JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Article

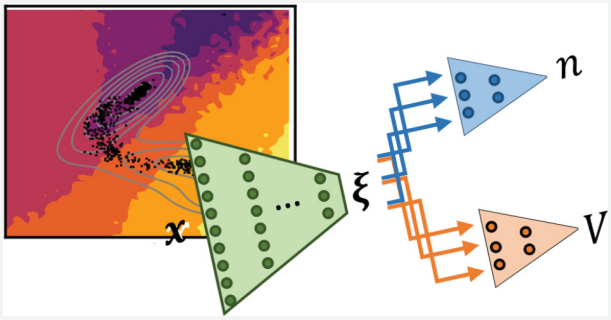
Multitask Machine Learning of Collective Variables for Enhanced Sampling of Rare Events

Lixin Sun,* Jonathan Vandermause, Simon Batzner, Yu Xie, David Clark, Wei Chen, and Boris Kozinsky*

Cite This: J. Chem. Theory Comput. 2022, 18, 2341−2353 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Computing accurate reaction rates is a central challenge in computational chemistry and biology because of the high cost of free energy estimation with unbiased molecular dynamics. In this work, a data-driven machine learning algorithm is devised to learn collective variables with a multitask neural network, where a common upstream part reduces the high dimensionality of atomic configurations to a low dimensional latent space and separate downstream parts map the latent space to predictions of basin class labels and potential energies. The resulting latent space is shown to be an effective low-dimensional representation, capturing the reaction progress and guiding effective umbrella sampling to obtain accurate free energy landscapes. This approach is successfully applied to model systems including a 5D Müller Brown model, a 5D three-well model, the alanine dipeptide in vacuum, and an Au(110) surface reconstruction unit reaction. It enables automated dimensionality reduction for energy controlled reactions in complex systems, offers a unified and data-efficient framework that can be trained with limited data, and outperforms single-task learning approaches, including autoencoders.



1. INTRODUCTION

Computing accurate reaction rates is one of the most important challenges in computational physics, chemistry, and biology. Reactions are rare events in which a system transitions from one metastable state to another. Reactions with high barriers can have time scales of microseconds or longer, such that conventional unbiased molecular dynamics (MD) is too slow to accumulate enough statistics on transitions to calculate accurate reaction rates. Enhanced sampling methods, such as umbrella sampling[1,2] and metadynamics,[3,4] address this challenge by accelerating sampling of phase space using biasing applied along several low-dimensional collective variable (CV) directions. These methods require one to first reduce the high-dimensional configuration space to a low-dimensional manifold of CVs in order to evaluate the free energy landscape where metastable states correspond to local minima basins and transition states correspond to high free energy separation ridges. The reaction rate can then be estimated within the transition state theory framework.

Good CVs need to discern different metastable states and transition paths; they can be simple geometrical variables such as atomic coordination numbers[5] or combinations of bond distances and bond angles.[6] Reaction coordinates (RCs) are a type of CV that need to be one-dimensional and that strictly preserve the reaction progress from one metastable state to another metastable state. Linear combinations of simple geometrical variables are usually not sufficient when the transition paths are complex, and poor choices of CVs result in inefficient sampling and inaccurate reaction rates. However, designing good CVs is a laborious trial-and-error process typically requiring intuition and prior knowledge of the relevant reaction mechanisms.[1−4]

In the spirit of data-driven analysis, a number of methods have recently been employed to design CVs and RCs with machine learning (ML),[7−21] but reactions with high barriers are often difficult to analyze with these ML approaches. For example, methods that identify slow CVs separate slow motions from fast vibrations by monitoring structural evolution over time[8,9,11,12] or by clustering metastable states.[22,23] However, these methods are not practical for high-barrier reactions because their training requires long MD trajectories with a sufficient number of transitions. These transitions are hard to obtain for high-barrier reactions unless enhanced sampling techniques with CVs are used in the first place. This

Received: February 8, 2021
Published: March 11, 2022

chicken-and-egg problem requires the development of iterative adaptive approaches.

