V. Supplement

A. Additional Details on Optimal Transport

An important generalization of Wasserstein geodesics is the Wasserstein *barycenter*. For absolutely continuous probability measures with finite second moment $\nu_1, \ldots, \nu_m$ and coefficients $\{\lambda_i\}_{i=1}^m \subset \mathbb{R}_{\geq 0}$ with $\sum_{i=1}^m \lambda_i = 1$, the associated Wasserstein barycenter is

$$\nu_\lambda = \arg\min_{\nu} \sum_{i=1}^m \lambda_i W_2^2(\nu, \nu_i),$$

where the minimization is taken over all absolutely continuous measures and the *Wasserstein-2* distance between two measures is

$$W_2(\nu, \nu_i) = \min_{T_\# \nu = \nu_i} \sqrt{\int_{\mathbb{R}^D} \|T(x) - x\|^2 d\nu(x)}.$$

The special case $m = 2$ coincides with the Wasserstein geodesic. While this paper focuses on modeling transitions via the $m = 2$ case, it is of interest in future work to consider the $m > 2$ case as a richer way of modeling transitions between metastable states.

B. Optimal Transport Between Laplace Distributions

Recall that a 1-dimensional Laplace distribution with parameters $\mu \in \mathbb{R}$ (mean) and $\sigma \in (0, \infty)$ (standard deviation), denoted $\mathrm{Lap}(\mu, \sigma)$, has density function

$$f(x) = \frac{1}{2\sigma} \exp\left(-\frac{|x-\mu|}{\sigma}\right). \tag{2}$$

The following results verify our approach to computing barycenters between two Laplace distributions.

Proposition 1. *Let $\nu_0 \sim \mathrm{Lap}(\mu_0, \sigma_0)$ and $\nu_1 \sim \mathrm{Lap}(\mu_1, \sigma_1)$. Then the optimal transport map between ν_0 and ν_1 is given by $T_{\nu_0 \to \nu_1}(x) = \frac{\sigma_1}{\sigma_0}(x - \mu_0) + \mu_1$.*

Proof. We first show that $(T_{\nu_0 \to \nu_1})_\# \nu_0 = \nu_1$. Recall that this requires showing for all Borel sets $A \subset \mathbb{R}$ that

$$\int_A f_1(x) dx = \int_{T^{-1}_{\nu_0 \to \nu_1}(A)} f_0(x) dx$$

where f_i is the density of ν_i. Making the change of variables $y = T_{\nu_0 \to \nu_1}(x)$ or equivalently $x = T^{-1}_{\nu_0 \to \nu_1}(y)$ yields that this is equivalent to

$$\int_A f_1(x) dx = \int_A f_0(T^{-1}_{\nu_0 \to \nu_1}(y))(T^{-1}_{\nu_0 \to \nu_1}(y))' dy.$$

It is therefore enough to show that the densities f_0, f_1 satisfy the condition

$$f_1(x) = f_0(T^{-1}_{\nu_0 \to \nu_1}(x)) \cdot (T^{-1}_{\nu_0 \to \nu_1}(x))'$$

where

$$f_0(x) = \frac{1}{2\sigma_0} \exp\left(-\frac{|x-\mu_0|}{\sigma_0}\right),$$
$$f_1(x) = \frac{1}{2\sigma_1} \exp\left(-\frac{|x-\mu_1|}{\sigma_1}\right).$$

Noting that

$$T^{-1}_{\nu_0 \to \nu_1}(x) = \frac{\sigma_0}{\sigma_1}(x - \mu_1) + \mu_0,$$
$$(T^{-1}_{\nu_0 \to \nu_1}(x))' = \frac{\sigma_0}{\sigma_1},$$

we indeed have

$$\begin{aligned} & f_0(T^{-1}_{\nu_0 \to \nu_1}(x)) \cdot (T^{-1}_{\nu_0 \to \nu_1}(x))' \\ =& \frac{1}{2\sigma_0} \exp\left(-\frac{|\frac{\sigma_0}{\sigma_1}(x-\mu_1) + \mu_0 - \mu_0|}{\sigma_0}\right) \cdot \frac{\sigma_0}{\sigma_1} \\ =& f_1(x). \end{aligned}$$

We now observe that $T_{\nu_0 \to \nu_1}(x)$ is the gradient of the convex function $\phi(x) = \frac{\sigma_1}{2\sigma_0}(x - \mu_0)^2 + \mu_1 x$. It is therefore the unique optimal transport map by Brenier's Theorem [42], [43]. □

With the optimal transport map between two Laplace distributions known, any barycenter between them has a simple closed form.

