



**CHALMERS**  
UNIVERSITY OF TECHNOLOGY

## **Thermodynamic Interpolation: A Generative Approach to Molecular Thermodynamics and Kinetics**

Downloaded from: <https://research.chalmers.se>, 2025-04-04 09:30 UTC

Citation for the original published paper (version of record):

Moqvist, S., Chen, W., Schreiner, J. et al (2025). Thermodynamic Interpolation: A Generative Approach to Molecular Thermodynamics and Kinetics. *Journal of Chemical Theory and Computation*, 21(5): 2535-2545. <http://dx.doi.org/10.1021/acs.jctc.4c01557>

N.B. When citing this work, cite the original published paper.

# Thermodynamic Interpolation: A Generative Approach to Molecular Thermodynamics and Kinetics

Published as part of *Journal of Chemical Theory and Computation special issue "Markov State Modeling of Conformational Dynamics"*.

Selma Moqvist, Weilong Chen, Mathias Schreiner, Feliks Nüske, and Simon Olsson\*



Cite This: *J. Chem. Theory Comput.* 2025, 21, 2535–2545



Read Online

ACCESS |



Metrics & More

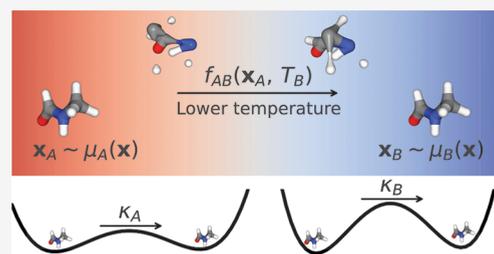


Article Recommendations



Supporting Information

**ABSTRACT:** Using normalizing flows and reweighting, Boltzmann generators enable equilibrium sampling from a Boltzmann distribution, defined by an energy function and thermodynamic state. In this work, we introduce thermodynamic interpolation (TI), which allows for generating sampling statistics in a temperature-controllable way. We introduce TI flavors that work directly in the ambient configurational space, mapping between different thermodynamic states or through a latent, normally distributed reference state. Our ambient-space approach allows for the specification of arbitrary target temperatures, ensuring generalizability within the temperature range of the training set and demonstrating the potential for extrapolation beyond it. We validate the effectiveness of TI on model systems that exhibit metastability and nontrivial temperature dependencies. Finally, we demonstrate how to combine TI-based sampling to estimate free energy differences through various free energy perturbation methods and provide corresponding approximated kinetic rates, estimated through generator extended dynamic mode decomposition (gEDMD).



## INTRODUCTION

Computing molecular properties and observables, such as free energies, is of great interest in numerous scientific and engineering applications. In statistical mechanics, we can express many of these observables directly through the partition function, or normalizing constant of statistical distribution over microscopic configurations of a molecular system.<sup>1</sup> Such a statistical distribution, or *ensemble* is defined by the macroscopic control variables, such as temperature, volume, pressure, and chemical potential, which are kept constant. The canonical Boltzmann distribution,  $\mu(\mathbf{x}) = \mathcal{Z}^{-1} \exp(-E(\mathbf{x})/kT)$  is for example characterized by constant temperature ( $T$ ), volume ( $V$ ), and particle number ( $N$ ).

Due to the high-dimensionality of molecular systems, direct computation of the partition function is intractable. Instead, we rely on simulation strategies such as Markov Chain Monte Carlo (MCMC) or Molecular Dynamics (MD) to draw samples from the distribution.<sup>2</sup> However, for molecular systems, the high-dimensional free energy landscapes lead to impractically long simulations, to ensure the generation of independent sampling statistics.

Numerous enhanced sampling methods are available, all aiming to accelerate sampling. These methods, modify the statistical ensemble to ensure faster traversal between free energy basins, or couple multiple thermodynamic ensembles *replicas*, or a combination of the two — albeit with the

constraint, that it is possible to reweigh the generated samples back to the correct ensemble.<sup>3,4</sup> Some influential examples include meta dynamics,<sup>5</sup> or conformational flooding,<sup>6</sup> replica-exchange and parallel tempering,<sup>7</sup> umbrella sampling,<sup>8</sup> Hémin et al. recently surveyed numerous other approaches.<sup>9</sup> Machine learning is having a dramatic impact on these strategies, in particular in helping identify collective variables,<sup>10–13</sup> which is lowering the need for manual trial-and-error optimization, and transformations which lower the number of replicas<sup>14</sup> needed to ensure effective simulation.

Current enhanced sampling methods allow us to compute stationary observables such as free energies. However, as they usually involve biasing the molecular dynamics, recovery of unbiased dynamics or kinetics, is only possible in certain situations.<sup>6,15,16</sup> Kinetic modeling using Markov state models (MSM),<sup>17–21</sup> Koopman operator approaches,<sup>22–24</sup> dynamic graphical models,<sup>25,26</sup> transition path sampling<sup>27–30</sup> and deep learning infused approaches,<sup>31–34</sup> take an alternative approach, leveraging either long unbiased simulations, or massively parallel short simulations collected using adaptive sampling

**Received:** November 16, 2024

**Revised:** February 13, 2025

**Accepted:** February 14, 2025

**Published:** February 24, 2025



strategies.<sup>35–38</sup> These approaches allow us to uncover unbiased dynamics and facilitate the calculation of unbiased stationary and dynamic observables, yet remain costly from a computational perspective.

Deep generative models enable the development of fundamentally new approaches to equilibrium sampling and sampling of stochastic dynamics.<sup>39–44</sup> An important example is Boltzmann generators (BG)<sup>45</sup> where an invertible deep neural network model, is trained to transform samples from a simple distribution, e.g., a normal distribution, to a complicated distribution, e.g., a Boltzmann distribution. In practice, this is often implemented using a *normalizing flow*.<sup>45–47</sup> Subsequent efforts to improve accuracy and transferability of BGs, leverage alternative neural networks, including diffusion models,<sup>48</sup> and continuous normalizing flows,<sup>49,50</sup> in a manner that account for molecular symmetries. A variation of this idea, called “Boltzmann Emulators” generate ensembles in a reduced configurational space, e.g., by parametrizing a generative model over torsion angles, and keeping bond-lengths and angles at idealized values, have also shown some success.<sup>48,51,52</sup>

The connections between statistical mechanics, and deep generative models have further fueled a zoo new methods for sampling. Many of these are either inspired by perturbative methods or perturbative in nature including thermodynamic maps,<sup>53</sup> which learn a coarse-grained multitemperature model from replica-exchange data. Other examples include mapping between cheap reference potentials and expensive quantum mechanical models,<sup>54</sup> computing free energies,<sup>55</sup> or decreasing the number of replicas in a replica exchange scheme.<sup>14</sup> More recently, a denoising diffusion model was used to perform a learned thermodynamic integration between in ideal gas and a Lennard-Jones liquid.<sup>56</sup> Finally, an early example proposed constructing normalizing flows that were steerable under temperature transformations,<sup>57</sup> allowing for some generalization across temperature.

In this work, we propose Thermodynamic Interpolation (TI) as a method for generating samples across multiple thermodynamic states, through a learned map between different Boltzmann distributions. We present two different TI methods **Ambient TI** and **Latent TI**. Ambient TI transforms between two thermodynamic states directly in the configurational space, while latent TI transform samples between thermodynamic ensembles through a normally distributed latent space distribution. We implement both TI approaches using new simulation-free training schemes developed for continuous normalizing flows,<sup>58</sup> and consequently, we can both transform samples between distributions and compute their change in log probabilities. Using temperature transformations as an example, we demonstrate that our ambient TI approach enables transformations between ensembles at different temperatures in configuration space. Our latent TI models similarly enable transformations between ensembles albeit through a shared latent space distribution, or reference state. Further, since the latent TI is implemented using a temperature conditioned BG we can generate samples at multiple different temperatures on-demand. We find both methods can be trained very efficiently with limited simulation data, and information is shared between thermodynamically similar ensembles. We evaluate our approach on a one-dimensional double-well model system and then scale it to MD simulations<sup>48</sup> of two molecules from the QM9 data set:<sup>59</sup> *N*-Methylformamide (*N*-Me) and 3-propan-2-ylhex-1-yne (3p2y1y). For all models, we achieve high sampling efficiency,

even outside the training data, allowing us to estimate equilibrium properties such as free energy differences and dynamic properties like kinetic transition rates at temperatures not encountered during training. We thus present a framework that enables flexible transformations between arbitrary thermodynamic states, along with access to corresponding probabilities. As a result, we believe our TI approach offers a robust and generalizable tool for accurately predicting thermodynamic and kinetic properties across a wide range of thermodynamic states.

## METHODS

**Estimation of Free Energies.** Free energy perturbation (FEP)<sup>60</sup> is a method to compute free energy difference  $\Delta F_{AB}$  between two thermodynamic states, *A* and *B*, through the identity

$$\mathbb{E}_{\mathbf{x}_A}[\exp(-(E_B(\mathbf{x}_A) - E_A(\mathbf{x}_A)))] = \exp(-\Delta F_{AB}) \quad (1)$$

where  $E_A$  and  $E_B$  are unit-less potential energy functions, and averaging is done for the Boltzmann distribution of state *A*,

$$\mu_A(\mathbf{x}) = \mathcal{Z}_A^{-1} \exp(-E_A(\mathbf{x})) \quad (2)$$

Unfortunately, this free energy estimator is inefficient if the overlap between states *A* and *B*, is low.

