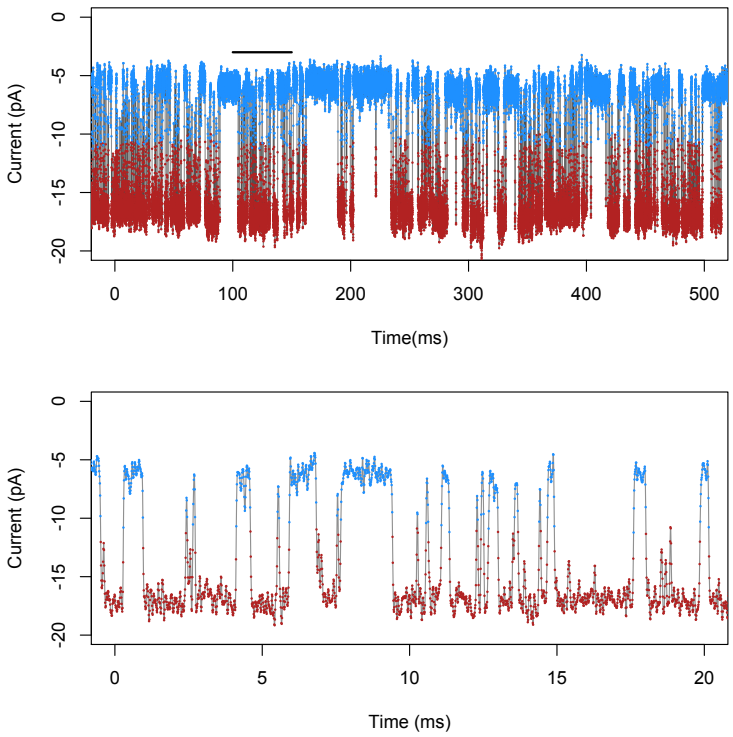

Figure S2: Using a two-state hidden Markov model to idealize single channel recordings. The time series is assumed to be drawn from a two-state Markov process where each state has a distinct emission distribution characterized by a Normal distribution with different means and variances. The model is fit using Gibbs sampling (see Theory) and the idealized trace (the hidden states) is shown as colors. Segments of the time series are shown at two different time scales.

2 Applications of Infinite Exponential Mixture Model

As is shown in the main text, the infinite exponential mixture (iEMM) model can be used to analyze multi-component mixture distributions without knowing beforehand the number of components. This is visualized in Figures 1, 3, and 7 of the main text for simulated data as well as dwell-time data recorded from a single BK channel. For the ion channel data in particular, Figure 3 of the main text shows that we can use the iEMM to analyze data from multiple holding voltages and learn how many states are visited by the channel. Figure 3 visualizes the results of this analysis: each data point is given a color according to which component it was likely drawn from and the densities of each component are shown as well as the aggregate density which is overlaid with the empirical histogram. However, this visualization does not convey our confidence in the number of inferred mixture components and we are left unable to make a strong statement regarding model selection. Figure S3 shows, for each of the same datasets, the approximated posterior distribution over the number of mixture components. We see that, in each case, the number of components is inferred with high confidence as the posterior distribution is sharply peaked at its modal value. The trace from +30mV (bottom right) yields the most uncertainty, with the posterior peaked at 4 components but with non-negligible probability mass at 5 components.

Further tests of this method come from analyzing simulated datasets. In the Discussion section of the main text, we compared our method with that of Landowne et al., and in particular, we used our method on parameter sets which were previously determined to be quite challenging (Figure 7 of main text). For clarity, we show in Figure S4 a more thorough comparison of our estimates with those from Landowne et al. For each parameter set (Boliden3 and Boliden4), we estimate the number of components, the time constant of each component and the weight parameter of each component. To facilitate comparison, Figure S4 shows the posterior distribution of each time constant parameter (shown as histograms) as well as the true parameter value (blue) and the parameter estimate reported in Landowne et al. (red). The parameter values estimated by Landowne et al. are all quite close to the true values. We cannot directly compare confidence intervals between the methods since Landowne et al. used 1000-fold larger sample sizes than we have and this would have a strong effect on parameter confidence.