Proposition 2. *Let $\nu_0 \sim \mathrm{Lap}(\mu_0, \sigma_0)$ and $\nu_1 \sim \mathrm{Lap}(\mu_1, \sigma_1)$. Then for any $t \in [0, 1]$, the Wasserstein barycenter between ν_0 and ν_1 with weights $(1 - t, t)$ is the Laplace distribution with parameters $\mu_t = (1 - t)\mu_0 + t\mu_1$ and $\sigma_t = (1 - t)\sigma_0 + t\sigma_1$.*

Proof. We note that the barycenter between ν_0 and ν_1 with weights $(1-t, t)$ coincides with the constant-speed Wasserstein geodesic [20] between ν_0 and ν_1 at time $t \in [0, 1]$, namely $\nu_t = (tT_{\nu_0 \to \nu_1} + (1 - t)\,\mathrm{Id})_\# \nu_0$. We compute the density f_t of ν_t as follows. Let

$$\begin{aligned} T_t(x) &= tT_{\nu_0 \to \nu_1}(x) + (1 - t)x \\ &= t\frac{\sigma_1}{\sigma_0}(x - \mu_0) + t\mu_1 + (1 - t)x \\ &= \frac{t\sigma_1 + (1 - t)\sigma_0}{\sigma_0} x + (t\mu_1 - \frac{t\sigma_1\mu_0}{\sigma_0}) \\ &= \frac{t\sigma_1 + (1 - t)\sigma_0}{\sigma_0} x + \frac{t\sigma_0\mu_1 - t\sigma_1\mu_0}{\sigma_0} \end{aligned}$$

so that

$$\begin{aligned} T_t^{-1}(x) &= \frac{\sigma_0}{t\sigma_1 + (1 - t)\sigma_0} \cdot \left(x + \frac{t\sigma_1\mu_0 - t\sigma_0\mu_1}{\sigma_0}\right) \\ (T_t^{-1})'(x) &= \frac{\sigma_0}{t\sigma_1 + (1 - t)\sigma_0}. \end{aligned}$$

We now compute:

$$
\begin{aligned}
&f_t(x) = f_0(T_t^{-1}(x)) \cdot (T_t^{-1}(x))' \\
&= \frac{1}{2\sigma_0} \exp\left(\frac{-|(x + \frac{t\sigma_1\mu_0 - t\sigma_0\mu_1}{\sigma_0}) \cdot \frac{\sigma_0}{t\sigma_1 + (1-t)\sigma_0} - \mu_0|}{\sigma_0} \right) \\
&\cdot \frac{\sigma_0}{t\sigma_1 + (1-t)\sigma_0} \\
&= \frac{1}{2(t\sigma_1 + (1-t)\sigma_0)} \\
&\cdot \exp\left(\frac{-|(x + \frac{t\sigma_1\mu_0 - t\sigma_0\mu_1}{\sigma_0}) \cdot \frac{\sigma_0}{t\sigma_1 + (1-t)\sigma_0} - \mu_0|}{\sigma_0} \right) \\
&= \frac{1}{2(t\sigma_1 + (1-t)\sigma_0)} \\
&\cdot \exp\left(\frac{-|(x + \frac{t\sigma_1\mu_0 - t\sigma_0\mu_1}{\sigma_0}) \cdot \frac{\sigma_0}{t\sigma_1 + (1-t)\sigma_0} - \frac{\mu_0(t\sigma_1 + (1-t)\sigma_0)}{t\sigma_1 + (1-t)\sigma_0}|}{\sigma_0} \right) \\
&= \frac{1}{2(t\sigma_1 + (1-t)\sigma_0)} \\
&\cdot \exp\left(\frac{-|(x + \frac{t\sigma_1\mu_0 - t\sigma_0\mu_1}{\sigma_0}) \cdot \sigma_0 - \mu_0(t\sigma_1 + (1-t)\sigma_0)|}{\sigma_0(t\sigma_1 + (1-t)\sigma_0)} \right) \\
&= \frac{1}{2(t\sigma_1 + (1-t)\sigma_0)} \\
&\cdot \exp\left(\frac{-|(\sigma_0 x + t\sigma_1\mu_0 - t\sigma_0\mu_1) - \mu_0(t\sigma_1 + (1-t)\sigma_0)|}{\sigma_0(t\sigma_1 + (1-t)\sigma_0)} \right) \\
&= \frac{1}{2(t\sigma_1 + (1-t)\sigma_0)} \\
&\cdot \exp\left(\frac{-|\sigma_0 x - (t\sigma_0\mu_1 + (1-t)\sigma_0\mu_0)|}{\sigma_0(t\sigma_1 + (1-t)\sigma_0)} \right) \\
&= \frac{1}{2(t\sigma_1 + (1-t)\sigma_0)} \cdot \exp\left(\frac{-|x - (t\mu_1 + (1-t)\mu_0)|}{(t\sigma_1 + (1-t)\sigma_0)} \right) \\
&= \frac{1}{2\sigma_t} \cdot \exp\left(\frac{-|x - \mu_t|}{\sigma_t} \right),
\end{aligned}
$$

