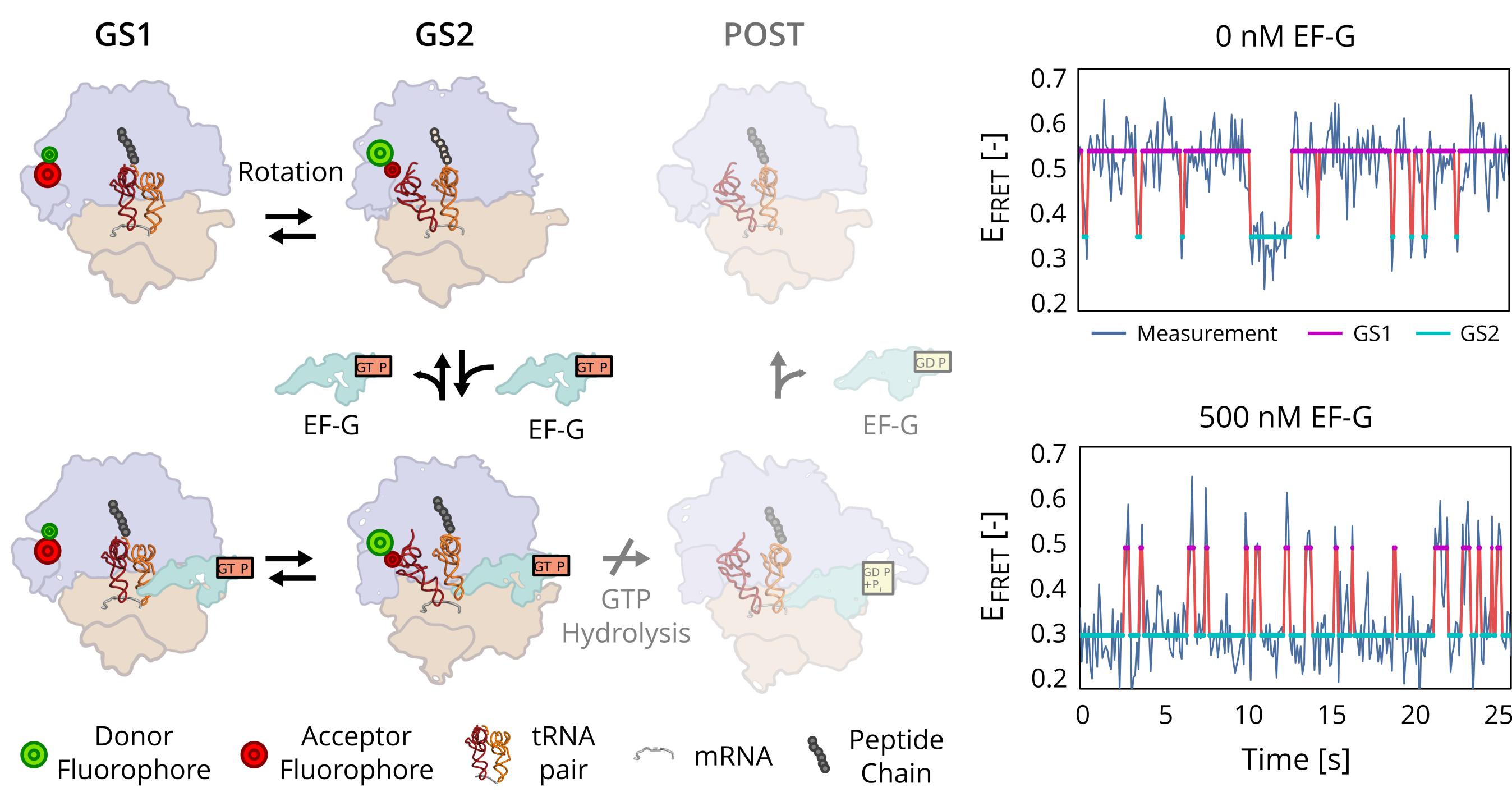


Hierarchically-coupled hidden Markov models for analysis of single-molecule experiments