To alleviate this, Jarzynski introduced targeted FEP (TFEP)<sup>61</sup> which augments the FEP by introducing a map  $f_{AB}: \mathbf{x}_A \rightarrow \mathbf{x}_B$  which aims to transform the state *A* to another state closer to *B*. If such map can be found, and is differentiable and invertible, then the free energy difference between *A* and *B* can be estimated only given samples  $\mathbf{x}_A \sim \mu_A$  by computing the average

$$\mathbb{E}_{\mathbf{x}_A}[\exp(-\varphi(\mathbf{x}_A))] = \exp(-\Delta F^{(\text{TFEP})}) \quad (3)$$

where

$$\varphi(\mathbf{x}_A) \triangleq E_B(f_{AB}(\mathbf{x}_A)) - E_A(\mathbf{x}_A) - \log \det(J_{f_{AB}}(\mathbf{x}_A)) \quad (4)$$

While this approach enjoys more favorable convergence properties, determining the map  $f_{AB}$  for a given application is not trivial. With the emergence of deep generative models, several approaches have been proposed to learn such maps for a variety of interesting applications.<sup>14,53–55,62</sup>

An alternative strategy to free energy difference estimation is based on Boltzmann generators (BG).<sup>45</sup> BGs are implemented with a normalizing flow, which is trained to transform the normally distributed latent space  $\mathbf{z} \sim \mathcal{N}(0, \text{Id}) \triangleq \mu_Z$  into molecular conformations  $f_{ZA}^{(\theta)}(\mathbf{z}) \sim \rho_A \approx \mu_A$ . In their paper Noé and co-workers,<sup>45</sup> show that the free energy difference between two metastable states *A* and *B* can be computed using two independent BGs, transforming from a latent state *Z* one into *A* and the other into *B*, using the expression

$$\Delta F^{(\text{BG})} = \mathbb{E}_Z[\mathcal{L}_{\text{KL}}^{(\text{BG},B)}(\mathbf{z})] - \mathbb{E}_Z[\mathcal{L}_{\text{KL}}^{(\text{BG},A)}(\mathbf{z})] \quad (5)$$

where the BG loss function  $\mathcal{L}_{\text{KL}}^{(\text{BG},X)}$  equal to the free energy of state *X* up to a constant. Using the definition of BG Kullback–Leibler loss we can rewrite the estimator as

$$\begin{aligned} \Delta F^{(\text{BG})} &= \mathbb{E}_{\mathbf{z}}[E_B(f_{ZB}(\mathbf{z})) - E_A(f_{ZA}(\mathbf{z})) + \log|\det J_{f_{ZA}}(\mathbf{z})| \\ &\quad - \log|\det J_{f_{ZB}}(\mathbf{z})|] \\ &\triangleq \mathbb{E}_{\mathbf{z}}[\psi(\mathbf{z})] \end{aligned} \quad (6)$$

As we show in the [Supporting Information](#) (free energy perturbation methods), it follows from the invertibility of the BG map that  $\psi$  is equivalent to the function (eq 4). This further suggests that

$$\mathbb{E}_{\mathbf{z}}[\exp(-\psi(\mathbf{z}))] = \frac{Z_B}{Z_A} \triangleq \exp(-\Delta F^{(\text{TFEP})}) \quad (7)$$

As implied by eqs 6 and 7, it is not necessary to have an explicit map between the two thermodynamic states, but rather it is sufficient with a map that transforms some initial distribution into the two target states A and B separately. Similarly to previous work,<sup>54,63</sup> we can apply Jensen's inequality to obtain a relationship between the BG and TFEP estimators, given as

$$\Delta F^{(\text{BG})} \geq \Delta F^{(\text{TFEP})} \quad (8)$$

The inequality 8 can also be seen as an expression of the Donsker-Varadhan (DV) variational principle for importance sampling of exponential expectations.<sup>64,65</sup> Here, we consider changes of probability measure induced by the push-forward under the invertible transformation  $f_{ZB} \circ f_{ZA}^{-1}$ . The DV principle states that 6 is an upper bound for the exponential expectation 7. In addition, the variance of 6 is decreasing with increasing right-hand side, leading to a (theoretical) one-shot estimator if equality is attained in 8. As such, the TFEP-estimator serves as a lower bound on the BG-estimator. A more in-depth derivation and discussion of this can be found in the [Supporting Information](#) (free energy perturbation methods).

**Continuous Normalizing Flows.** Normalizing flows is a class of deep generative models, where we learn a map  $f^{(\theta)}: \Omega_0 \rightarrow \Omega_1$ , parametrized by  $\theta$  where  $\Omega_0, \Omega_1 \subset \mathbb{R}^d$  and  $d \in \mathbb{N}$ , that transforms an initial distribution  $\rho_0: \Omega_0 \rightarrow \mathbb{R}_+$  into a target distribution  $\rho_1: \Omega_1 \rightarrow \mathbb{R}_+$ . To obtain sample probabilities, the map must be both smooth and invertible, a *diffeomorphism*. In general, the map can be between any two distributions, for example Boltzmann distributions at different temperatures.

A diffeomorphic map can be constructed as either a composition of several smooth and invertible partial transformations<sup>66</sup> or as the solution to an initial value problem, known as a continuous normalizing flow (CNF).<sup>67</sup> In other words, we learn the velocity field  $\mathbf{b}^{(\theta)}: [0, 1] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  such that ordinary differential equation (ODE)

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{b}^{(\theta)}(t, \mathbf{x}(t)), \quad \mathbf{x}(0) = \mathbf{x}_0 \quad (9)$$

with initial condition  $\mathbf{x}_0 \sim \rho_0$ , approximates  $\rho_1$  when integrated in time from 0 to 1. We call the resulting map  $f_{01}^{(\theta)}$ . The ODE coupled with the initial distribution  $\rho_0$  gives rise to a time-dependent probability density  $\rho(t, \mathbf{x}(t))$  described by the continuity equation,

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{b}^{(\theta)}) = 0, \quad \rho(0, \mathbf{x}(0)) = \rho_0(\mathbf{x}_0) \quad (10)$$

We can solve both equations jointly to get both the transformed sample  $f_{01}^{(\theta)}(\mathbf{x}_0) = \mathbf{x}_1 \sim \rho_1$  and its corresponding change in logarithmized probability or Jacobian determinant  $\log|\det J_{f_{01}^{(\theta)}}(\mathbf{x}_0)|$ , where the latter is obtained by integrating

$$\log \rho(t, \mathbf{x}(t)) = \log \rho(0, \mathbf{x}(0)) - \frac{\int_0^t \nabla \cdot \mathbf{b}^{(\theta)}(s, \mathbf{x}(s)) ds}{\log|\det J_{f_{01}^{(\theta)}}(\mathbf{x}_0)|} \quad (11)$$

using an off-the-shelf numerical solver.

Here we use the Stochastic Interpolant (SI) framework to learn the velocity field  $\mathbf{b}^{(\theta)}$ .<sup>58</sup> By interpolating between samples from the initial and target distributions in a stochastic manner, one can define a stochastic process

$$\mathbf{x}(t) = I(t, \mathbf{x}_0, \mathbf{x}_1) + \gamma(t)\mathbf{z}, \quad t \in [0, 1], \quad \mathbf{z} \sim \mathcal{N}(0, \text{Id}) \quad (12)$$

where the interpolant  $\mathbf{x}(t)$  satisfies  $\mathbf{x}(0) = \mathbf{x}_0$  and  $\mathbf{x}(1) = \mathbf{x}_1$ . The interpolant in eq 12 defines a path for moving samples from  $\rho_0$  to  $\rho_1$  in finite time, with the guarantee that at time  $t = 0$  the sample is distributed according to  $\rho_0$ , and at time  $t = 1$  the sample is distributed according to  $\rho_1$ . However, at intermediate times  $t \in (0, 1)$  the interpolant  $\mathbf{x}(t)$  is completely characterized by the functions  $I$  and  $\gamma$ . Since these can be chosen in any way that respects the SI boundary conditions,<sup>58</sup> the velocity of the interpolant,  $\dot{\mathbf{x}}(t)$ , is easily obtained as long as one uses appropriate choices of  $I$  and  $\gamma$ . A vector field  $\mathbf{b}$  can be defined as the expected velocity of the interpolant (eq 12)

$$\mathbf{b}(t, \mathbf{x}(t)) = \mathbb{E}[\dot{\mathbf{x}}(t)|\mathbf{x}(t)] \quad (13)$$

where we note that  $\mathbf{b}$  describes the velocity of the individual samples in eq 9 and the density in eq 10. In order to learn a parametrized version,  $\mathbf{b}^{(\theta)}$ , of the velocity  $\mathbf{b}$ , we minimize the regression-based objective

$$\begin{aligned} L[\mathbf{b}^{(\theta)}] &= \mathbb{E}_{\mathbf{x}_0, \mathbf{x}_1, t} \left[ \frac{1}{2} \|\mathbf{b}^{(\theta)}(t, \mathbf{x}(t))\|^2 \right. \\ &\quad \left. - (\partial_t I(t, \mathbf{x}_0, \mathbf{x}_1) + \dot{\gamma}(t)\mathbf{z}) \cdot \mathbf{b}^{(\theta)}(t, \mathbf{x}(t)) \right] \end{aligned} \quad (14)$$

where times  $t$  are drawn as  $t \sim \mathcal{U}[0, 1]$  and  $\partial_t$  denotes a time-derivative.

