distribution of observables and characterization of rates of interconversion between states [11].

Hidden Markov models (HMMs) [?], which use temporal information in addition to the observable to determine which *hidden* state the system is currently in, have provided an effective solution to this problem [?]. In an HMM, the observed signal is assumed to come from a realization of an underlying Markov chain, where the system makes history-independent transitions among a set of discrete states with probabilities governed by a transition or rate matrix. The experimenter does not know which state the system is in, and can only measure some observable whose value is determined by a probability distribution of observables characterizing each state (which may overlap). Given a set of data, maximum likelihood estimates (MLEs) of the model parameters (transition rates and state observable distributions) and sequence of hidden states corresponding to the observed data can be determined by standard methods [12, 13].

Unfortunately, this approach has a number of serious limitations. Single-molecule experiments often suffer from limited statistics; the events of interest (transitions between states) may occur only a few times during the course of the measurement. As a result, while the MLE may give the most likely set of model parameters, there may be enormous uncertainty in these parameters, and the MLE provides no simple way to characterize them. These uncertainties may also be highly *correlated*, in that certain combinations of parameters may be well-determined in a complex way, despite individual parameters being poorly determined. The high cost (both in terms of instrument and experimenter time) of collecting additional data also means that it is not a simple task to judge *how much* data need be collected to test a particular hypothesis in a statistically meaningful way.

Here, we present a resolution to this issue in terms of a *Bayesian* extension of hidden Markov models applicable to single molecule experiments. By sampling over the posterior distribution of model parameters and hidden state assignments instead of simply finding the most likely values, the experimenter is able to accurately characterize the correlated uncertainties in both the model parameters (transition rates and state observable distributions) and hidden state sequences corresponding to observed data. Additionally, prior information (either from independent measurements or physical constraints) can be easily incorporated. The framework we present is based on Gibbs sampling [14, 15], allowing simple swap-in replacement of models for observable distributions, extension to multiple observables, and alternative models for state transitions. Additionally, the Bayesian method provides a straightforward way to model the statistical outcome and assess the utility of additional experiments given some preliminary data, allowing the experimenter a powerful tool for assessing whether the cost of collecting additional data is outweighed by their benefits.

This papers is organized as follows. In Section II, we present the mathematical framework behind hidden Markov models in the context of single-molecule experiments, and describe their Bayesian extensions. Section III describes the algorithms used in this study for computing both maximum-likelihood (MLE-HMM) and Bayesian (BHMM) hidden Markov models from experimental data. We apply these algorithms to optical force microscopy data collected on an RNA hairpin in Section IV, and discuss extensions of the method in Section V.

use MLHMM everywhere else.

[JDC: This introduction does not yet reference other applications of HMMs to single-molecule experiments or Bayesian HMM extensions.]

TABLE I. Important symbols and their elements.

O	$o_t^{(n)}$	observed temporal traces
S	$s_t^{(n)}$	hidden sequence of states
T	T_{ij}	transition probability for Δt
E	$\mathbf{e}_s$	state observable distribution parameters
$\boldsymbol{\Theta}$		model parameters $\boldsymbol{\Theta} \equiv \{\mathbf{T}, \mathbf{E}\}$
M	m	number of hidden states
N	n	number of independent observed traces
T_n	t	length of observed trace n
$\boldsymbol{\rho}$	ρ_i	initial state probability distribution
$\boldsymbol{\pi}$	π_i	equilibrium state probability
$\varphi(o\|\mathbf{e})$		state observable probability distribution

II. HIDDEN MARKOV MODELS

We now describe the basic theory behind the maximum likelihood estimate for a hidden Markov model (~~MLHMM~~) (MLE-HMM) and corresponding Bayesian extension (BHMM). While any scheme for computing the maximum-likelihood estimator or sampling from the Bayesian posterior can be used to generate these models, the algorithms used in this work are described in detail in Section III. Because of the complexity of mathematical notation ~~below,~~ we summarize important symbols used throughout in Table I.

A. Preliminaries

Suppose we observe N independent temporal traces, where some observable $O(x)$ that is a function of molecular configuration x is observed at temporal intervals Δt. This observable may be, for example, the measured force or extension of a polymer in a force microscopy experiment, an observed FRET efficiency, or an ion current measured by patch-clamp electrophysiology. While we restrict ourselves to consideration of scalar functions $O(x)$, the extension to multidimensional probes (or multiple probes) is straightforward.

Let trace n be denoted by $o_t^{(n)}$, where $t \in \{0, 1, 2, \ldots, T_n\}$, collected with uniform sampling interval Δt. We allow the system under observation to either start from equilibrium at the beginning of the observation period (if sufficient time has been allowed for the system to reach equilibrium), or from an out-of-equilibrium initial configuration (such as preparing a protein system by mechanically unfolding it prior to starting observation).

We presume the system under study has M kinetically metastable states, in the sense that they persist for many observation intervals Δt but may not represent the lowest free energy (most populous) state of the system[1]. We treat these as

[1] In the language of chemical kinetics, we require that the molecular relaxation time within a state $\tau_{\mathrm{mol}} \ll \Delta t$ but the typical reaction time for transitioning between states $\tau_{\mathrm{rxn}} \gg \Delta t$ [16–18].