as desired. □

We remark that in general, Wasserstein geodesics and barycenters preserve shape properties of the distributions being interpolated, unlike linear mixtures; see Figure 5.

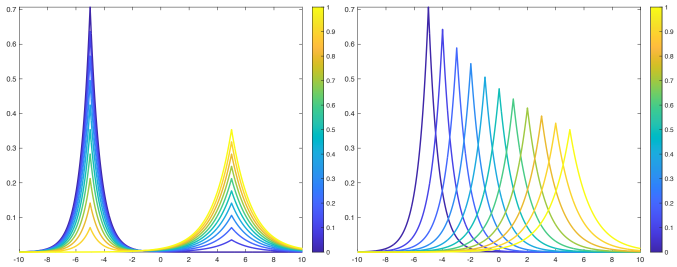

Fig. 5. Interpolation between two Laplace distributions via linear (left) and W_2 (right) geodesics. Time runs from 0 (blue) to 1 (yellow).

C. Additional Data Set Details

1) Synthetic Toy Models: The performance of BarT was compared to that of SIMPLE over a wide set of tuning parameters. In addition, the performance of BarT was evaluated on a batch of trajectories with varied transition lengths and variances. For each transition length, this length was applied to all the transition segments within a trajectory, with the rest of the trajectory consisting of 5000 points distributed among 20 segments, alternating between the two major states (μ = 100 and μ = 200). The metastable segment lengths were randomly generated, under the constraint that no metastable segment was allowed to be shorter than 100 points. This randomization was done using the randfixedsum program, which relies on the decomposition of the (n – 1)-dimensional solution space into a number of different types of simplexes, which are selected with probabilities proportional to their volumes, and within which vectors are distributed uniformly, to generate an arbitrary number of n-dimensional length vectors [44]. In other words, no rejection sampling is performed, and lengths are randomly generated in a way that ensures every solution fits the imposed constraint. All of the trajectories generated in this way are available at [22]. A comparison of BarT, SIMPLE, and PELT on different tuning parameter values and two different noise levels is shown in Figure 7. An illustrative example showing the robustness of BarT to transition length is provided in Figure 6.

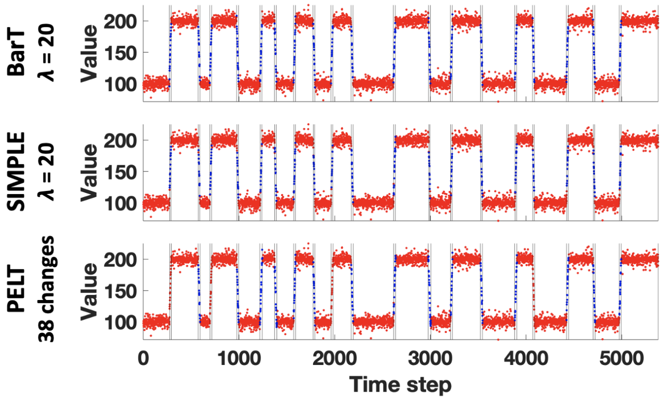

Fig. 6. A synthetic, 1-D, two-state trajectory with $\sigma = 5$ and transition length = 20. BarT reproduces the ground truth change point placements near-perfectly, with SIMPLE performing almost as well, only slightly underestimating the transition lengths. PELT results in correct labeling of all but four segments. Segments labeled as metastable are shown in red, and segments labeled as transitions are shown in blue.