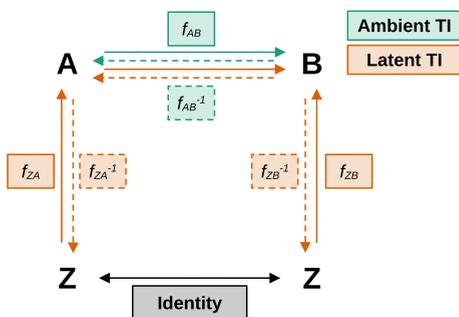
**SE(3)-Equivariant Message Passing Neural Networks.** The energy of a closed molecular system is symmetric under E(3) group action. These symmetries imply that models of the potential energy and the corresponding Boltzmann distribution should be *invariant* to global 3D rotations, translations, and inversion.<sup>49,68</sup> Consequently, when we learn functions to approximate potential energies or Boltzmann distributions from data we can limit our hypothesis space to functions that satisfy these symmetries without incurring any approximation error.<sup>69</sup> A consequence of this is improved data-efficiency and better generalization beyond the training set.

More formally, we call a function  $f$  G-invariant if  $f(T_g \mathbf{x}) = f(\mathbf{x})$  and G-equivariant if  $S_g f(\mathbf{x}) = f(T_g \mathbf{x})$ , where  $S_g$  and  $T_g$  are linear representations of the group element  $g \in G$ . We can learn a G-invariant probability density model by combining a G-invariant distribution  $\rho_0$  with a H-equivariant velocity field,<sup>49</sup> where H is a subgroup of G. In other words, applying equivariant perturbations to an invariant probability density will leave it invariant. Practically, the neural network parametrizing our velocity field  $\mathbf{b}^{(\theta)}$  should be constructed in a way that is equivariant with respect to the symmetries displayed by a molecular system.

The energy of a molecule generally exhibits reflection and permutation symmetry. However, as we here work with classical MD data, where chirality is usually constrained we follow previous work and instead of the E(3)-group we thus

consider the SE(3)-group, which is comprised of only rotations and translations. A neural network architecture fulfilling this requirement is the ChiroPaiNN architecture, which we employ to build SE(3)-equivariant flow models.<sup>70</sup> While the architecture we use here is equivariant under the permutation group, we do not exploit it in our experiments on molecules.

**Thermodynamic Interpolation.** In this work, we present thermodynamic interpolation (TI), an approach which builds on the idea of targeted free energy perturbation<sup>61</sup> where a diffeomorphic map is used to transform between different thermodynamic states with the aim to compute free energy differences. Here, we learn such maps using two different approaches: latent and ambient TI (Figure 1)



**Figure 1.** Illustration of the two thermodynamic interpolation approaches. In the ambient case (green) one begins with Boltzmann distributed samples  $\mathbf{x}_A \sim \mu_A$  at the thermodynamic state A. These are transformed into samples  $f_{AB}^{(\theta)}(\mathbf{x}_A) \sim \rho_B$  drawn from the surrogate distribution at state B, using the learned aTI map  $f_{AB}^{(\theta)}$ . Finally, when estimating observables, samples are weighed using importance weights  $w^a$  so that they correspond to the thermodynamic state B. This is done to correct for the bias introduced by the surrogate. One can also first learn a lTI surrogate model to generate approximately Boltzmann distributed samples  $f_{ZA}^{(\theta)}(\mathbf{z}) \sim \rho_A$ , compute the corresponding weights  $w^l$  to weigh  $f_{ZA}^{(\theta)}(\mathbf{z})$  into  $\mathbf{x}_A$ , and then transform  $\mathbf{x}_A$  into  $\mathbf{x}_B \sim \mu_B$  with aTI (orange) and the corresponding weights  $w^a$ .

**Latent TI.** transforms samples between thermodynamic states through a latent space equipped with a normal distribution. We implement a latent TI (lTI) model using a temperature conditioned BG. Due to modeling errors, the BG will sample from an approximation,  $\rho_{A'}$ , of the true Boltzmann distribution,  $\mu_{A'}$ . To compute unbiased estimates we therefore need to weigh samples by their importance weights<sup>45</sup>

$$w_{\text{BG}}(\mathbf{z}) = \exp \left( -\frac{1}{kT_A} E(f_{ZA}^{(\theta)}(\mathbf{z})) - \log \mu_Z(\mathbf{z}) + \log \det J_{f_{ZA}^{(\theta)}}(\mathbf{z}) \right) \quad (15)$$

In lTI, we make the map temperature-conditioned  $f_{ZA}^{(\theta)}(\cdot; T_A)$ , in a similar spirit to recent work aiming to compute phase diagrams using normalizing flows.<sup>71</sup>

To train our temperature conditioned map we use make of the ‘one-sided interpolant’

$$I(t, \mathbf{z}, \mathbf{x}_A) = (1-t)\mathbf{z} + t\mathbf{x}_A = \mathbf{x}(t) \quad (16)$$

linearly interpolating normally distributed noise  $\mathbf{z} \sim \mu_Z$  with Boltzmann distributed configurations  $\mathbf{x}_A \sim \mu_A$  at temperature

$T_A$ . By combining this interpolant, with the loss (eq 14), which further averages is over the thermodynamic states A, in our case, temperatures,

$$L[\mathbf{b}^{(\theta)}] = \mathbb{E}_{t, \mathbf{z}, \mathbf{x}_A, A} \left[ \frac{1}{2} \left| \mathbf{b}^{(\theta)}(t, \mathbf{x}(t); T_A) \right|^2 - \partial_t I(t, \mathbf{z}, \mathbf{x}_A) \cdot \mathbf{b}^{(\theta)}(t, \mathbf{x}(t); T_A) \right] \quad (17)$$

we can learn a parametrized version  $f_{ZA}^{(\theta)}$  of the lTI map  $f_{ZA}$  which is conditioned on temperature. Then we can generate samples at an arbitrary temperature  $T_A$  by sampling the latent space distribution and transforming them with the learned forward map  $f_{ZA}^{(\theta)}$ , and compute unbiased observables by weighing the samples by their importance weights (eq 15).

**Ambient TI.** In many cases, two thermodynamic ensembles may be more similar to each other than to the reference state encoded in a latent space. Consequently, we propose ambient TI (aTI) as an approach to learn a direct map between thermodynamic states  $f_{AB}$  in the configuration space. Compared to lTI, learning a direct map will avoid two step transformations. For example computing free energy changes between two thermodynamic ensembles using lTI would require sampling positions and their corresponding probabilities at two temperatures. With a direct map we could sample at one temperature and directly compute the free energy from the change in probability associated of the aTI map using these samples. Additionally, we can also combine an aTI map with a Parallel Tempering protocol, akin to previous work,<sup>14</sup> to avoid having to simulate multiple replicas to ensure efficient sampling. As illustrated in Figure 1, high-temperature (e.g., A) samples could then be mapped directly into lower temperatures (e.g., B) using aTI.

To train aTI maps  $f_{AB}$  between thermodynamic states A and B we make use of the ‘two-sided interpolant’<sup>58</sup>

$$I(t, \mathbf{x}_A, \mathbf{x}_B, \mathbf{z}) = (1-t)\mathbf{x}_A + t\mathbf{x}_B + \gamma(t)\mathbf{z} = \mathbf{x}(t) \quad (18)$$

where  $\mathbf{z} \sim \mu_Z$ ,  $\mathbf{x}_A \sim \mu_A$  and  $\mathbf{x}_B \sim \mu_B$ . Here,  $\mu_A$  and  $\mu_B$  represent Boltzmann distributions with identical energy functions  $E$  but at different temperatures  $T_A$  and  $T_B$ , corresponding to states A and B respectively. We detail the choice of  $\gamma$ , in the Supporting Information (hyperparameters for ADW system and hyperparameters for molecular systems). As for the lTI, we can combine this interpolant, with the loss (eq 14), where we now take the expectation with respect to source and target state temperatures, samples drawn at both temperatures ( $\mathbf{x}_A$  and  $\mathbf{x}_B$ ),  $t$  and  $\mathbf{z}$ ,

$$L[\mathbf{b}^{(\theta)}] = \mathbb{E}_{t, \mathbf{z}, \mathbf{x}_A, \mathbf{x}_B, A, B} \left[ \frac{1}{2} \left| \mathbf{b}^{(\theta)}(t, \mathbf{x}(t); T_A, T_B) \right|^2 - \partial_t I(t, \mathbf{x}_A, \mathbf{x}_B, \mathbf{z}) \cdot \mathbf{b}^{(\theta)}(t, \mathbf{x}(t); T_A, T_B) \right] \quad (19)$$

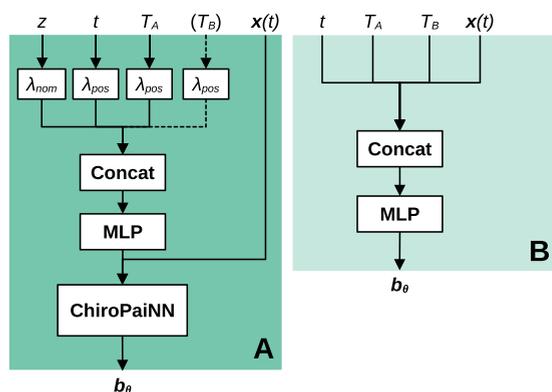
Note that our model of the velocity field  $\mathbf{b}^{(\theta)}$  now depends on source and target temperatures.