Jan-Willem van de Meent¹ Jonathan E. Bronson² Frank Wood³ Ruben L. Gonzalez, Jr.⁴ Chris H. Wiggins⁵

¹janwillem.vandemeent@columbia.edu ²jonathan.bronson@gmail.com ³fwood@robots.ox.ac.uk ⁴rlg2118@columbia.edu ⁵chris.wiggins@columbia.edu

smFRET measurements in the Ribosome



In smFRET experiments, the anti-correlated intensity of two fluorophores reports on an intra-molecular distance, enabling direct observation of the distinct steps in kinetic pathways, such as shown here for translocation of the bacterial ribosome.

Commonly a single experiment yields observations for hundreds of individual molecules, which report on the same process but have variable photophysical properties and kinetics.

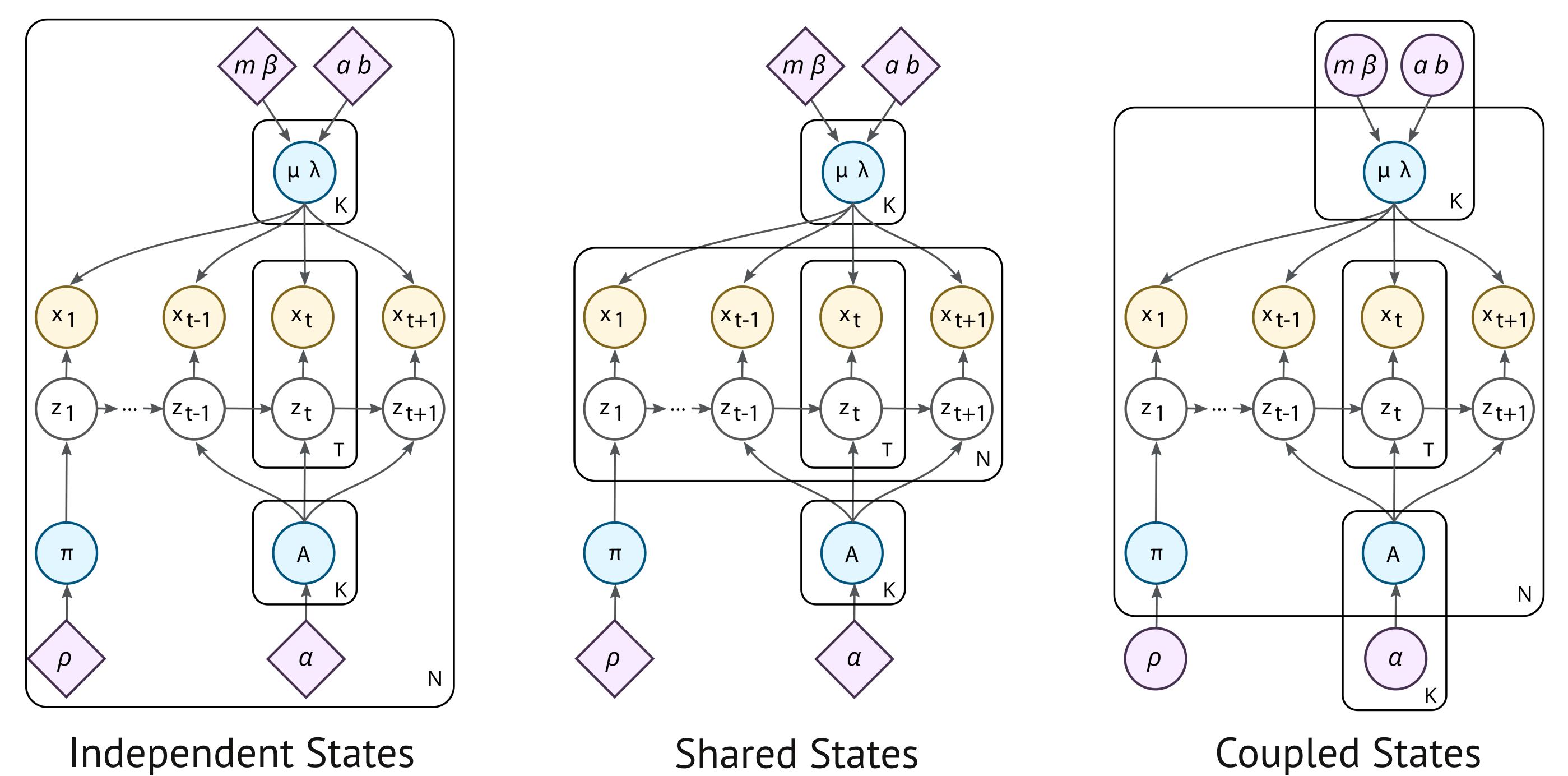
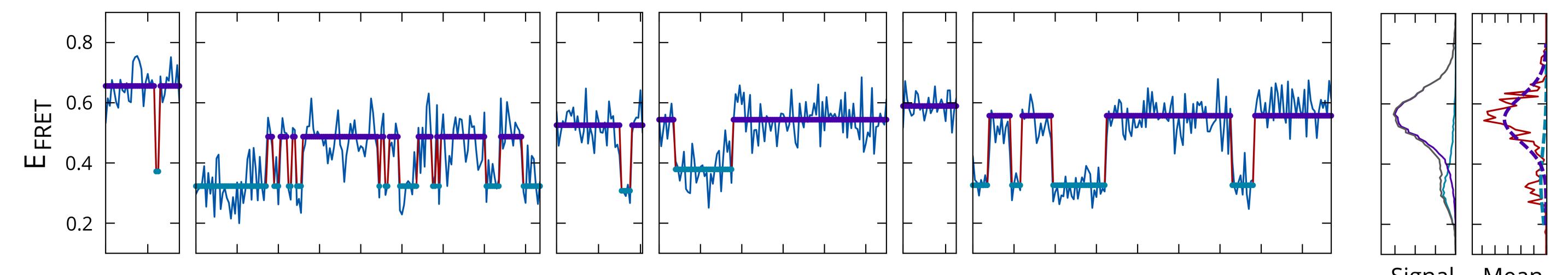
From this collection of noisy time series an experimentalist wishes to identify a set of conformational states, the transitions that occur between these states, and the typical associated kinetic rates.