There are also methods that learn one-dimensional RCs $\xi(\mathbf{x}) = f(q(\mathbf{x}))$ from the committor function $q(\mathbf{x}) \in [0, 1]$ which describes the progression of a reaction between two predefined basins A and B.[19,24,25] The training data typically come from transition path sampling (TPS)[24,26,27] or relaxation trajectories, in which most configurations are near the transition state and have relatively high free energies. These methods maximize a likelihood of the committor function in the transition path ensemble[18] or transition state ensemble.[28,29] Because the committor function perfectly preserves the reaction progress, it is often considered the ideal reaction coordinate and used to grade the quality of other reaction coordinates.[28,30,31]

However, most committor learning frameworks require a numerical estimation of the commitor function, $q(\mathbf{x})$, for supervised training, which can be data inefficient. For example, to obtain one q value for one configuration $\mathbf{x}$, Ma and Dinner[29] used 100 trajectories of MD simulation, which is 250 ns in total, to obtain a training set of 2100 configurations $\mathbf{x}$ and their $q(\mathbf{x})$. In other words, 125,000 configurations were computed to obtain 2100 training points, with less than 2% data utilization. Moreover, in these methods, the q values need to change smoothly from 0 to 1 in order to provide sufficient information for the training.

For high-barrier reactions, the committor function has a sharp change from 0 to 1 around the transition hypersurface, while in the majority of the phase space, the committor is close to 0 or 1. Numerical estimation of the committor function is computed by the counts of trajectories that commit to a basin divided by the total number of trajectories. In a high-barrier reaction, any configuration that is slightly away from the transition state will only commit to the same basin, which means the estimated values are almost always 0, 1, and values around 0.5. This makes it difficult to accurately estimate the committor function and the CVs derived from it, as discussed in more detail in Section 2.3.

Essentially, these two groups of methods both suffer from limited training data or slow convergence in statistics for high-barrier reactions. Therefore, we identify two requirements for approaches to learning robust CVs. First, the method needs to learn from a limited amount of training data, including configurations from basins and transition states. Second, the CVs need to capture the distinction and progression of intermediate states along the reaction path.

We introduce an approach to simultaneously fulfill the above goals using a multitask machine learning model. Multitask machine learning models consist of a common upstream part that processes the input data and separate downstream parts that produce several outputs, where a joint loss function is used for training. The output of the common upstream part is called latent space. The latent space of the optimized model will then encode the union of all the downstream information. This training strategy has been used in the field of image classification[32,33] and natural language processing[34] for dimensionality reduction and improving generalization performance.

In this work, we represent the simultaneous requirements of the CVs as multiple loss functions, design separate downstream parts for each loss function, and use the latent space as CVs. Unlike previous methods, which typically ignore potential energies, here the multitask learning exploits the potential energy label and uses it as a way to measure the reaction progress for high-barrier reactions. The model is trained with a combination of short MD trajectories, including relaxation from the transition state to the basins and ones that are confined to the basins with no transitions. The learning algorithm is applied to several model systems, including a 5D Müller-Brown model, a 5D three-well model, and the alanine dipeptide. The latent space is shown to be an effective low-dimensional representation of atomic configurations, identifying the important dimension for the reactions and yielding accurate reaction free energies. In addition, the multitask learning framework is shown to be more data efficient than conventional commitor learning methods and to outperform single-task learning frameworks, such as an autoencoder.

2. MULTITASK LEARNING APPROACH

2.1. Architecture. In the multitask learning framework, both the network architecture and the loss function should be designed to reflect the training data's nature and the RC/CV learning objectives. We can break CV learning into three tasks: (T1) dimension reduction, (T2) separating basins, and (T3) preserving atomic structural evolution from basins to TS hypersurface. The multitask neural network contains three parts corresponding to these three tasks.

An encoder is designed as the common upstream part to handle T1. This encoder is a neural network whose hidden layers have progressively fewer nodes, as the layer is closer to the latent layer output. It takes as input the Cartesian coordinates of atomic configurations x and maps them to a low-dimensional latent space ξ. For T2, as discussed in Section 2.3, we assign a basin label n to each $\mathbf{x}$, so that one of the downstream networks is a classifier trained with supervised learning; and for T3, the potential energy labels are exploited. In a high-barrier reaction, the system has to go through low potential energy states before climbing to the higher potential energy transition states. Therefore, potential energies V can be used as an indicator of the reaction progress. The later discussion in Section 4 will show that potential energy outperforms the geometry-based indicator. Here, the second downstream part is a network that predicts potential energy.