When applying aTI, we can provide samples to the initial state in several ways. For example, we can run conventional, or enhanced sampling, MD simulations at  $T_A$ , or we can sample a surrogate model which generates sample from  $\rho_A \approx \mu_A$ . By applying the learned map we can then transform the samples to samples from the surrogate  $\rho_B$ , and we can recover unbiased estimates of observables by computing the importance weights

$$w^a(\mathbf{x}_A, T_A, T_B) = \exp\left(-\frac{1}{kT_A}E(\mathbf{x}_A) - \frac{1}{kT_B}E(f_{AB}^{(\theta)}(\mathbf{x}_A)) + \log\det J_{f_{AB}^{(\theta)}}(\mathbf{x}_A)\right) \quad (20)$$

When initial conditions are generated with a surrogate, we further need to account for the approximate nature of the surrogate, e.g., using the important weights if its a (temperature conditioned) BG  $w_{BG}$ . We then use eq 20 to compute the aTI weights  $w^a$ .

**Implementation Details.** The learned velocity fields  $\mathbf{b}^{(\theta)}$  depend on a range of inputs. We illustrate the information flow in Figure 2. For the molecular systems, we embed the



**Figure 2.** Thermodynamic interpolant neural architecture. (A) Latent and ambient TI for high-dimensional molecular-type systems. In the latent case we are only interested in sampling at a single temperature  $T_A$  and do not input the second temperature  $T_B$  to the cPaiNN model. In the ambient case, both temperatures  $T_A$  and  $T_B$  are included, along with atom numbers  $z$ . (B) Ambient TI for low-dimensional systems. In the lower-dimensional case we do not encode the temperatures or times into positional embeddings  $\lambda_{pos}$ , and a simple MLP can be used instead of the ChiroPaiNN model.

temperature using the positional embeddings  $\lambda_{pos}(T) = \{\lambda_{pos}^n(T)\}_{n=0}^N$  of dimension  $N$ , where

$$\lambda_{pos}^n(T) = \begin{cases} \cos\left(\left(1 + \frac{n}{2}\right)\frac{\pi\hat{T}}{l_0}\right) & \text{for even } n \\ \sin\left(\left(1 + \frac{n-1}{2}\right)\frac{\pi\hat{T}}{l_0}\right) & \text{for odd } n \end{cases} \quad (21)$$

for normalized temperatures  $\hat{T} = \frac{T - \bar{T}}{T_{\max} - T_{\min}}$  and where  $l_0$  is a hyperparameter. By changing  $l_0$ , embeddings of different temperatures can be made more or less similar to each other. We embed the interpolation time  $t$  and atom numbers  $z$  are positional and nominal embeddings, following previous work.<sup>70</sup> For the asymmetric double-well system we use the temperatures directly. Specific details on hyperparameter for experiments are available in the [Supporting Information](#) (Hyperparameters for ADW system and hyperparameters for molecular systems)

**Generator Extended Dynamic Mode Decomposition.** Modeling slow processes such as protein folding and ligand-binding/unbinding to a target protein is challenging as it relies on extensive unbiased MD simulation data. However, recent

advances allow for data-driven estimation of the Koopman operator through extended dynamic mode decomposition (EDMD),<sup>72,73</sup> which can be extended to the Koopman generator by using generator EDMD.<sup>24</sup> With statistical estimates of Koopman operators or generators, for example expressed in some feature space, we analyze slow processes through their eigenvectors and eigenvalues, akin to MSMs.

For lag time  $t \geq 0$ , the Koopman operator maps  $\phi$  to the conditional expectation

$$\mathcal{K}^t \phi(x) = \mathbb{E}[\phi(\mathbf{x}_t) | \mathbf{x}_0 = x] \quad (22)$$

where  $\mathbf{x}_t$  is a (stochastic) dynamical system in the space  $\mathbb{X} = \mathbb{R}^d$ , and  $\phi$  is a function of the state space. By taking the time derivative of  $\mathcal{K}^t \phi$  at  $t = 0$ , we obtain the Koopman generator

$$\mathcal{L}\phi = \lim_{\tau \rightarrow 0} \frac{1}{\tau} [\mathcal{K}^\tau - Id]\phi \quad (23)$$

If  $\mathbf{x}_t$  follows a SDE,  $\mathcal{L}$  becomes a second-order linear differential operator. Specifically, for overdamped Langevin dynamics with potential energy  $E$  and at temperature  $T$ ,

$$d\mathbf{x}_t = -\nabla E(\mathbf{x}_t) dt + \sqrt{2k_B T} dW_t \quad (24)$$

the corresponding generator is

$$\mathcal{L}\phi(x) = -\nabla E \cdot \nabla \phi + kT \Delta \phi \quad (25)$$

where  $\nabla \phi$  is the gradient and  $\Delta \phi$  is the Laplacian.

In essence, gEDMD uses statistical samples from the equilibrium distribution, or invariant measure, associated with some stochastic dynamics, to estimate an infinitesimal time-continuous generator of the dynamics.

Given a basis set  $\{\phi_i\}_{i=1}^n$  and data  $\{\mathbf{x}_i\}_{i=1}^m$  sampled from a probability distribution, for example the Boltzmann distribution  $\mu$ , the finite-dimensional estimate for the generator  $\mathcal{L}$  is represented as a matrix  $\mathbf{L}$ , which can be calculated from the solution to a system of linear equations

$$\mathbf{L} = \mathbf{G}^{-1} \mathbf{A} \quad (26)$$

As the sample size  $m \rightarrow \infty$ , the matrices  $\mathbf{G}$  and  $\mathbf{A}$  can be expressed as expectations:

$$\begin{aligned} \mathbf{G} &\rightarrow \mathbb{E}[\phi(\mathbf{x}) \otimes \phi(\mathbf{x})], \\ \mathbf{A} &\rightarrow \mathbb{E}[\phi(\mathbf{x}) \otimes \mathcal{L}\phi(\mathbf{x})] \end{aligned} \quad (27)$$

Here,  $\phi(\mathbf{x}) = [\phi_1(\mathbf{x}), \dots, \phi_n(\mathbf{x})]$  are the vectors of basis functions evaluated at  $\mathbf{x}$ , and  $\mathcal{L}\phi(\mathbf{x})$  is the corresponding generator applied to the basis functions. Using a finite sample of data, we approximate  $\mathbf{G}$  and  $\mathbf{A}$  with the empirical averages

$$\begin{aligned} \mathbf{G} &= \frac{1}{m} \sum_{i=1}^m \phi(\mathbf{x}_i) \otimes \phi(\mathbf{x}_i), \text{ and} \\ \mathbf{A} &= \frac{1}{m} \sum_{i=1}^m \phi(\mathbf{x}_i) \otimes \mathcal{L}\phi(\mathbf{x}_i) \end{aligned} \quad (28)$$

For reversible overdamped Langevin dynamics, the expression for  $\mathbf{A}$  simplifies to

$$\mathbf{A}^* = -kT \mathbb{E}_{\mathbf{x} \sim \mu} [\nabla \phi(\mathbf{x}) \cdot \nabla \phi(\mathbf{x})] \quad (29)$$

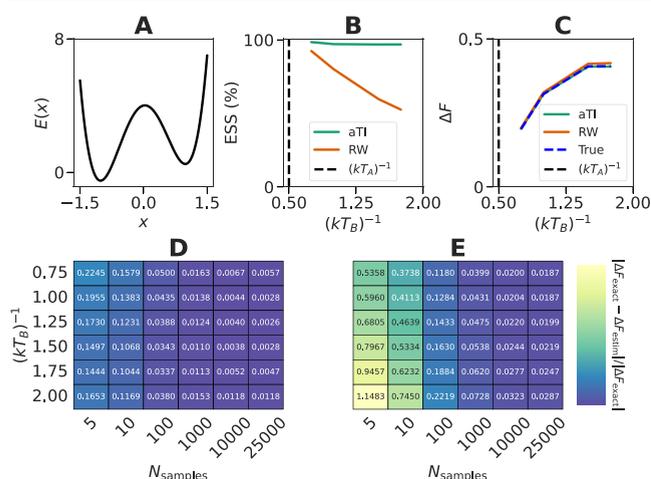
where  $\nabla \phi(\mathbf{x})$  is the Jacobian matrix of the basis set  $\phi$ . The reversible estimator 29 retains the symmetry and positive-definiteness of the generator matrix  $\mathbf{A}$ . The key implication is that we can estimate the generator  $\mathcal{L}$  directly from equilibrium

samples drawn from the Boltzmann distribution  $\mu$  corresponding to the stationary state of the generator. The eigenvalues of this generator correspond to the kinetic transition rates between metastable states, providing a bound on the system's slow kinetics. This approach allows us to approximate slow processes and rare transitions without requiring time-correlated trajectories, making it an efficient method for studying molecular kinetics based purely on equilibrium data.