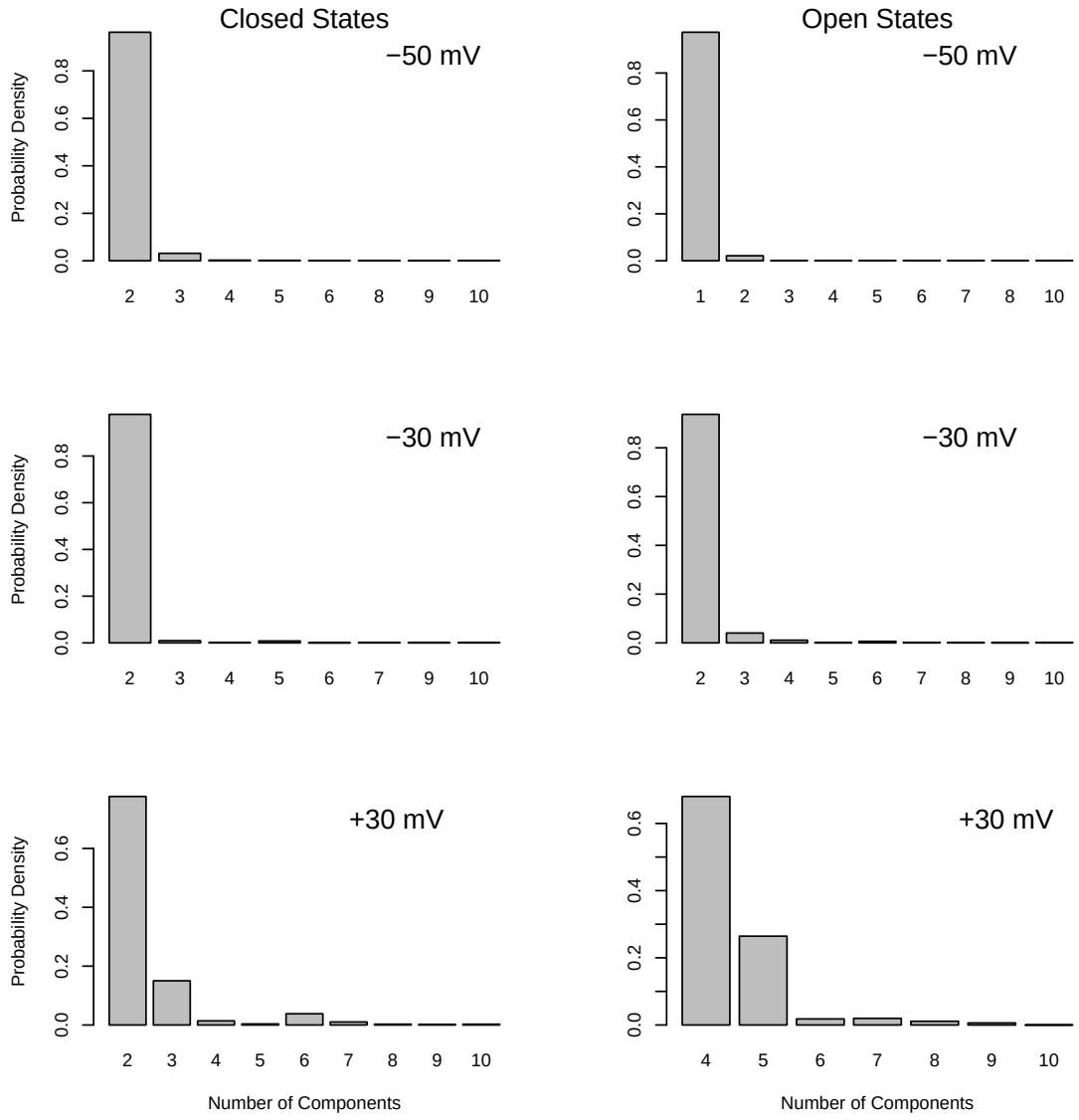

Figure S3: Application of infinite exponential mixture model to BK data. Analysis of dwell times from BK recordings at various holding voltages. For each trace, the posterior distribution over the number of mixture components is shown.