2) Langevin Dynamics Trajectory: The two-well potential landscape is shown in Figure 8 below. In addition, the performance of BarT in a high-sensitivity regime is demonstrated in Figure 9. It can be seen that under this tuning setting, both local fluctuations and transitions between the major states are identified as changes. Unlike SIMPLE, BarT and PELT are able to identify gradual transitions throughout the dataset.

3) Alanine Dipeptide MD: In a nutshell, an all-atom molecular dynamics simulation consists of repeatedly solving the

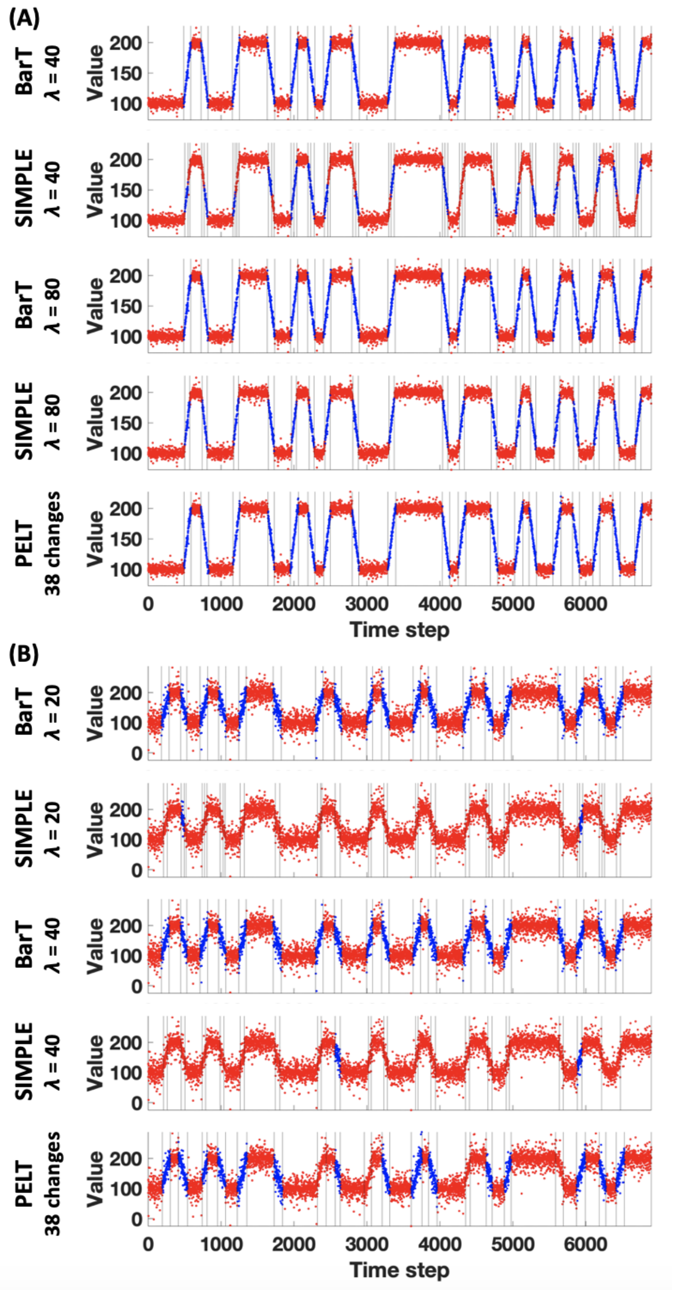

Fig. 7. Comparative performance of BarT, SIMPLE, and PELT on (A) a low-noise (σ = 5) and (B) a high-noise (σ = 20) synthetic trajectory, with two different tuning parameter values for SIMPLE and BarT (PELT is tuned to the ground-truth number of changes). Note that SIMPLE both under- and over-segments depending on tuning, whereas BarT remains robust.