## RESULTS

We evaluate the performance of the ambient and latent TI methods by applying them to study three different systems: a one-dimensional asymmetric double-well potential (see the Supporting Information, ADW data set generation, for data set details) and MD simulation data<sup>48</sup> of two molecules from the QM9 data set<sup>59</sup>: *N*-Methylformamide (*N*-Me) and 3-propan-2-ylhex-1-yne (3p2y1y). We use the learned maps to generate low-temperature conformational ensembles from high-temperature ensembles, compute free energy differences, and estimate the temperature dependence of kinetic exchange rates. Specific hyperparameter choices for all our experiments can be found in the Supporting Information (hyperparameters for ADW system and hyperparameters for molecular systems).

**Generating Low Temperature Samples from High-Temperature Ensembles.** We perform the first evaluation on a 1D asymmetric double well system (Figure 3A). Here we train an aTI model on MD generated samples at temperatures  $(kT_{\text{train}})^{-1} \in \{0.5, 1.25, 2.0\}$  and transform from  $(kT_A)^{-1} = 0.5$  into different target temperatures not seen by the model during training. We achieve high sampling efficiency, measured through the Kish effective sample size (ESS)<sup>74</sup> (Figure 3B).



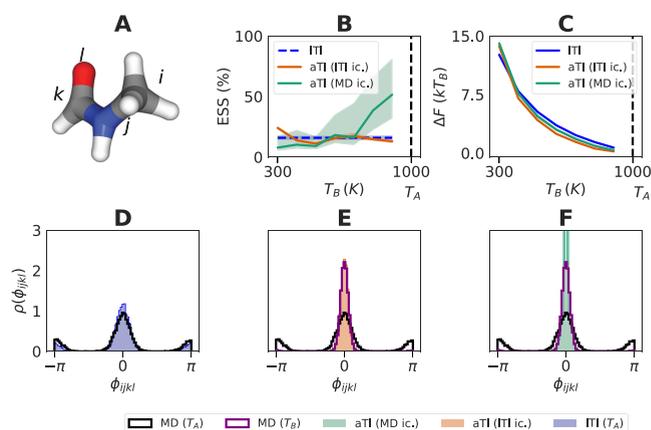
**Figure 3.** Results for asymmetric double well potential. (A) The energy landscape of the 1D asymmetric double well model system. (B) Effective sample sizes (ESS) for the aTI model compared to direct reweighing. Here, the free energy was estimated using  $\Delta F^{(\text{TPEP})}$ . (C) Estimated differences in Helmholtz free energy  $\Delta F$  for the ambient TI (aTI) model and using direct reweighing compared to true reference values. (D, E) Heatmaps of relative errors in  $\Delta F$ , for the aTI model and direct reweighing baseline respectively, as a function of the number of samples used in the estimator and the transformation target  $(kT_B)^{-1}$ . In (B–E), all transformations are made from  $(kT_A)^{-1} = 0.5$ , where the aTI model was trained on data at  $(kT_{\text{train}})^{-1} \in \{0.5, 1.25, 2.0\}$ . (A–E) To avoid high variance in the free energy estimators we made use of a filtering strategy discussed further in the Supporting Information (IQR-filtering of outliers).

The ESS estimates the effective number of statistically independent samples, as

$$n_{\text{eff}} = \frac{(\sum_i w_i)^2}{\sum_i w_i^2} \quad (30)$$

where  $w_i$  is the importance weight of sample  $x_i$  generated from a surrogate.

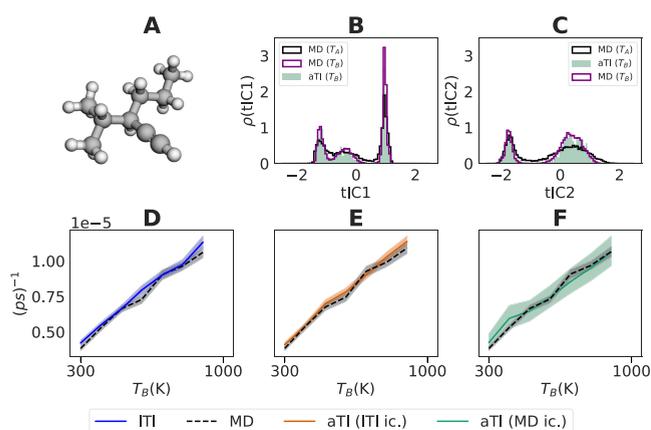
We extend this approach to the molecular systems *N*-Me and 3p2y1y, training molecule-specific aTI models on the replica-exchange molecular dynamics from two molecules from the MDQM9 data set.<sup>48</sup> At high temperatures, *N*-Me displays the multimodal distribution of a torsion (visualized in Figure 4A), which collapses into a unimodal distribution at low



**Figure 4.** *N*-Me ensemble, effective sample sizes, and free energies. (A) A visualization of the *N*-Me molecule with labels for the atoms  $i, j, k, l$  which form the torsion angle  $\phi_{ijkl}$ . (B) The effective sample size (ESS) plotted against the target temperature  $T_B$ . (C) Estimated differences in Helmholtz free energy  $\Delta F$  between temperatures  $T_A = 1000$  K and  $T_B \in \{300, 400, 500, 600, 700, 800, 900$  K} for the latent TI (ITI), and ambient TI (aTI) applied to ITI and MD-simulated initial conditions. Here, the free energy was estimated using  $\Delta F^{(\text{TPEP})}$ . (D–F) Marginal histograms of the torsion angle between the four atoms  $i, j, k$  and  $l$ , depicted in the small molecule. (D) Marginal histograms of torsions corresponding to ITI samples and MD data at  $T_A = 1000$  K. (E) Results of applying aTI to initial conditions generated through ITI, so that the system temperature is lowered from 1000 to 300 K, compared to reference MD simulated data. (F) Results of applying aTI to MD simulated initial conditions, so that the system temperature is lowered from 1000K to 300 K, compared to reference MD simulated data. (B–E) To avoid high variance on the free energy estimators we made use of a filtering strategy discussed further in the Supporting Information (IQR-filtering of outliers).

temperatures (Figure 4D–F). For 3p2y1y, two metastable states in the high temperature ensemble split into three distinct states transitioning from high to low temperature ensembles, as visualized by time-lagged independent components<sup>75–77</sup> (TICA) (Figure 5) (see details in the Supporting Information, scaling to larger systems).

Using our aTI models we generate low-temperature ensembles, starting from high-temperature samples. We generate these low-temperature ensembles in two ways: first, we transform samples generated with conventional MD simulations at high temperature. Second, we generate samples using a BG-type surrogate of the high-temperature ensemble,  $T_A = 1000$  K, here a ITI model. We compare these results against an unbiased low-temperature MD simulations (Figure



**Figure 5.** 3p2y1y ensemble and kinetics without reweighing. (A) A visualization of the 3p2y1y molecule. (B) The aTI torsion angles projected onto the first TICA ( $\tau = 2$  ps) dimension tIC1, plotted in a histogram. Along with the torsion angles, we also show reference MD values at the initial and target temperatures  $T_A$  and  $T_B$ . (C) The aTI torsion angles projected onto the second TICA dimension tIC2, plotted in a histogram. Along with the torsion angles, we also show reference MD values at the initial and target temperatures  $T_A$  and  $T_B$ . (D–F) Kinetic rate results estimated using gEDMD method for the larger molecule system, plotted against different target temperatures  $T_B \in \{300, 400, 500, 600, 700, 800, 900, 1000\text{ K}\}$  for the ITI, and the aTI approaches, compared with MD samples.

4D–F). To test aTIs ability to generalize beyond seen data, we train seven aTI models, such that each one is “blind” to the target temperature  $T_B$  during training, i.e.,  $T_{\text{train}} \in \{300, 400, 500, 600, 700, 800, 900, 1000\text{ K}\} \setminus T_B$ . In this way, we either have an interpolation or an extrapolation at test-time. We find that aTI models accurately generalizes a transformation from the initial Boltzmann distribution to a different thermodynamic state characterized by the target temperature  $T_B$ , indicated by the match of the histograms along the torsion angles Figure 4D–F) and high Kish<sup>74</sup> ESS (Figure 4B). We note that these ESS are lower for the higher dimensional molecular system compared to the 1D model potential, due to the increased complexity in the map, and improving these results are expected in line with the continued improvement in training strategies. Scaling up to the larger molecule 3p2y1y, we qualitatively find very similar results with distributions in the reduced TICA coordinates closely matching the reference simulations (Figure 5B,C).