Therefore, the multitask neural network has three separate networks: an encoder, a potential energy predictor (PEP), and a basin classifier (Figure 1). The encoder maps the x to the

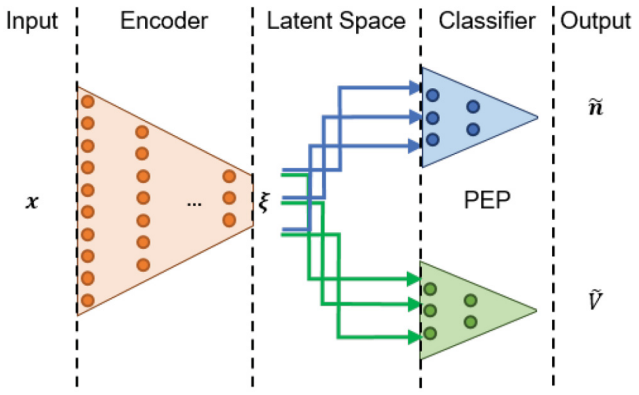


Figure 1. Illustration of the multitask neural network. The input x is the Cartesian coordinates of the atomic configuration. The latent space ξ_{mt} is the output of the encoder. The classifier and the PE networks predict the basin label $\tilde{\mathbf{n}}$ and potential energy $\tilde{V}$ from the latent space output.

https://doi.org/10.1021/acs.jctc.1c00143
J. Chem. Theory Comput. 2022, 18, 2341−2353

latent space $\xi_{mt} = \mathrm{Enc}(\mathbf{x})$. From the latent space ξ_{mt}, the PEP predicts the potential energy $\tilde{V} = h(\xi_{mt})$, and the classifier network predicts the basin label encoded via a one-hot vector $\tilde{n} = [\tilde{n}_\alpha] = \mathbf{g}(\boldsymbol{\xi}_{mt})$, where α denotes the basin label, and $\tilde{n}_\alpha \in [0,1]$. In the following, a tilde is used to indicate predicted quantities, and the absence of a tilde is used to indicate actual true quantities.

2.2. Learning Objectives. The learning objective L is a joint loss function that sums several loss functions L_p with coefficients c_p.

$$L = \sum_p c_p L_p, \quad p = \mathrm{clf}, \mathrm{pe}, \mathrm{reg}, \ldots \tag{1}$$

$$L_{clf} = -\sum_i w_i [n_i \log(\tilde{n}_i)) + (1 - n_i) \log(1 - \tilde{n}_i)] \tag{2}$$

$$L_{pe} = \sum_i u_i (\tilde{V}_i - V_i)^2 \tag{3}$$

These components are the classification error L_{clf}, potential energy error L_{pe}, and regularization on the encoder weights L_{reg}. All terms in eqs 2 and 3 sum over all atomic configurations i in the training data.

The classification error uses the cross-entropy loss, as in eq 2. This term is used to guarantee that different basins are linearly separable in the latent space. The weights of configurations w_i are normalized such that their sum is unity for each basin.

Although the above loss function is written for a two-basin scenario, assuming the true basin label to be integer $n = 0, 1$ and the predicted label as a scalar $\tilde{n} \in [0,1]$, it can be extended to multiple metastable states with multiclass classification cross entropy.

The potential energy loss term is used to preserve the reaction progress from low energy basins to high energy transition states in the latent space. Here, we choose the form for the potential energy loss as the L2 norm. In eq 3, the weights u_i are adjusted such that each potential energy interval $[E_{min}, E_{min} + \Delta E]$, $[E_{min} + \Delta E, E_{min} + 2\Delta E]$, ..., $[E_{max} - \Delta E, E_{max}]$ contributes equally to the loss function.

2.3. Basin Classification Learning. While obtaining the potential energy labels V is straightforward, obtaining the basin labels is not trivial, because the transition state (TS) hypersurface is not known in advance.