Newtonian equations of motions for all atoms in a system of interest. Based on sets of parameters, commonly referred to as force fields, which describe the energetics of particle interaction, one may compute the forces acting on each atom in the system. Based on these forces, one may calculate the overall resulting force on each atom, which may be used to find its acceleration. As an iterative process taking place over very small discrete time steps, in each iteration, the forces, accelerations, velocities, and ultimately positions, are updated, causing the system to move and explore different conformations. To minimize numerical errors, the positions and velocities are typically updated at staggered time intervals, and for this reason the "standard" MD integrator is also referred to as the leapfrog integrator. Additional algorithms may be used to constrain chemical bonds (i.e., distances between atoms forming such bonds) to their known lengths, or stabilize the simulation by maintaining a constant temperature or pressure. An illustration of the MD simulation process is shown in Figure 10.

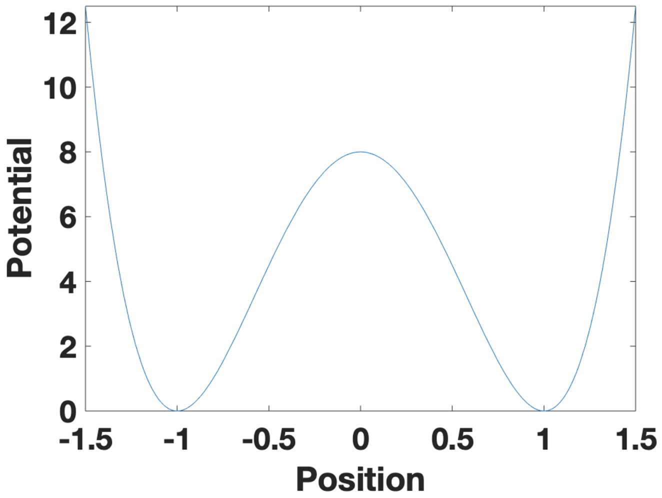

Fig. 8. The double-well potential used in the single-particle Langevin dynamics simulation. Note that adjusting the barrier height and well separation may allow the user to alter the transition frequency and speed.

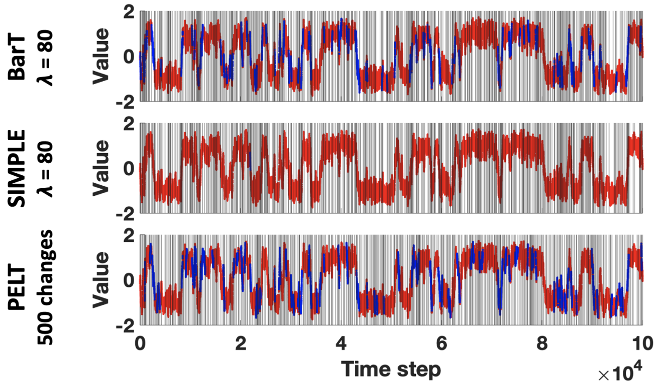

Fig. 9. The performance comparison of BarT, SIMPLE, and PELT on the Langevin dynamics data set in a high-sensitivity (λ = 80 for SIMPLE and BarT, 500 changes for PELT) regime. Change points are marked with vertical lines; segments labeled metastable are shown in red, and those labeled as transitions are shown in blue.

In this work, we apply the MD approach to a commonly-used model system, alanine dipeptide (Figure 11). Alanine dipeptide consists of a single amino acid, alanine, characterized by a single methyl-group side chain, connected by amide (CONH) bonds on the N- and C-terminal sides to an acetyl cap, and an N-methyl cap, respectively. The two rotatable bonds on the alanine backbone describe the conformational transitions of this system, taking what is fundamentally at least a 30-dimensional data set (3 Cartesian coordinates for each of

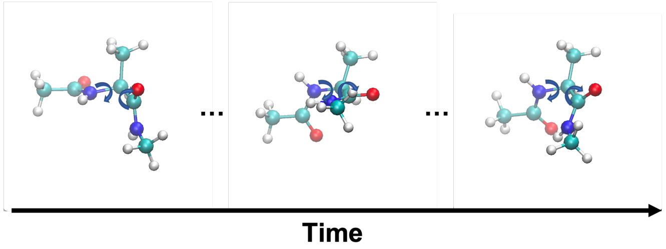

Fig. 10. An illustration of an MD simulation of alanine dipeptide. The geometry of the system evolves over time, and different conformations are sampled. The raw simulation trajectory consists of positions of all atoms in 3-D space (shown as spheres in the ball-and-stick model), which can be used to compute lower-dimension internal coordinates (angles ϕ and ψ shown as blue arrows).

the ten heavy atoms) prone to numerical errors (since a rotation + translation transformation would have to be carried out on each frame to align it to a reference before similarity between frames could be evaluated) to a 2-dimensional data set based on *internal* coordinates, which do not need to account for the system's translation around the solvent box.