Finally, we find that training a single model that learns using multiple temperatures, provide much better predictions across the temperature range than specialized ITI and aTI models (Supporting Information, single vs multitemperature training, Figure S3). This observation suggests that our TI models learn to share information between different thermodynamic ensembles, and thereby be applicable in cases with only limited data akin to what has been reported for multiensemble Markov models.<sup>78</sup>

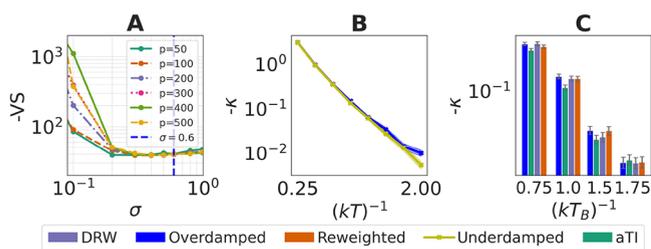
**Predicting Free Energies Changes upon Temperature Change.** Since CNFs allow for exact evaluation of changes in sample probability, we can use any of the free energy estimators discussed above to estimate the free energy difference between states A and B. Using the obtained low-temperature samples and their corresponding changes in probability, we estimate free energy differences between the thermodynamic state at the initial and target temperatures with the TFEP estimator  $\Delta F^{(\text{TFEP})}$ .

For the ADW system, we compare our free energy estimates against an estimate based on numerical integration of the Boltzmann distributions which acts as a highly accurate benchmark. We further compare against a benchmark where we directly reweigh samples from the reference state A to B using importance sampling. Both aTI and RW accurately reproduce the numerical benchmark (Figure 3B), however, even for this simple system the ESS drops rapidly with the difference in  $T_A$  and  $T_B$ , whereas the aTI based approach maintains a near perfect sample efficiency (Figure 3C).

To illustrate the practical impact of the ESS on free-energy estimates, we compute the relative errors in free energy as a function of samples from the reference state A for aTI (Figure 3D) and direct reweighing (Figure 3E). Unsurprisingly, we find that we can get highly accurate estimates with as little as five samples using aTI whereas comparable errors would need 20-fold more samples in the direct reweighing case. These results further underline, the potential of TI models as a data-efficient way to learn maps between thermodynamic states and compute free energy differences. Moving on to the molecule N-Me we get consistent estimates of the free energy differences using ITI and aTI (Figure 4C). While the estimates of the aTI estimator using samples from a surrogate (ITI) and MD as initial condition give us comparable  $\Delta F$  estimates, we note that the surrogate appears to limited the ESS (Figure 4B). For a comparison of the two free energy estimators  $\Delta F^{(\text{TFEP})}$  and  $\Delta F^{(\text{BG})}$ , see the Supporting Information (free energy perturbation methods, Figure S1), where we empirically find the bound (eq 7) to hold, with a gap suggesting that the learned aTI and ITI maps are not perfect, in line with the lower ESS (Figure 4B).

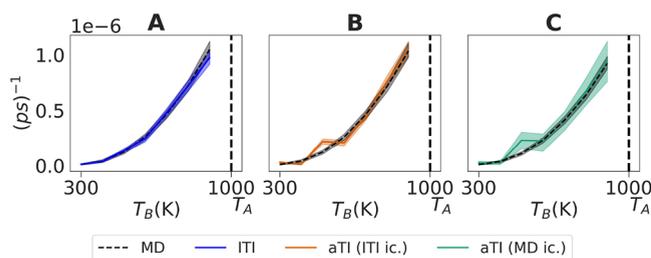
**Generator Estimation To Approximate Molecular Kinetics across Temperatures.** Next, we leverage our ability to efficiently generate unbiased statistics across a temperature range to show how kinetics change across a temperature range. Following previous work, we use a kernel-based gEDMD estimator<sup>79</sup> with Random Fourier Features (RFFs).<sup>80,81</sup> We are particularly interested in characterizing the temperature dependence of slow processes, corresponding to the smallest eigenvalues of the generator, which correspond directly to relaxation rates associated with exchange between metastable states. The estimation of the gEDMD models with RFFs requires optimizing hyperparameters, including the bandwidth and number of features, as illustrated in (Figure 6A). Details of this model selection process are provided in the Supporting Information (gEDMD model selection).

For ADW, we focus on the slowest rate, which is associated with exchange between the two major states (Figure 3A). Using gEDMD we compute these rates across four temperatures comparing aTI samples with and without reweighing using importance weights, to direct reweighing and overdamped Langevin simulations. Broadly, the predicted rates all align with reweighted samples yielding most closely following the overdamped reference (Figure 6B,C). Since using overdamped Langevin (Brownian) dynamics is uncommon on molecular applications as assumed with the estimator of the generator, we provide rates exacted from MSMs trained on simulations conducted with overdamped and underdamped simulations illustrating that rates extracted from overdamped simulations bound the corresponding rates from underdamped simulations from below (Supporting Information, MSM for different systems). For the molecular systems, we compute kinetic rates in a similar fashion: assuming the generator is



**Figure 6.** gEDMD analysis of asymmetric double well system. (A) VAMP scores<sup>22</sup> as a function of the kernel bandwidth  $\sigma$  for different numbers of Fourier features  $p$ . We select  $p = 50$  and  $\sigma = 0.6$  as the model parameters. (B) Kinetic rates calculated under the Brownian assumption using gEDMD methods for both overdamped and underdamped MD samples. (C) Kinetic rates estimated using gEDMD for MD samples, aTI predictions, and reweighted results. The aTI model was trained on data at  $(kT_{\text{train}})^{-1} \in \{0.5, 1.25, 2.0\}$ . Transformations are made from  $(kT_A)^{-1} = 0.5$  to  $(kT_B)^{-1} \in \{0.75, 1.0, 1.5, 1.75\}$ .

Brownian and using a torsion angle as features resolving the main meta-stabilities of the systems. We compare the ITI, and aTI applied to the ITI and MD initial conditions, illustrating that the predicted temperature dependence of the kinetic rates are consistent with reference values computed from MD samples, both for the N-Me molecule (Figure 7C) and the



**Figure 7.** Kinetic rate results estimated using gEDMD for N-Me, plotted against different target temperatures  $T_B \in \{300, 400, 500, 600, 700, 800, 900 \text{ K}\}$  for (A) the latent TI (ITI), and the two ambient TI (aTI) approaches with initial conditions generated using ITI (B) and MD (C). We show rates computed using MD samples with a dashed line. The reference temperature  $T_A$  is shown with a vertical dashed line.

3p2y1y molecule (Figure 5D–F). Strikingly, we observed that although our model imperfectly learns the map between thermodynamic states the predicted rates do not differ dramatically if we reweigh samples to the exact target (Figures 5D–F and 6C), suggesting that gEDMD is robust to some modeling bias introduced by generative surrogates.

## CONCLUSIONS

In this work, we have introduced TI as a flexible approach to equilibrium sampling of Boltzmann distributions across different thermodynamic states, and as a practical approach for free energy estimation through TFEP and for analyzing the kinetic dependence of thermodynamic transformations. We present two different instances of TI, one a directly mapping between states in the configurational space — ambient TI — and an alternative approach mapping between thermodynamic states through a latent reference state — latent TI. The two approaches allow end-users to approach free energy estimation through two routes, ITI being more flexible whereas aTI, being more computationally affordable. The first ITI, allows for either

reweighing of sampling statistics at a source state to a target state, or the direct generation of samples at a target state. The second, aTI, allows for the direct transformation and reweighing of sampling statistics from a source state to a target state. The samples in the target ensemble can be directly used to compute observables in the target ensemble.

We audition TI using transformations between different temperatures, and find both ITI and aTI models show promise of efficient generation of statistics from ensembles at multiple different temperatures and accurately estimate changes in free energy. Further, we find that the model interpolates and extrapolates to temperatures not seen during training. Due to the high data and sampling efficiency of TI we envision its use to overcome slow sampling at low temperatures, for example through integration with Parallel Tempering schemes akin to recent work.<sup>14</sup> Combining TI with gEDMD we can study the temperature dependence of kinetics on thermodynamic transformations. We illustrate this through predicting the temperature dependence of exchange between two metastable states in the molecule N-Me and by analyzing the slowest process in the 3p2y1y molecule where we recover qualitative kinetics using biased samples.

A major bottleneck for scalable deployment of TI depends on the specific application. The calculation of change in log-probability relies on the computation of Jacobian determinants of a velocity field during sampling, which scales linearly with the dimension of the system and the number of integration time-steps. Consequently, for larger systems where transformations are more complicated and dimensions are higher, exact computation of free energies is not tractable in the current architectures. Consequently, the explicit treatment of solvent remains computationally impractical. However, there is a steady development in the machine learning community to make architectural as well as algorithmic improvements to speed up such calculations, by enforcing structure in the Jacobian or improving conditioning of the learned velocity fields using ideas from optimal transport. Both of these ideas will be necessary to ensure the competitiveness of TI and related approaches compared to current approaches. In the meantime, as we illustrate, our proposed TI approaches still yield semiquantitative predictions for larger systems, even if reweighing is ignored. Further, future work might benefit from coarse-graining to reduce the number of particles. However, here explicitly dealing with non-Markovian dynamics might be necessary to obtain accurate models.<sup>82</sup> As such, we believe that TI is an important step toward learnable transformations between thermodynamic states, enabling calculations of free energy and kinetics.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jctc.4c01557>.