Here we employ HMM analysis with a hierarchical coupling to learn a set of similar yet not identical parameters for each time series using an empirical Bayes (type II maximum likelihood) procedure.

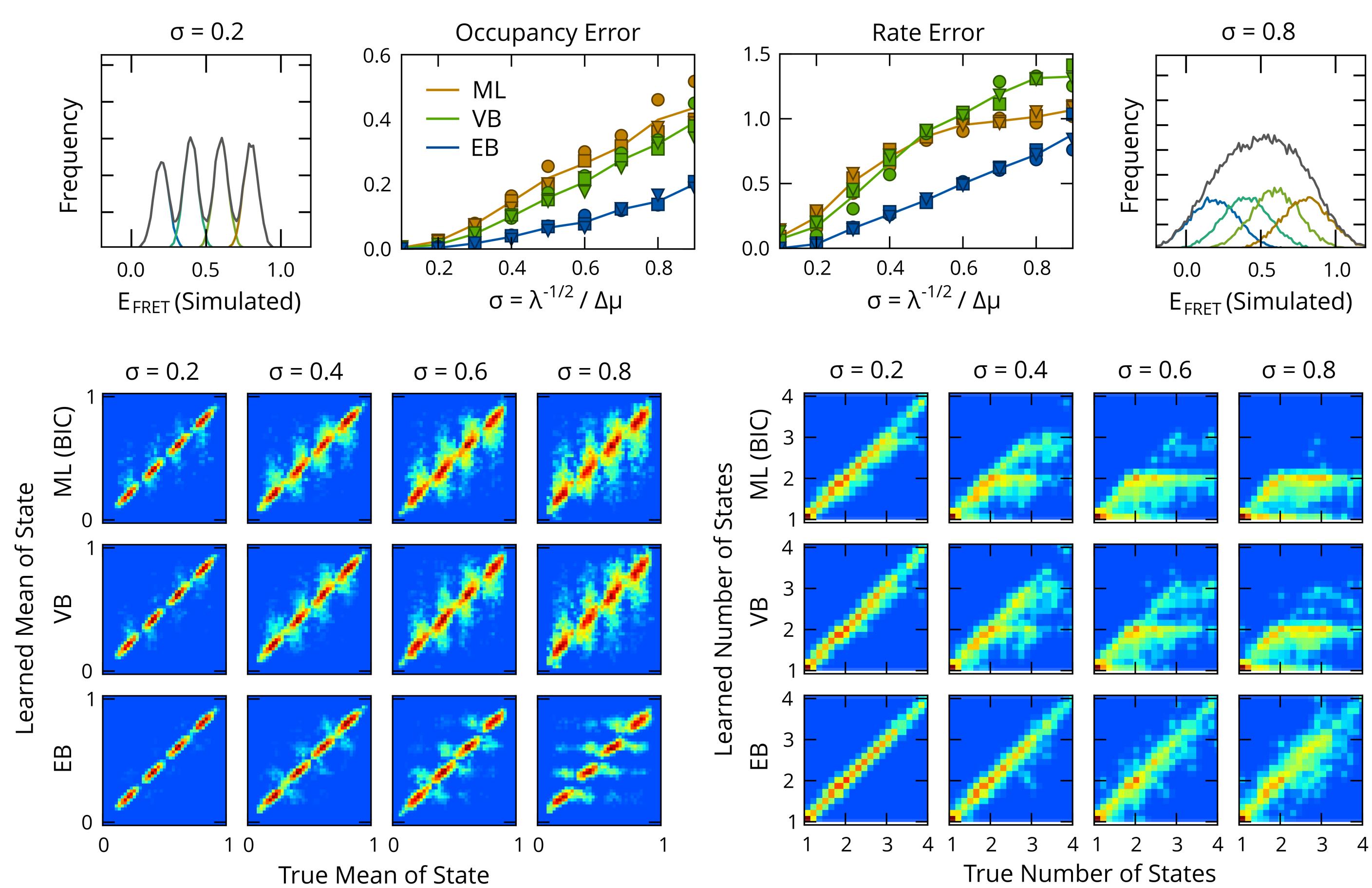
Learning HMMs from multiple time series

Existing HMM approaches in the single-molecule field learn independent HMMs for each time series and use ad-hoc heuristics to identify groups of similar states in an experiment.