As mentioned in the Introduction, maximizing the likelihood related to the committor function can be used to learn an RC $\xi_{mt}(\mathbf{x}) = f(q(\mathbf{x}))$.[18,31] In these methods, only two-basin reactions are concerned. For each TPS shooting point configuration $\mathbf{x}$, the probability of starting at $\mathbf{x}$ and arriving first at the $n = 1$ basin is defined as the committor function

$$p_1(n = 1|\boldsymbol{x}) \equiv q(\boldsymbol{x}) \tag{4}$$

The committor can be estimated by maximizing the join likelihood $\mathcal{L} = \prod_i p_1(n = 1|\boldsymbol{x}_i)$ over all training data i.

Mathematically, the loss function of these methods is equivalent to the cross-entropy classification loss. The committor estimator can hence be a basin classifier that classifies atomic configurations by the predicted basin labels $\tilde{n}(x) = q(x)$ (see the proof in Supplementary Section 1). By definition, the classification boundary $\tilde{n} = 0.5$ is the transition state hypersurface separating the two basins.

However, these frameworks are very data-inefficient for high-barrier reactions. Only the shooting point configurations are utilized which constitute less than 1% of all computed configurations, and it is also hard to statistically estimate the committor function in the remaining region where q is close to 0 (or 1). It requires $\sim 1/q$ (or $\frac{1}{1-q}$) trajectories for a good estimation of q. Otherwise, with only a small number of trajectories, the estimator cannot discern any slight change of q, because the labels for the same $\mathbf{x}$ will be either all zero or all one.

In this work, data utilization is 100%: all computed configurations are used for training, regardless of shooting or nonshooting point configurations. In the presence of a high energy barrier, we can utilize the considerations that (1) the reaction transitions do not occur in a short MD simulation and (2) basin recrossing rarely happens in a short relaxation trajectory starting near the TS.

Therefore, in an unbiased MD simulation, trajectories are trapped at the basin, and all configurations are labeled by the corresponding starting basin. For short relaxation trajectories, each configuration is labeled by the ending basin. For example, a one-way shooting move starts from a chosen high potential energy configuration (shooting point $(\mathbf{x}_{sp}, \mathbf{v}_{sp})$) close to the TS with randomly assigned velocities. The system commits/relaxes toward one of the basins. If a shooting move commits to basin A, all the configurations between the shooting point and the end point are labeled as A ($n = 0$). For configurations close to transition state hypersurfaces, the same configuration can appear in several trajectories with different $(\mathbf{x}_{sp}, \mathbf{v}_{sp})$ that commit to different basins and thus are labeled differently. This method can then be used to statistically sample the basin label for each configuration.

The resulting arrival probability learned with eq 2 is different from the committor q. Because the basin label of nonshooting point configurations depends on the shooting point $(\mathbf{x}_{sp}, \mathbf{v}_{sp})$, the learned probability distribution p_2 is

$$p_2(n = 1|\mathbf{x}) = c \int_\Omega q(\mathbf{x}_{sp}, \mathbf{v}_{sp}) d\mathbf{x}_{sp} d\mathbf{v}_{sp} \tag{5}$$

where Ω contains the starting configurations that lead to a trajectory that arrives at $\mathbf{x}$ before committing to a basin, and c is the normalizing factor. However, in the case of a one-dimensional reaction tube, it can be shown that p_2 is a monotonic transformation of q (in Supplementary Section 1). More importantly, this monotonic dependency is inherited by the latent space variable ξ when the classifier monotonically transforms ξ to p_2 with $g(\xi) = p_2$. In other words, ξ is a monotonic transformation of q if g is a monotonic function of ξ. To ensure this, it is sufficient that all the classifiers used in Sections 3 and 4 are purely linear.

Nonetheless, p_2 is still a good approximation to q numerically in the TS region where $q \approx 0.5$, as confirmed by the results in Sections 3 and 4. This is because the majority of the data around the TS region are shooting point configurations. Thanks to this correlation, the decision boundary of our classifier $p_2 = 0.5$ is close to actual transition states, where $p_1 = 0.5$. In the remaining region where q is close to 0 (or 1), p_2 suffers from the same numerical accuracy problem as $q = p_1$. Because both p_1 and p_2 will be either 0 or 1 in these regions, the learned CVs may not be able to differentiate different intermediate states from the basin to the TS, mapping them to the same CV value; but later in the

https://doi.org/10.1021/acs.jctc.1c00143
J. Chem. Theory Comput. 2022, 18, 2341−2353

discussion, it will be shown that the potential energy label can help remedy such numerical issues by separating these intermediate states while preserving their order in the reaction progress.