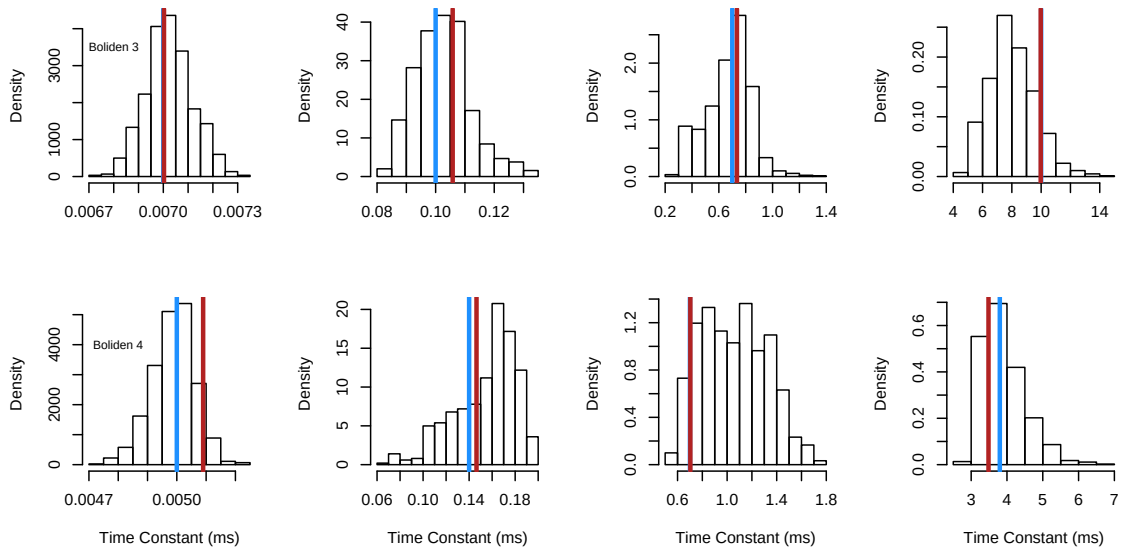

Figure S4: Application of infinite exponential mixture model to challenging datasets. The results of parameter inference are shown for the Boliden3 and Boliden4 parameter sets. Posterior distributions of time constant parameter are shown as histogram. The true parameter values are shown in blue and the point estimates from Landowne et al. are shown in red.

3 Sensitivity to Dirichlet Process Parameters

We now discuss the effects of Dirichlet process parameters on model inference. Recall that random probability measure, G, is a draw from a Dirichlet process as, $G \sim \text{DP}(\alpha, H)$. The Dirichlet process has two parameters, scalar α and probability measure H. Base measure H serves as the expectation of $G(A)$ (on any interval A) such that $\text{E}[G(A)] = H(A)$. Parameter α alters the variability of G around the expectation H, $\text{Var}[G(A)] = \frac{H(A)(1-H(A))}{\alpha+1}$, such that when α is large, G settles near H with low variance. With respect to the stick-breaking representation of the Dirichlet process, α tunes the expected size of the weights. Since the weights are related to iid draws from a Beta(1,α) distribution, large α results in many weights which are relatively small and a small value of α results in fewer weights which each occupy larger probability mass. Therefore, when using a Dirichlet process prior for model inference, the value of α will have an effect on the number of inferred model components. One approach to handling this complication is to incorporate uncertainty in α into the model by putting a parametric prior on α and marginalizing this uncertainty through the course of MCMC sampling (3). In the applications explored in the paper, we are primarily interested in applying these methods to distinct subsets of data, each of which represents an independent measurement or a measurement in a different experimental condition. In this way, we are most interested in comparing the inference results across different data subsets, where the inference algorithm is fixed in each case. Then, differences between the models inferred from each subset can be meaningfully compared, regardless of the uncertainty in α. Therefore, our strategy for choosing DP parameter values is to choose values which have accurate and reliable performance with simulated data and then fix these parameters for analysis of an entire dataset. In all cases, the relevant algorithm parameters used are reported in Figures 1 through 7.