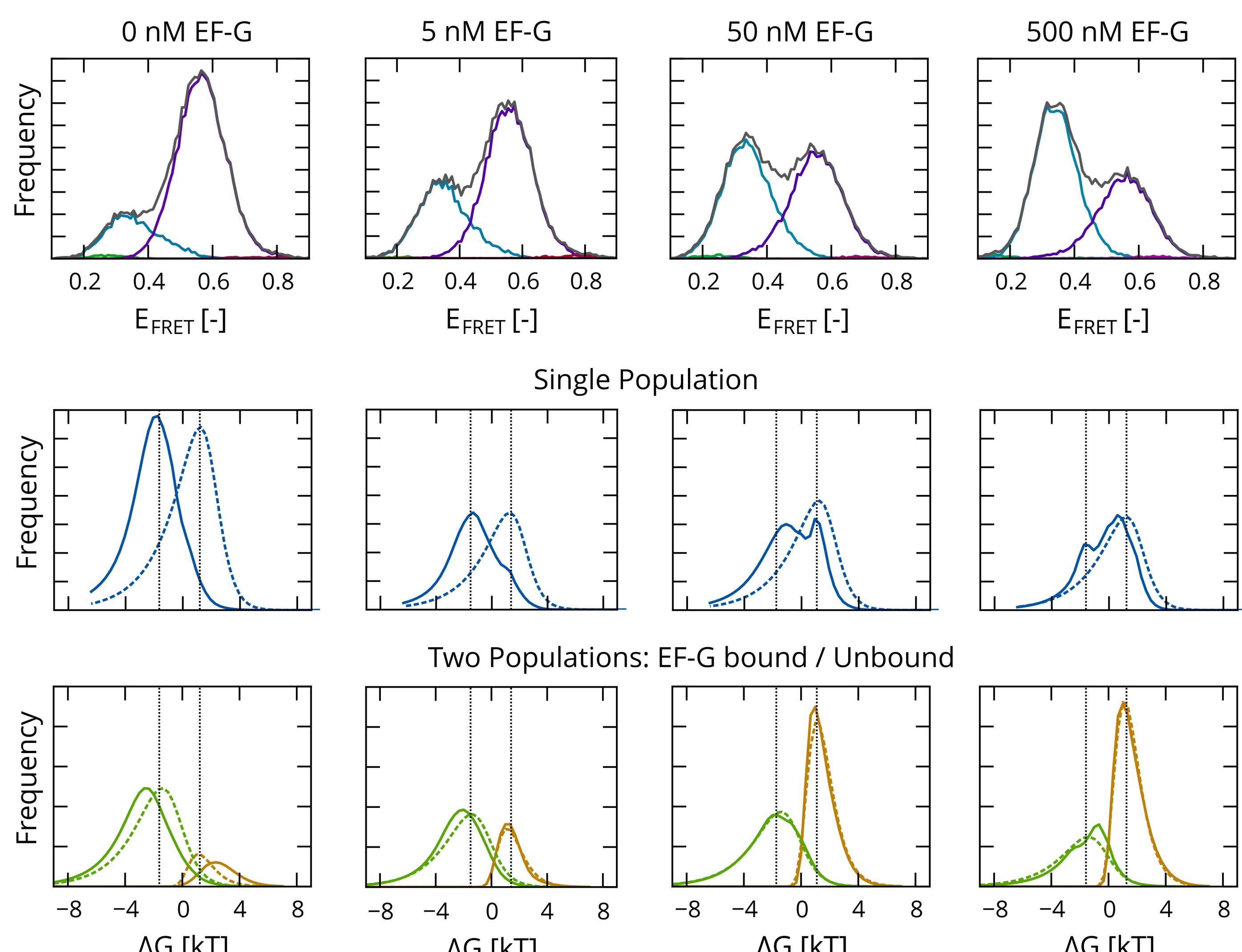
Our approach is to learn a set of priors for each state, which capture the common features as well as variability of the associated parameters in the ensemble of time series from an experiment.