Additional qualitative results on larger systems, details on free energy perturbation methods, interquartile range (IQR) filtering, data set generation, hyper parameters, model selection, and Markov state modeling details (PDF)

## AUTHOR INFORMATION

### Corresponding Author

**Simon Olsson** – Department of Computer Science and Engineering, Chalmers University of Technology and University of Gothenburg, SE-41296 Gothenburg, Sweden; [orcid.org/0000-0002-3927-7897](https://orcid.org/0000-0002-3927-7897); Email: [simonols@chalmers.se](mailto:simonols@chalmers.se)

### Authors

**Selma Moqvist** – Department of Computer Science and Engineering, Chalmers University of Technology and University of Gothenburg, SE-41296 Gothenburg, Sweden

**Weilong Chen** – Department of Computer Science and Engineering, Chalmers University of Technology and University of Gothenburg, SE-41296 Gothenburg, Sweden

**Mathias Schreiner** – Department of Computer Science and Engineering, Chalmers University of Technology and University of Gothenburg, SE-41296 Gothenburg, Sweden

**Feliks Nüske** – Max-Planck-Institute for Dynamics of Complex Technical Systems, Magdeburg 39106, Germany; [orcid.org/0000-0003-2444-7889](https://orcid.org/0000-0003-2444-7889)

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.jctc.4c01557>

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This work was partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation. We further acknowledge funding from the Knut and Alice Wallenberg Foundation project: “From atom to organism: Bridging the scales in the design of ion channel drugs.” Preliminary results were enabled by resources provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS) at Alvis (project: NAISS 2024/22-33), partially funded by the Swedish Research Council through grant agreement no. 2022-06725. The authors thank Juan Viguera Diez for sharing MDQM9-nc dataset ahead of publication and help with evaluation and Ross Irwin and Johann Flemming Gloy for helpful discussions.

## REFERENCES

- (1) Chandler, D. *Introduction to Modern Statistical Mechanics*; Oxford University Press, 1987.
- (2) Frenkel, D.; Smit, B. *Understanding Molecular Simulation*; Elsevier, 2002.
- (3) Ferrenberg, A. M.; Swendsen, R. H. Optimized Monte Carlo data analysis. *Phys. Rev. Lett.* **1989**, *63*, 1195–1198.
- (4) Shirts, M. R.; Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **2008**, *129*, 124105.
- (5) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562–12566.
- (6) Grubmüller, H. Predicting slow structural transitions in macromolecular systems: Conformational flooding. *Phys. Rev. E* **1995**, *52*, 2893–2906.
- (7) Marinari, E.; Parisi, G. Simulated Tempering: A New Monte Carlo Scheme. *Europhysics Letters (EPL)* **1992**, *19*, 451–458.
- (8) Torrie, G.; Valleau, J. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **1977**, *23*, 187–199.
- (9) Hénin, J.; Lelièvre, T.; Shirts, M. R.; Valsson, O.; Delemotte, L. Enhanced Sampling Methods for Molecular Dynamics Simulations [Article v1.0]. *Living J. Comput. Mol. Sci.* **2022**, *4*, 1583.
- (10) Chen, W.; Ferguson, A. L. Molecular enhanced sampling with autoencoders: On-the-fly collective variable discovery and accelerated free energy landscape exploration. *J. Comput. Chem.* **2018**, *39*, 2079–2102.
- (11) Wang, Y.; Ribeiro, J. M. L.; Tiwary, P. Past–future information bottleneck for sampling molecular reaction coordinate simultaneously with thermodynamics and kinetics. *Nat. Commun.* **2019**, *10*, 3573.
- (12) Herringer, N. S. M.; Dasetty, S.; Gandhi, D.; Lee, J.; Ferguson, A. L. Permutationally Invariant Networks for Enhanced Sampling (PINES): Discovery of Multimolecular and Solvent-Inclusive Collective Variables. *J. Chem. Theory Comput.* **2024**, *20*, 178–198.
- (13) Sultan, M. M.; Pande, V. S. tICA-Metadynamics: Accelerating Metadynamics by Using Kinetically Selected Collective Variables. *J. Chem. Theory Comput.* **2017**, *13*, 2440–2447.
- (14) Invernizzi, M.; Krämer, A.; Clementi, C.; Noé, F. Skipping the Replica Exchange Ladder with Normalizing Flows. *J. Phys. Chem. Lett.* **2022**, *13*, 11643–11649.
- (15) Tiwary, P.; Parrinello, M. From Metadynamics to Dynamics. *Phys. Rev. Lett.* **2013**, *111*, No. 230602.
- (16) Voter, A. F. A method for accelerating the molecular dynamics simulation of infrequent events. *J. Chem. Phys.* **1997**, *106*, 4665–4677.
- (17) Schütte, C.; Fischer, A.; Husinga, W.; Deuffhard, P. *A Hybrid Monte Carlo Method for Essential Molecular Dynamics*, 1998.
- (18) Prinz, J.-H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schütte, C.; Noé, F. Markov models of molecular kinetics: Generation and validation. *J. Chem. Phys.* **2011**, *134*, 174105.
- (19) Swope, W. C.; Pitera, J. W.; Suits, F. Describing Protein Folding Kinetics by Molecular Dynamics Simulations. 1. Theory. *J. Phys. Chem. B* **2004**, *108*, 6571–6581.
- (20) Buchete, N.-V.; Hummer, G. Coarse Master Equations for Peptide Folding Dynamics. *J. Phys. Chem. B* **2008**, *112*, 6057–6069.
- (21) Husic, B. E.; Pande, V. S. Markov State Models: From an Art to a Science. *J. Am. Chem. Soc.* **2018**, *140*, 2386–2396.
- (22) Wu, H.; Noé, F. Variational approach for learning Markov processes from time series data. *J. Nonlinear Sci.* **2020**, *30*, 23–66.
- (23) Nüske, F.; Boninsegna, L.; Clementi, C. Coarse-graining molecular systems by spectral matching. *J. Chem. Phys.* **2019**, *151*, No. 044116.
- (24) Klus, S.; Nüske, F.; Peitz, S.; Niemann, J.-H.; Clementi, C.; Schütte, C. Data-driven approximation of the Koopman generator: Model reduction, system identification, and control. *Physica D: Nonlinear Phenomena* **2020**, *406*, No. 132416.
- (25) Olsson, S.; Noé, F. Dynamic graphical models of molecular kinetics. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 15001–15006.
- (26) Hempel, T.; Olsson, S.; Noé, F. Markov field models: Scaling molecular kinetics approaches to large molecular machines. *Curr. Opin. Struct. Biol.* **2022**, *77*, No. 102458.
- (27) Dellago, C.; Bolhuis, P. G.; Chandler, D. Efficient transition path sampling: Application to Lennard-Jones cluster rearrangements. *J. Chem. Phys.* **1998**, *108*, 9236–9245.
- (28) Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L. Transition Path Sampling: Throwing Ropes Over Rough Mountain Passes, in the Dark. *Annu. Rev. Phys. Chem.* **2002**, *53*, 291–318.
- (29) Jung, H.; Covino, R.; Arjun, A.; Leitold, C.; Dellago, C.; Bolhuis, P. G.; Hummer, G. Machine-guided path sampling to discover mechanisms of molecular self-organization. *Nature Computational Science* **2023**, *3*, 334–345.
- (30) Plainer, M.; Stärk, H.; Bunne, C.; Günnemann, S. *Transition Path Sampling with Boltzmann Generator-based MCMC Moves*, 2023.
- (31) Mardt, A.; Pasquali, L.; Wu, H.; Noé, F. VAMPnets for deep learning of molecular kinetics. *Nat. Commun.* **2018**, *9*, 5.
- (32) Chen, W.; Sidky, H.; Ferguson, A. L. Nonlinear discovery of slow molecular modes using state-free reversible VAMPnets. *J. Chem. Phys.* **2019**, *150*, 214114.