2.4. Iterative Training Procedure. The neural network and training framework are implemented with Tensorflow 1.14.[35] The encoder, classifier, and PEP are trained together by the Adam algorithm[36] with a learning rate which is reduced by 5% every 20 epochs. The number of nodes used for the encoder and PEP are listed in Supplementary Tables 1 and 2. In the first 100 epochs, the prefactors c_p are varied randomly with a uniform distribution as follows:

$$c_p \sim \mathrm{Unif}(0, M_p), \quad p = \mathrm{clf}, \mathrm{pe}, \mathrm{reg} \tag{6}$$

The magnitude M_p is chosen such that L_{clf} and L_{pe} contribute equally to the initial loss function value, while regularization terms contribute less than 5%. The choice of starting learning rates, numbers of epochs, and the magnitudes M_p are listed in Supplementary Table 3.

The initial training data is obtained from short unbiased MD simulations, including those that are trapped in a basin or those from TPS. However, depending on the complexity of the reaction, it can be difficult to collect sufficient training data purely using short MD simulations. For example, TPS trajectories may be strongly constrained by the initial path and thus not covering all relevant configurational space, or there are unknown competing reaction paths or basins not included in the initial training data. In order to solve this problem, we introduce an iterative training procedure to collect more training data and converge the latent space for free energy calculations.

As shown in Figure 2, the training procedure includes the following steps:

1. Collect initial training configurations $\{\boldsymbol{x}_i^{(0)}, n_i^{(0)}, V_i^{(0)}\}$ from short unbiased MD simulations.
2. Train a multitask network with latent space $\xi^{(m)} = \mathrm{Enc}^{(m)}(x)$, PEP $\tilde{V}^{(m)} = \mathrm{h}^{(m)}(\xi)$, and classifier $\tilde{\mathbf{n}}^{(m)} = \mathrm{g}^{(m)}(\xi)$, where m is the iteration number.
3. Expand the exploration of the configuration space with biased simulations using $\xi^{(m)}$ as the CV. In this work, around 50−100 umbrella sampling simulations are used, where the biases of each simulation are centered at the grid points in latent space close to the training data that are collected in the previous iterations.
4. Short MD simulations (10−100 ps) are run by restarting from the umbrella sampling simulations in Step 3 but without the bias. Each simulation is run until it (a) reaches a known basin or (b) the number of time steps exceeds a predefined maximum value. All structures in the same simulation are then assigned a label corresponding to the destination basin or "unknown", for cases (a) and (b), respectively. The basin class and potential energy labels of these new configurations are then collected $\{\boldsymbol{x}_i^{(m+1)}, n_i^{(m+1)}, V_i^{(m+1)}\}$.
5. Compute the misclassification rate and potential energy error using $\tilde{V} = \mathrm{h}^{(m)}(\mathrm{Enc}(\boldsymbol{x}_{i(m+1)}))$ and $\tilde{\mathbf{n}} = \mathrm{g}^{(m)}(\mathrm{Enc}(\boldsymbol{x}_{i(m+1)}))$ and compute the free energy landscape on $\xi^{(m)}$. If the errors are high or the free energy landscape is vastly different from the last iteration, meaning that convergence is not reached, add $\{\mathbf{x}_i^{(m+1)}, n_i^{(m+1)}, V_i^{(m+1)}\}$ to the training data and repeat Steps 2−5.
6. Once the CV $\xi^{(m)}$ is converged, use it to estimate the free energy with umbrella sampling (or another enhanced sampling method).