That backbone torsion angles are the fundamental descriptor of this system's motions is particularly convenient, as we can readily visualize the conformations sampled during the simulations on a Ramachandran plot. Strictly speaking, a Ramachandran plot is a plot representing the *free energy* of all backbone conformations of an amino acid residue in (ϕ, ψ) space [45]. Conformations with a lower free energy are more thermodynamically stable, and, in the context of an MD simulation, more frequently sampled. For natural, L-chirality amino acids, such as the ones that exist in the human body, the *canonical* regions are taken to be ones corresponding to common secondary structure motifs in proteins: right-handed α helix (lower light blue area in the leftmost panel of 4A), β sheet (upper light blue area in the same figure), polyproline II (red area), and left-handed α helix (central blue area in the $\phi > 0$ region). The remaining regions which are energetically allowed lie between these canonical areas, and we proposed that they may be categorized as transitions, rather than metastable states in their own right.

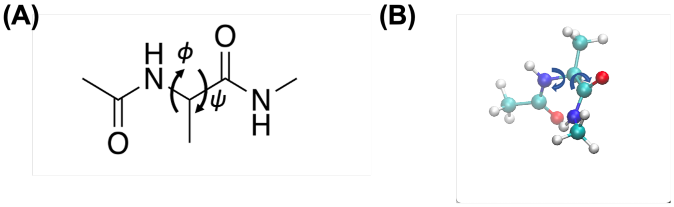

Fig. 11. (A) The chemical formula of capped alanine dipeptide. Note that each node represents a carbon atom with enough hydrogen atoms bound to it to make a total of four chemical bonds. The dihedral torsion angles corresponding to the main conformational transitions, ϕ and ψ, are indicated by curved arrows around the torsion bond. (B) A ball-and-stick model of the alanine dipeptide molecule. Nitrogen atoms are shown in dark blue, oxygen atoms are shown in red, carbon atoms are shown in cyan, and hydrogen atoms are shown in white.

The dihedral angle ϕ is characterized by fast local fluctuations, which make this time series particularly challenging to perform change detection on; in the end, very few trajectory segments are actually marked as transitions. This time series, along with BarT change assignments for λ = 10, is shown in Figure 12.

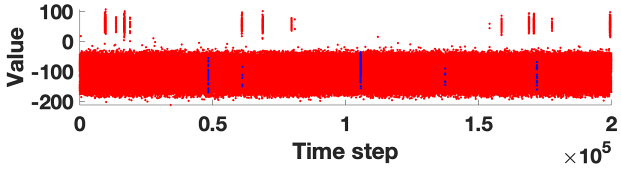

Fig. 12. The time series corresponding to the values of the dihedral angle ϕ over the course of the 200 ns MD simulation. Points marked as belonging to metastable states are shown in red; points marked as belonging to transitions are shown in blue.

D. Algorithm Details

In order to optimize (1), we use the following methods.

1) Initial Placement: Analogously to the approach of [17] we use the approach of [35] to perform a pruned dynamic programming computation to compute the set of change points which maximizes the penalized log-likelihood function (1). This procedure recursively calculates the penalized log-likelihood for all times t as $F(t) = \max_s F(s) + \hat{l}(Y_{s+1}, ..., Y_t) - \lambda p_s$, where Y is the univariate time series, p is the penalty function, and λ is a tuning parameter, with the maximum taken over all times $s < t$ not prohibited by the pruning scheme from being the last change point before time t. For more details on the pruning scheme, see [35].

2) Optimization: Following the initial change point placement, once the univariate change point placements are stabilized, we perform a shift-and-merge iterative sequence to avoid being "stuck" in a local minimum of the objective likelihood function. In each iteration of this sequence, change points are taken individually, in chronological order, and the log-likelihood of the dataset being described by an updated change point set, with the individual change point deleted or moved anywhere between its neighboring change points, is recalculated. The log-likelihood is calculated for both models (Laplace distribution or barycenter) for each segment, and the maximum of the two log-likelihoods is taken as a segment's log-likelihood contribution. The process continues until the change point placement is fully converged or a set number of iterations has been reached. Throughout the optimization process, the penalty function is also updated to account for the number of change points and, if α is set to a value other than 1, concurrent change points in multiple variables. Note, however, that while other variables are considered for the simultaneity-based penalty modification calculation, only *one change point in one variable* is moved in each step of the shift-and-merge process. This approach is thus able to capture both one- and multidimensional changes, which makes it readily applicable to diverse types of free energy landscapes.