It is important to conduct sensitivity analysis to determine how changes in α affect model inference. As an example, a Dirichlet process mixture of exponentials was used to model data simulated from a mixture of two exponentials where the components differed in time-scale by ten-fold ($N = 200$ data points). Figure S5 shows the result of this model inference for several fixed values of α. It is clear that over this range of α, the effect on the inferred models is negligible as the two component mixture is correctly inferred in each case. For the biophysical applications in the Results section, we fix $\alpha = 1$, which, when compared across distinct data subsets, is able to distinguish when a small number of components are in the data. For the Hierarchical Dirichlet process models (iHMM and iAMM), we incur an additional parameter γ, which also tunes the variability of a Dirichlet process around its base measure. Again, we choose to fix $\gamma = 1$, since this low value leads to good performance with simulated data. With the sticky-iAMM, we have an additional parameter κ which biases probability mass onto the diagonal elements of a transition matrix π. We fix $\kappa = 100$, which places a very weak prior on elements of π, since the traces used for analysis have 10^5 data points. Nonetheless, this weak prior is able to deter states which have zero dwell time and effectively accomplishes the goal of the sticky-iAMM. Despite uncertainty in these algorithm parameters, our strategy is to fix them to be small values which perform well with simulated data, because the primary goal is to compare between data sets given fixed values of these parameters.

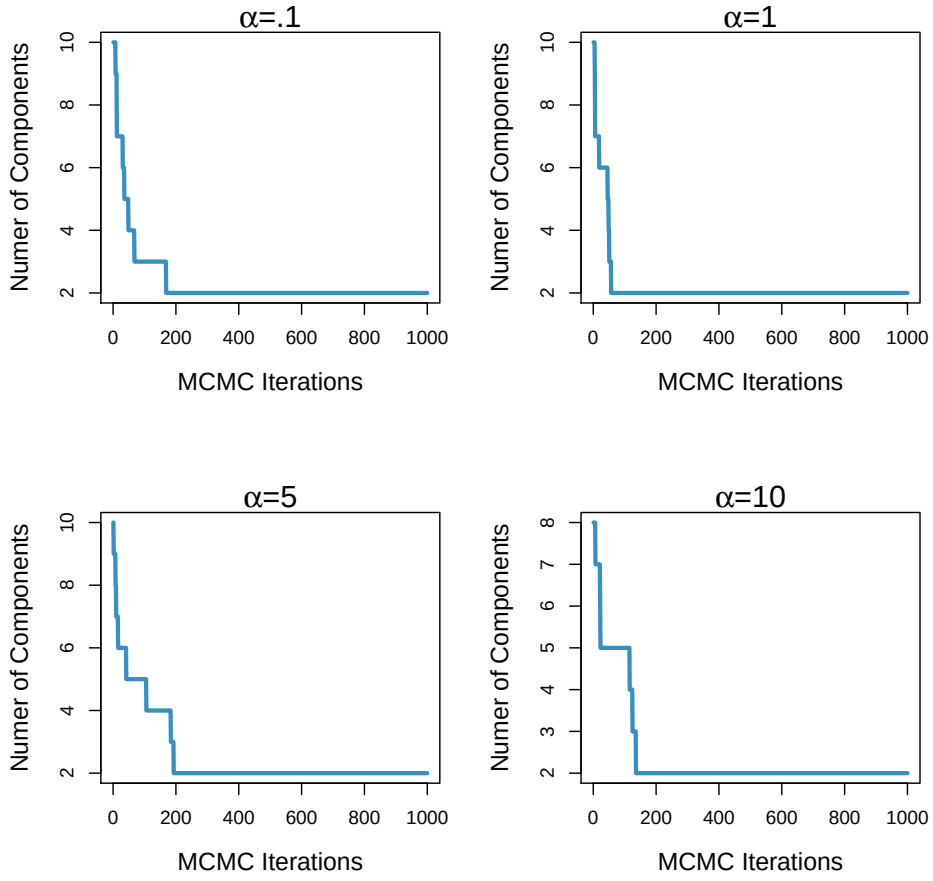

Figure S5: Sensitivity of Dirichlet process mixture models to values of α. Data was simulated as drawn from a mixture of two Exponential distributions which differ in time-scale by 10. The result of model inference for several fixed values of α. It is clear that over this range of α, the effect on the inferred models is negligible as the two component mixture is correctly inferred in each case.

4 Applications of infinite HMM

The infinite hidden Markov model can be used to analyze stochastic single molecule time series. In the main text, we used the iHMM to analyze data from electrophysiology, single molecule FRET, and single molecule photobleaching. Here, we discuss in more detail the benefits of using a nonparametric Bayesian approach for these time series. Figures S6 and S7 show the results of analyzing data from FRET and from photobleaching, respectively. In each case, several distinct traces are shown and the data points are colored according to which hidden state they are likely drawn from (Left columns). Since we use a Dirichlet process prior on the number of hidden states, we consider an infinite number of hidden states, yet through the course of Gibbs sampling, we integrate out this infinite measure. As a result, we gain a quantification of the posterior distribution over the number of hidden states likely to have generated the data. The Right columns in Figures S6 and S7 show this posterior distribution for each data trace. The posterior maximum provides a point estimate of the most probable number of hidden states, and the entire distribution provides a quantification of confidence in any given interpretation of the data. In this way, we not only consider the set of all possible models, but gain a confidence in any particular model of the data.