- (33) Mardt, A.; Hempel, T.; Clementi, C.; Noé, F. Deep learning to decompose macromolecules into independent Markovian domains. *Nat. Communications* **2022**, *13*, 7101.
- (34) Wu, H.; Mardt, A.; Pasquali, L.; Noe, F. *Deep Generative Markov State Models*, 2019; <https://arxiv.org/abs/1805.07601>.
- (35) Bowman, G. R.; Ensign, D. L.; Pande, V. S. Enhanced Modeling via Network Theory: Adaptive Sampling of Markov State Models. *J. Chem. Theory Comput.* **2010**, *6*, 787–794.
- (36) Doerr, S.; De Fabritiis, G. On-the-Fly Learning and Sampling of Ligand Binding by High-Throughput Molecular Simulations. *J. Chem. Theory Comput.* **2014**, *10*, 2064–2069.
- (37) Bogetti, A. T.; Mostofian, B.; Dickson, A.; Pratt, A.; Saglam, A. S.; Harrison, P. O.; Adelman, J. L.; Dudek, M.; Torrillo, P. A.; DeGrave, A. J.; Adhikari, U.; Zwier, M. C.; Zuckerman, D. M.; Chong, L. T. A Suite of Tutorials for the WESTPA Rare-Events Sampling Software [Article v1.0]. *Living J. Comput. Mol. Sci.* **2019**, *1*, No. 10607.
- (38) Betz, R. M.; Dror, R. O. How Effectively Can Adaptive Sampling Methods Capture Spontaneous Ligand Binding? *J. Chem. Theory Comput.* **2019**, *15*, 2053–2063.
- (39) Schreiner, M.; Winther, O.; Olsson, S. *Implicit Transfer Operator Learning: Multiple Time-Resolution Surrogates for Molecular Dynamics*, 2023; <https://arxiv.org/abs/2305.18046>.
- (40) Diez, J. V.; Schreiner, M.; Engkvist, O.; Olsson, S. *Boltzmann priors for Implicit Transfer Operators*, 2024.
- (41) Klein, L.; Foong, A. Y. K.; Fjelde, T. E.; Mlodozeniec, B.; Brockschmidt, M.; Nowozin, S.; Noé, F.; Tomioka, R. *Timewarp: Transferable Acceleration of Molecular Dynamics by Learning Time-Coarsened Dynamics*, 2023; <https://arxiv.org/abs/2302.01170>.
- (42) Hsu, T.; Sadigh, B.; Bulatov, V.; Zhou, F. Score Dynamics: Scaling Molecular Dynamics with Picoseconds Time Steps via Conditional Diffusion Model. *J. Chem. Theory Comput.* **2024**, *20*, 2335–2348.
- (43) Jing, B.; Stärk, H.; Jaakkola, T.; Berger, B. *Generative Modeling of Molecular Dynamics Trajectories*, 2024.
- (44) dos Santos Costa, A.; Mitnikov, I.; Pellegrini, F.; Daigavane, A.; Geiger, M.; Cao, Z.; Kreis, K.; Smidt, T.; Kucukbenli, E.; Jacobson, J. *EqiJUMP: Protein Dynamics Simulation via SO(3)-Equivariant Stochastic Interpolants*, 2024.
- (45) Noé, F.; Olsson, S.; Köhler, J.; Wu, H. Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning. *Science* **2019**, *365*, No. eaaw1147.
- (46) Rezende, D. J.; Mohamed, S. *Variational Inference with Normalizing Flows*, 2016.
- (47) Tabak, E. G.; Vanden-Eijnden, E. DENSITY ESTIMATION BY DUAL ASCENT OF THE LOG-LIKELIHOOD. *Communications in Mathematical Sciences* **2010**, *8*, 217–233.
- (48) Diez, J. V.; Atance, S. R.; Engkvist, O.; Olsson, S. Generation of conformational ensembles of small molecules via surrogate model-assisted molecular dynamics. *Mach. Learn.: Sci. Technol.* **2024**, *5*, No. 025010.
- (49) Köhler, J.; Klein, L.; Noé, F. Equivariant flows: exact likelihood generative learning for symmetric densities. *Int. Conf. Mach. Learn.* **2020**, 5361–5370.
- (50) Klein, L.; Noé, F. *Transferable Boltzmann Generators*, 2024; <https://arxiv.org/abs/2406.14426>.
- (51) Jing, B.; Corso, G.; Chang, J.; Barzilay, R.; Jaakkola, T. *Torsional Diffusion for Molecular Conformer Generation*, 2022.
- (52) Zheng, S.; et al. Predicting equilibrium distributions for molecular systems with deep learning. *Nature Machine Intelligence* **2024**, *6*, 558–567.
- (53) Herron, L.; Mondal, K.; Schneekloth, J. S.; Tiwary, P. Inferring phase transitions and critical exponents from limited observations with Thermodynamic Maps. *Proc. Natl. Acad. Sci. U. S. A.* **2023**, *121* (52), No. e2321971121.
- (54) Rizzi, A.; Carloni, P.; Parrinello, M. Targeted Free Energy Perturbation Revisited: Accurate Free Energies from Mapped Reference Potentials. *J. Phys. Chem. Lett.* **2021**, *12*, 9449–9454.
- (55) Wirnsberger, P.; Ballard, A. J.; Papamakarios, G.; Abercrombie, S.; Racanière, S.; Pritzel, A.; Jimenez Rezende, D.; Blundell, C. Targeted free energy estimation via learned mappings. *J. Chem. Phys.* **2020**, *153*, 144112.
- (56) Máté, B.; Fleuret, F.; Bereau, T. *Neural Thermodynamic Integration: Free Energies from Energy-based Diffusion Models*, 2024; <https://arxiv.org/abs/2406.02313>.
- (57) Dibak, M.; Klein, L.; Krämer, A.; Noé, F. Temperature steerable flows and Boltzmann generators. *Phys. Rev. Res.* **2022**, *4*, No. L042005.
- (58) Albergo, M. S.; Boffi, N. M.; Vanden-Eijnden, E. Stochastic interpolants: A unifying framework for flows and diffusions. *arXiv preprint arXiv:2303.08797* 2023.
- (59) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A. Quantum chemistry structures and properties of 134 kilo molecules. *Sci. Data* **2014**, *1*, 140022.
- (60) Zwanzig, R. W. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *J. Chem. Phys.* **1954**, *22*, 1420–1426.
- (61) Jarzynski, C. Targeted free energy perturbation. *Phys. Rev. E* **2002**, *65*, No. 046122.
- (62) Willow, S. Y.; Kang, L.; Minh, D. D. L. Learned mappings for targeted free energy perturbation between peptide conformations. *J. Chem. Phys.* **2023**, *159*, 124104.
- (63) Jarzynski, C. Nonequilibrium Equality for Free Energy Differences. *Phys. Rev. Lett.* **1997**, *78*, 2690–2693.
- (64) Deuschel, J.-D.; Stroock, D. W. *Large deviations*; American Mathematical Soc., 2001; Vol. 342.
- (65) Hartmann, C.; Richter, L.; Schütte, C.; Zhang, W. Variational characterization of free energy: Theory and algorithms. *Entropy* **2017**, *19*, 626.
- (66) Dinh, L.; Sohl-Dickstein, J.; Bengio, S. *Density estimation using Real NVP*, 2017; <https://arxiv.org/abs/1605.08803>.
- (67) Chen, R. T.; Rubanova, Y.; Bettencourt, J.; Duvenaud, D. K. Neural ordinary differential equations. *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 6572–6583.
- (68) Thomas, N.; Smidt, T.; Kearnes, S.; Yang, L.; Li, L.; Kohlhoff, K.; Riley, P. *Tensor field networks: Rotation- and translation-equivariant neural networks for 3D point clouds*, 2018.
- (69) Bietti, A.; Venturi, L.; Bruna, J. *On the Sample Complexity of Learning under Invariance and Geometric Stability*, 2021.
- (70) Schreiner, M.; Winther, O.; Olsson, S. Implicit Transfer Operator Learning: Multiple Time-Resolution Models for Molecular Dynamics. *Adv. Neural Inf. Process. Syst.* **2023**, 36449–36462.
- (71) Schebek, M.; Invernizzi, M.; Noé, F.; Rogal, J. *Efficient mapping of phase diagrams with conditional Boltzmann Generators*, 2024.
- (72) Williams, M. O.; Kevrekidis, I. G.; Rowley, C. W. A Data-Driven Approximation of the Koopman Operator: Extending Dynamic Mode Decomposition. *J. Nonlinear Sci.* **2015**, *25*, 1307–1346.
- (73) Klus, S.; Nüske, F.; Koltai, P.; Wu, H.; Kevrekidis, I.; Schütte, C.; Noé, F. Data-Driven Model Reduction and Transfer Operator Approximation. *J. Nonlinear Sci.* **2018**, *28*, 985–1010.
- (74) Kish, L. *Survey Sampling*; John Wiley & Sons, Inc.: New York, 1965.
- (75) Molgedey, L.; Schuster, H. G. Separation of a mixture of independent signals using time delayed correlations. *Phys. Rev. Lett.* **1994**, *72*, 3634–3637.
- (76) Ziehe, A.; Müller, K.-R. *ICANN 98*; Springer: London, 1998; pp 675–680.
- (77) Pérez-Hernández, G.; Paul, F.; Giorgino, T.; De Fabritiis, G.; Noé, F. Identification of slow molecular order parameters for Markov model construction. *J. Chem. Phys.* **2013**, *139*, No. 015102.
- (78) Wu, H.; Paul, F.; Wehmeyer, C.; Noé, F. Multiensemble Markov models of molecular thermodynamics and kinetics. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, E3221–E3230.
- (79) Klus, S.; Nüske, F.; Hamzi, B. Kernel-based approximation of the koopman generator and Schrödinger operator. *Entropy* **2020**, *22*, 722.
- (80) Rahimi, A.; Recht, B. Random features for large-scale kernel machines. *Adv. Neural Inf. Process. Syst.* **2007**, *20*, 1177–1184.

- (81) Nüske, F.; Klus, S. Efficient approximation of molecular kinetics using random Fourier features. *J. Chem. Phys.* **2023**, *159*, No. 074105.
- (82) Cao, S.; Qiu, Y.; Kalin, M. L.; Huang, X. Integrative generalized master equation: A method to study long-timescale biomolecular dynamics via the integrals of memory kernels. *J. Chem. Phys.* **2023**, *159*, 134106.