3) Labeling: Once the change point placements are finalized, the log-likelihood of each trajectory segment corresponding to either a Laplace distribution or a weighted barycentric interpolation between Laplace distributions corresponding to its neighboring segments is calculated, and the greater of the two is taken as the segment "label". This scheme was also used to label and evaluate segments derived from SIMPLE change point detection. For multivariate data, segments are labeled in each dimension separately.

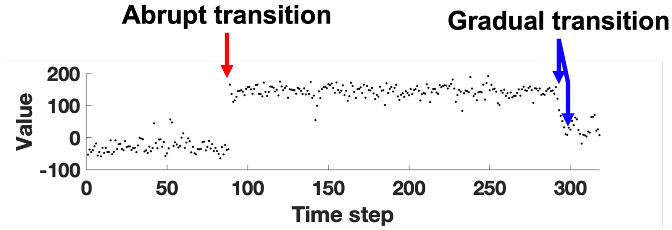

Fig. 13. A trajectory snippet from the ψ time series of the alanine dipeptide MD trajectory. An abrupt transition is marked with a red arrow, and a gradual transition is marked with two blue arrows (denoting the beginning and the end of the transition).

4) Parameter Tuning: Due to the low dimensionality of the data sets used in this work (one-dimensional toy model and Langevin dynamics simulation, and an effectively two-dimensional alanine dipeptide MD simulation trajectory), the simultaneity tuning parameter α was left at 1 for all experiments. The alanine dipeptide trajectory was analyzed using $\alpha = 0.7$ as well, but the results were not found to be qualitatively different. For the synthetic data sets, the λ tuning parameter was adjusted for both algorithms in increments of 10 until optimal change point placement and/or labeling were observed. For the alanine dipeptide dataset, a λ value of 10 was used in line with prior work [12]. On data sets without *a priori* knowledge of optimal tuning parameters, we recommend setting the initial value of λ = 20, inspecting the results, and adjusting the parameter in small increments as needed. A good initial value for α in multivariate data is 0.7, which is in line with literature [17].

E. Additional Results

1) PELT Parameter Tuning: As previously discussed, PELT may be run in "minimum improvement threshold" mode; the user-set threshold serves a purpose analogous to that of the λ parameter in BarT and SIMPLE, effectively associating a cost with placing additional candidate change points. An overview of results with this parameter manually tuned is shown in Figure 14. With manual parameter tuning, it is possible to obtain results that are qualitatively similar to those obtained with a fixed number of change points (Figure 14, panels A–D). However, the parameter adjustment is not as straightforward as it is for the two penalized likelihood estimation approaches, with "good" parameter ranges being somewhat difficult to estimate, and optimal parameter values varying by orders of magnitude dependent on data set features (compare Figure 14 panels D–G to Figures 7, 3, and 9).