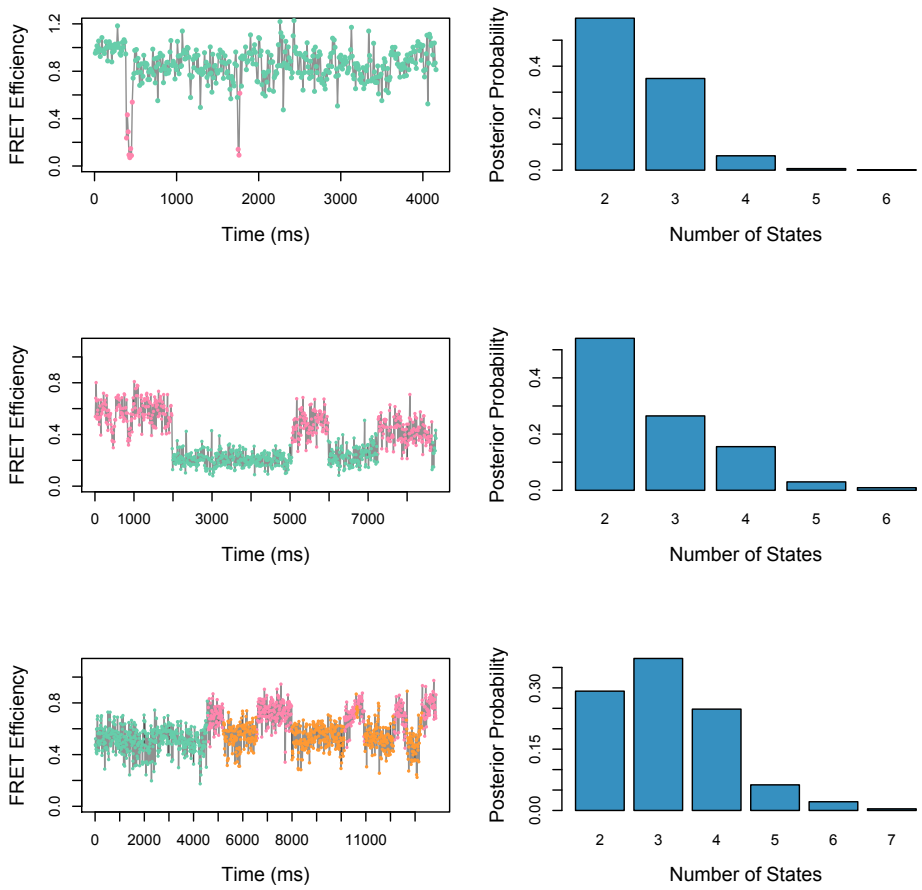

Figure S6: Application of iHMM to single molecule FRET. (Left column) Example traces of FRET efficiency over time. Sudden conformational changes are evident, but it it difficult to know the number of states and precise moment of state changes in these noisy traces. Colors indicate which hidden state each data point is assigned to. (Right column) Posterior distributions over number of hidden states inferred for each trace. The iHMM is able to decipher the number of the number of conformational states represented in these noisy time series. Algorithm parameters: $\alpha = 1, \gamma = 1$.