Accuracy on Simulated Data



Subpopulation Analysis



Generative Model

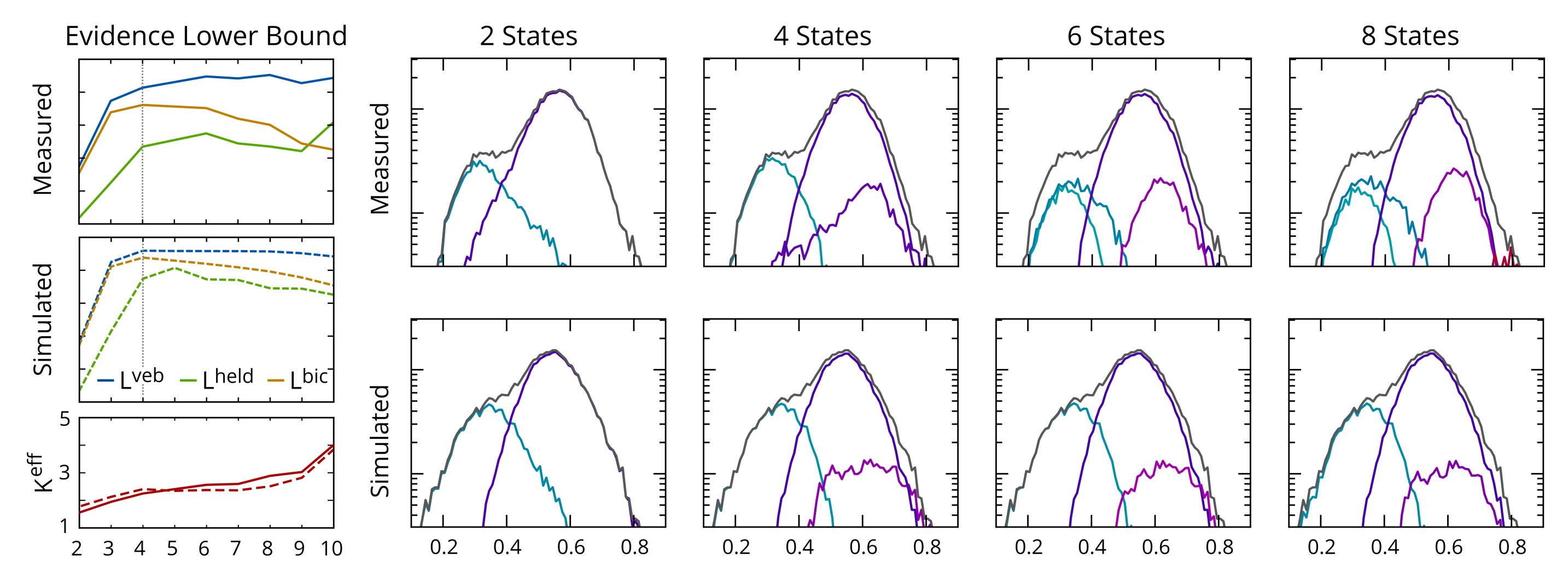
$$\begin{aligned} x_{n,t} | z_{n,t}=k &\sim \text{Normal}(\mu_{n,k}, \lambda_{n,k}) \\ z_{n,t} | z_{n,t-1}=k &\sim \text{Discrete}(A_{n,k}) \\ z_{n,0} &\sim \text{Discrete}(\pi_n) \end{aligned}$$

$$\begin{aligned} \mu_{n,k} &\sim \text{Normal}(m_k, \beta_k \lambda_{n,k}) \\ \lambda_{n,k} &\sim \text{Gamma}(a_k, b_k) \\ A_{n,k} &\sim \text{Dirichlet}(\alpha_k) \\ \pi_n &\sim \text{Dirichlet}(\rho) \end{aligned}$$

Variational Empirical Bayes Estimation

$$\begin{aligned} q(z_n) &= \underset{q(z_n)}{\operatorname{argmin}} D_{\text{KL}}[q(z_n)q(\theta_n) \| p(z_n, \theta_n | x_n, \psi)] \\ q(\theta_n) &= \underset{q(\theta_n)}{\operatorname{argmin}} D_{\text{KL}}[q(z_n)q(\theta_n) \| p(z_n, \theta_n | x_n, \psi)] \\ \psi &= \underset{\psi}{\operatorname{argmin}} \sum_n D_{\text{KL}}[q(\theta_n) \| p(\theta_n | \psi)] \end{aligned}$$

Model Selection



References

- Bronson, J. E., Hofman, J. M., Fei, J., Gonzalez, R. L., & Wiggins, C. H. Graphical models for inferring single molecule dynamics. *BMC bioinformatics*, 11 Suppl 8, S2, 2010.
- Bronson, J. E., Fei, J., Hofman, J. M., Gonzalez, R. L., & Wiggins, C. H. Learning rates and states from biophysical time series: a Bayesian approach to model selection and single-molecule FRET data. *Biophysical Journal*, 97(12), 3196-205, 2009.
- Corduneanu, Adrian and Bishop, CM. Variational Bayesian model selection for mixture distributions. *AISTATS*, 2001.
- Kass, R.E. and Steffey, D. Approximate Bayesian inference in conditionally independent hierarchical models (parametric empirical Bayes models). *JASA*, 84(407), 1989.

Source Code: <http://ebfret.github.io>