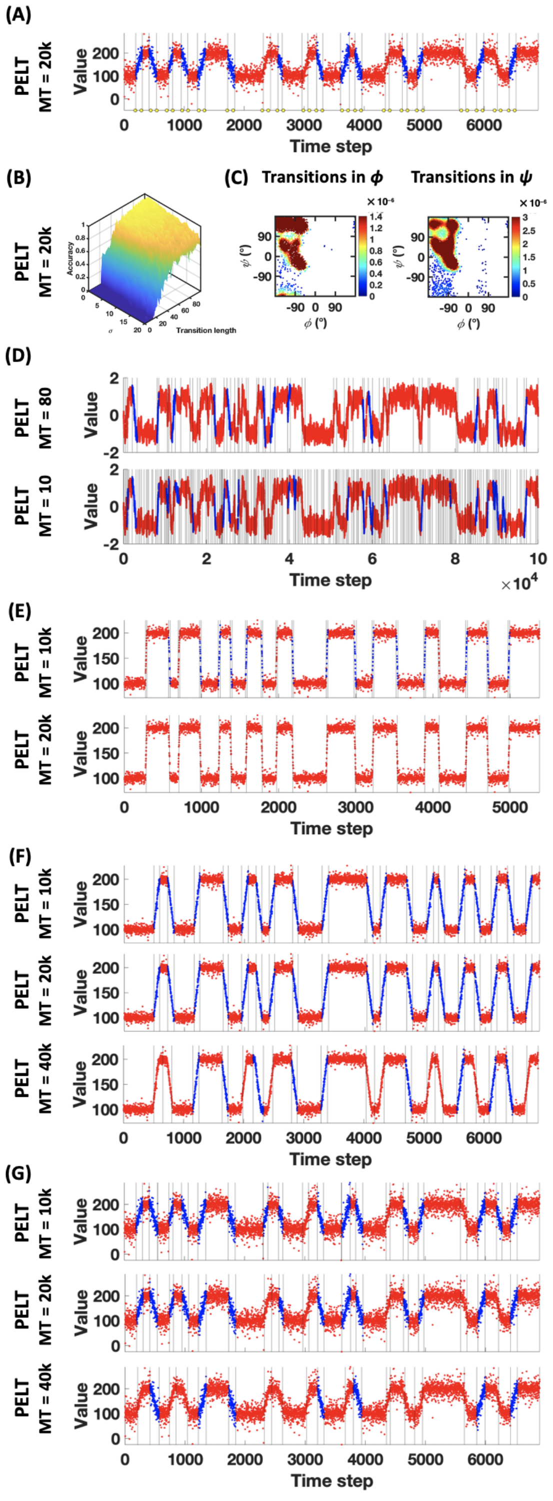

Fig. 14. An overview of performance of PELT in "minimum improvement threshold" mode. (A) Representative trajectory from the synthetic data set (μ_1 = 100, μ_2 = 200, σ = 20, transition length = 100). (B) Accuracy on the entire synthetic data set. (C) Alanine dipeptide data set; MT = 10,000. (D) Langevin dynamics trajectory with two different sensitivity settings. (E) Representative short-transition trajectory from the synthetic data set (μ_1 = 100, μ_2 = 200, σ = 5, transition length = 20). (F, G) Synthetic trajectories with transition length = 100, and varying σ (5 and 20, respectively), with varying tuning parameter settings. Metastable segments are labeled in red, and transition segments are labeled in blue in all cases.

2) Change point placement performance: On the synthetic data set with known ground truth, which allowed for quantitative performance determination, we evaluated the change point placement of all tested algorithms by computing the sum of distances of all placed change points to their nearest ground-truth change point. BarT takes the performance crown here, with SIMPLE's oversegmentation becoming its undoing in the case of long transitions, and PELT struggling with short transitions, particularly in high-variance cases, where the test statistic improvement from placing change points in the transition regions (as opposed to within metastable segments) is not high enough. A visual overview of placement accuracy is shown in Figure 15.

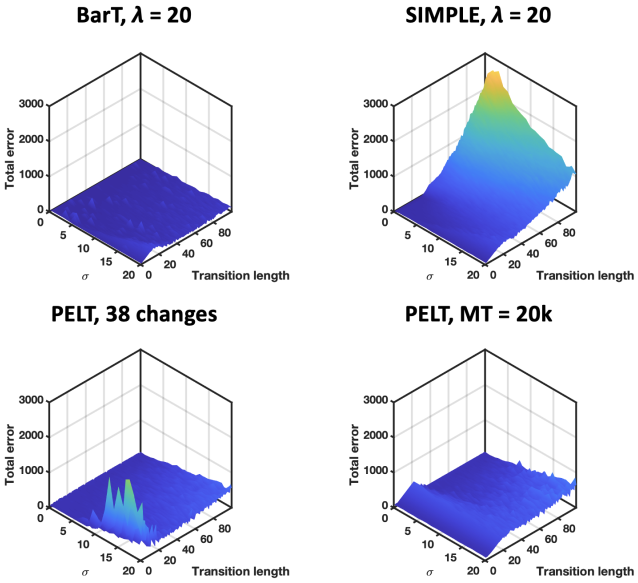

Fig. 15. Change point placement accuracy for all tested algorithms. The total error represents the sum of distances of all change points placed to their nearest ground-truth changes.