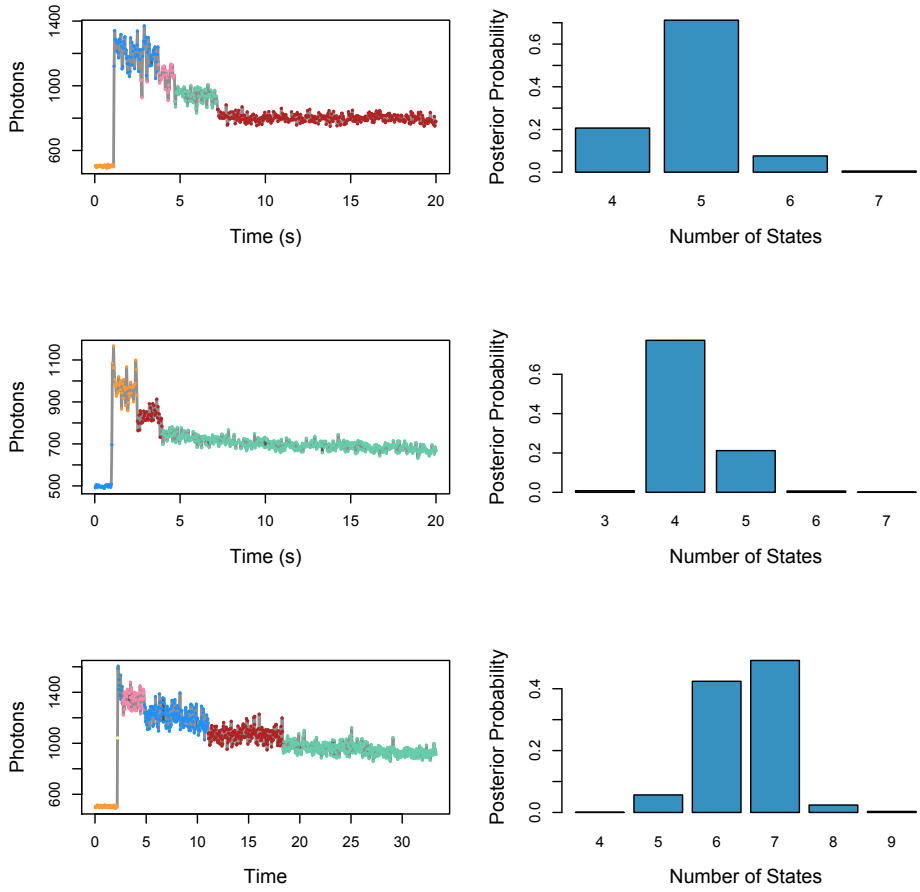

Figure S7: Application of iHMM to single molecule photobleaching. (Left column) Example traces of photon counts over time. Sudden photobleaching events are evident, but it it difficult to know the number of bleaching steps in the presence of noise. Colors indicate which hidden state each data point is assigned to. (Right column) Posterior distributions over number of hidden states inferred for each trace. The iHMM is able to decipher the number of the number of bleaching events and also provides a quantification of confidence. Algorithm parameters: $\alpha = 1, \gamma = 1$.

5 Extended Description of iHMM and iAMM

Here we describe in full detail the sampling methods underlying the iHMM and iAMM. We first describe a Gibbs sampling scheme for parameter inference with finite HMMs and then describe the implementation used for the iHMM.

For the hidden Markov model examples, we imagine our observations are normally distributed random variables and that each hidden state corresponds to a distinct mean θ_i and precision τ_i, such that $y_t \sim N(\theta_i, \frac{1}{\tau_i})$. Again, let A_i denote the set of all t for which $s_t = i$. For the means, θ_i, we use a conjugate prior normal distribution $N(a, b)$. For each θ_i,

$$p(\theta_i|...) \propto N(M, V) \tag{1}$$

$$\text{where} \quad M = \frac{ab + \tau \sum_{t \in A_i} y_t}{|A_i|\tau + b} \tag{2}$$

$$V = \frac{1}{|A_i|\tau + b} \tag{3}$$

With a conjugate gamma prior,$p(\tau_i) = \text{Ga}(c, d)$, on the precisions, τ_i,

$$p(\tau_i|...) \propto \text{Ga}(A, B) \tag{4}$$

$$\text{where} \quad A = \frac{d + |A_i|}{2} \tag{5}$$

$$B = \frac{1}{bc + \frac{1}{2}\sum (y_t - \theta_i)^2}. \tag{6}$$

Sampling the transition matrix, π, is simple conditioned on the previous samples of hidden states $s_1, ..., s_N$. First, we use the standard Dirichlet distribution prior for rows of the transition matrix, ie. $p(\pi_i) = \text{Dir}(m, ..., m)$. Let matrix N track the number of transitions between hidden states i and j such that $N_{i,j} = \sum_t I(s_t = j|s_{t-1} = i)$. Then each row of the transition matrix is sampled as,

$$p(\pi_i|...) \propto \text{Dir}(N_{i,1} + m, ..., N_{i,K} + m). \tag{7}$$

Finally, the hidden states, s_t, are sampled using the forward-filter-backward-sampler method (4). First we construct the $K \times N$ forward matrix F in the following way. For each datapoint, y_t, first compute vector O which quantifies the conditional probability of observing y_t given the emission distributions of each hidden state,

$$O = \begin{bmatrix} p(y_t|\theta_1, \tau_1) \\ p(y_t|\theta_2, \tau_2) \\ . \\ . \\ . \\ p(y_t|\theta_K, \tau_K) \end{bmatrix}. \tag{8}$$

We then combine the observation probabilities, the transition probabilities, and the occupancy probabilities from the previous time step,

$$L = (O \times \pi) \bullet F_{,t-1} \tag{9}$$

$$F_{,t} = \frac{L}{\sum L}. \tag{10}$$

Having computed F deterministically, we use Gibbs sampling on the backwards pass. Starting at time step N, we move backwards through each time step t, and combine F with the transition probability

$$L = F_{,t} \bullet \pi_{,s_{t+1}} \tag{11}$$

$$\vec{p} = \frac{L}{\sum L}. \tag{12}$$

We sample s_t from the resulting multinomial distribution,

$$p(s_t|...) \propto \text{Mult}(\vec{p}). \tag{13}$$

The result of this forward-backward sampler is a new sample of $s_1, s_2, ..., s_N$. For any hidden Markov model of fixed size, K, this Gibbs sampler allows us to calculate posterior distributions of all relevant parameters.

Generalizing this model to the infinite case will proceed similarly as with a mixture model. Again, the problem is that we now wish to consider the probability of transitions to each of an infinite number of hidden states, a computation that we cannot perform in our existing Gibbs sampler. However, using the hierarchical Dirichlet process hidden Markov model, we

can sample from both the currently instantiated hidden states as well as the infinitely many other hidden states which have yet to be sampled (5),

$$p(s_t = j|\mathbf{s}^-, \beta, \alpha, \mathbf{y_N}) \propto \begin{cases} (N_{s_{t-1},j} + \alpha\beta_j)\frac{N_{s_{t+1}}+\alpha\beta_{s_{t+1}}}{N_{k,}+\alpha} & j \leq k^-, k^- \neq s_{t-1} \\ (N_{s_{t-1},j} + \alpha\beta_j)\frac{N_{s_{t+1}}+1+\alpha\beta_{s_{t+1}}}{N_{k,}+1+\alpha} & j = s_{t-1} = s_{t+1} \\ (N_{s_{t-1},j} + \alpha\beta_j)\frac{N_{s_{t+1}}+\alpha\beta_{s_{t+1}}}{N_{k,}+1+\alpha} & j = s_{t-1} \neq s_{t+1} \\ \alpha\beta_j\beta_{s_{t+1}} & j = k^- + 1 \end{cases} \tag{14}$$

The sampling scheme works well, but it was noted that since Markov-type models will inherently have very high correlation between the latent variables, this form of Gibbs sampling could mix very slowly. To remedy this, (6) proposed the beam sampler for iHMMs. This implementation combines the dynamic programming approach described previously (forward-filter backward-sampler) with the slice sampling approach of (7). As described previously, the model is augmented to include latent variables $u_1, ..., u_N$ in order to limit the computation to a finite number of hidden states (at each iteration of MCMC). Once the appropriate number of states, k^*, is computed from $\vec{u}$, then we proceed with the Gibbs sampler just described for finite HMMs. Again, throughout the course of MCMC, resampling $\vec{u}$ results in fluctuations in the number of hidden states represented such that the aggregate of all MCMC samples results in integration over the infinite number of states. Sampling for β is performed using standard sampling methods for hierarchical Dirichlet process models (5). For our analyses of single molecule time series, we have utilized this beam sampling approach, and refer the reader to (6) for additional details.

In the iHMM, it was assumed that each hidden state corresponds to a distinct emission distribution, $p(y_t|\theta_i)$. In some cases, we might want to model a degeneracy such that multiple hidden states share the same emission distribution. In this aggregated Markov model (8), we imagine that the hidden states appear as aggregated into one of A distinct emission distributions such that $A < K$. We augment the iHMM with an indicator variable, $a_t \in \{1, 2, ..., A\}$, that specifies which aggregate each data point is drawn from such that $y_t \sim p(y_t|\theta_{a_t})$. This does very little to change the Gibbs sampler described above for HMMs and iHMMs. We simply need to sample each a_t in proportion to $[p(y_t|\theta_1, \tau_1), p(y_t|\theta_2, \tau_2), ..., p(y_t|\theta_A, \tau_A)]$. In the applications here, this model is applied to data from single ion channel recordings and A is fixed to be two. For each a_t, we sample

$$a_t \sim \text{Mult}(p(y_t|\theta_1, \tau_1), p(y_t|\theta_2, \tau_2)) \tag{15}$$

Intuitively, the likelihood $p(y_t|\theta_1, \tau_1)$ would correspond to, say, the probability of observing y_t given the channel was in an open state (any open state) at time t and $p(y_t|\theta_2, \tau_2)$ would correspond the likelihood of y_t given a closed state. The addition of the latent variables a_t has added minimal complexity to the Gibbs sampler for HMMs, and everything else remains the same, including the beam sampling. It is our intention with the iAMM that the number of aggregates, A, is known beforehand

and we mean to infer the number of hidden states within each aggregate. It would be possible to treat the number of aggregates as unknown and model both A and π nonparametrically, but we do not know of any interesting use for such a thing, so do not explore this possibility.

The use case for the iAMM is the analysis of single ion channel recordings, for which we add one additional feature to the model. Previous authors extended the infinite hidden Markov model framework by allowing for a strong preference for models with state-persistence (9). That is, we assume the time-scale of system dynamics is significantly slower than the data sampling rate. In this way, we are interested in solutions to the data where the system stays in each state for many time samples and we are intentionally not interested in models where states have zero dwell-time before transitioning. This certainly seems to be the case with ion channels, where from dwell-time distributions, we imagine that the channel tends to stay in each state for multiple time samples (at least). Following (9), we employ a sticky-iAMM by biasing probability mass onto the diagonal elements of the transition matrix π. By ensuring non-zero probability mass on the diagonal of π, we exclude models where states transition arbitrarily quickly to other states. To achieve this, we make a slight alteration to the algorithm described in the previous section. We add a hyper-parameter κ, the magnitude of which tunes the stickiness of the resulting Markov model. Each row of π is drawn from a Dirichlet process, with the diagonal elements biased by κ,

$$\pi_j \sim \mathrm{DP}(\alpha + \kappa, \frac{\alpha\beta + \kappa\delta_j}{\alpha + \kappa}), \tag{16}$$

and the rest of the algorithm remains the same. Incorporating uncertainty in κ into the sampling model should be possible in principle (3), but we prefer to use a fixed value. In experiments with simulated data, $\kappa = 100$ works well, and we use this same value for all ion channel data analyzed.

References

1. Rosales, R., 2004. MCMC for hidden Markov models incorporating aggregation of states and filtering. Bulletin for Mathematical Biology 66:1173–1199.
2. Siekmann, I., L. Wagner, D. Yule, C. Fox, D. Bryant, E. Crampin, and J. Sneyd, 2011. MCMC estimation of Markov models for ion channels. Biophysical Journal 100:1919–29.
3. Escobar, M., and M. West, 1995. Bayesian Density Estimation and Inference Using Mixtures. Journal of the American Statistical Association 90:577–588.
4. Scott, S., 2002. Bayesian Methods for Hidden Markov Models: Recursive Computing in the 21st Century. Journal of the American Statistical Association 97:337–351.
5. Teh, Y., M. Jordan, M. Beal, and D. Blei, 2006. Hierarchical Dirichlet Processes. Journal of the American Statistical Association 101:1566–1581.
6. van Gael, J., Y. Saatci, Y. Teh, and Z. Ghahramani, 2008. Beam Sampling for the Infinite Hidden Markov Model. Proceedings of the

25th International Conference on Machine Learning 1088–1095.

7. Walker, S., 2007. Sampling the Dirichlet mixture model with slices. Simulation and Computation 36:45–54.
8. Kienker, P., 1989. Equivalence of aggregated Markov models of ion-channel gating. Proceedings of the Royal Society of London B 236:269–309.
9. Fox, E., E. Sudderth, M. Jordan, and A. Willsky, 2011. A Sticky HDP-HMM with Application to Speaker Diarization. Annals of Applied Statistics 5:1020–56.