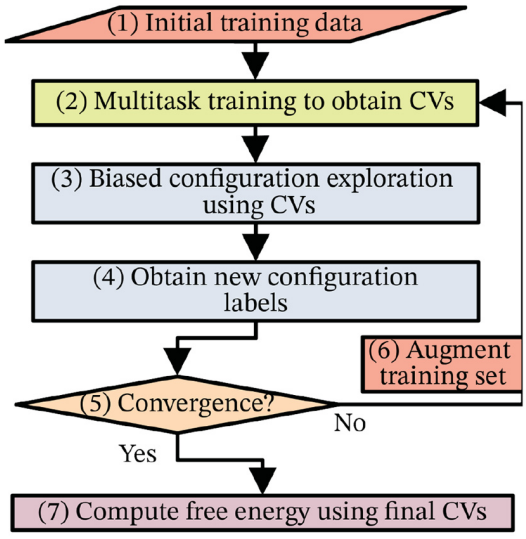


Figure 2. Iterative training workflow to explore and compute free energy landscapes.

More details of the implementation and the training protocol can be found in Supplementary Sections 2 and 3 and the Harvard Dataverse repository.[37−40]

3. CASE STUDIES

In this section, the multitask learning framework is applied to three model systems: a 5D Müller-Brown model, a 5D three-well model, and the alanine dipeptide in vacuum. These three model systems all have well-defined ideal CVs that can be used to accelerate sampling and compute accurate free energies. Our goal is to examine the model's ability to learn complex reaction paths, and hence the CVs in these models are nonlinearly related to the input features, i.e., Cartesian coordinates.

3.1. 5D Müller-Brown and Three-Well Models. We first consider two variations of a model parametrized by five dimensions. The potential energy $V_{5d}(x_1, x_2, ..., x_5)$ is taken as a nonlinear transformation from a two-dimensional function V_{2d} as follows

$$V(x_1, x_2, ..., x_5) = V_{2d}(\tilde{x}, \tilde{y}) \tag{7}$$

$$\tilde{x} = \sqrt{x_1^2 + x_2^2 + 10^{-7}x_5^2} \tag{8}$$

$$\tilde{y} = \sqrt{x_3^2 + x_4^2} \tag{9}$$

The 2D subspace $(\tilde{x}, \tilde{y})$ unambiguously determined the potential energy, while the derived 5D model is made largely degenerate in energy.

The underlying 2D potential function is defined as

$$V_{2d}(\tilde{x}, \tilde{y}) = \sum_{i=1}^{4} A_i \exp[\alpha_i(\tilde{x} - a_i)^2 + \beta_i(\tilde{x} - a_i)(\tilde{y} - b_i) + \gamma_i(\tilde{y} - b_i)^2)] - D(\tilde{x} - d)^3 - E(\tilde{y} - e)^3 \tag{10}$$

Two sets of coefficients, listed in Supplementary Tables 4 and 5, are used to generate a double-well model (so-called Müller-

https://doi.org/10.1021/acs.jctc.1c00143
J. Chem. Theory Comput. 2022, 18, 2341−2353

Brown model) and a three-well model. These coefficients are originally devised by Müller and Brown[41] and Metzner et al.,[42] but we scaled them up to increase the reaction barrier height for dynamics at the temperature T = 300 K. The Müller-Brown model (Figure 3(a)) has two metastable states (A and B) and

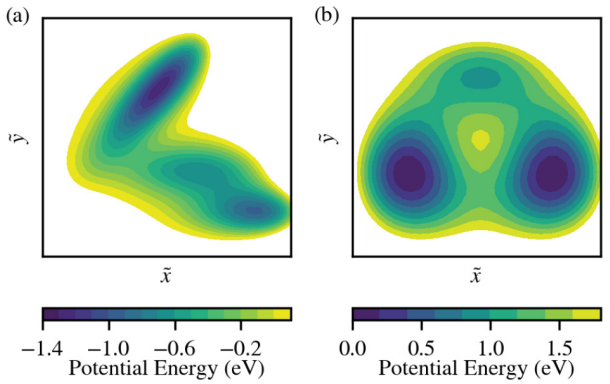


Figure 3. Potential energy landscape of the (a) Müller-Brown model and (b) three-well model in the $(\tilde{x}, \tilde{y})$ subspace.

one minimum energy path between the two basins with a transition barrier of 0.9 eV $(\tilde{3}0k_BT)$ from A to B and 0.6 eV $(\tilde{2}3k_BT)$ from B to A. In the three-well model (Figure 3(b)), two basins (A and B) have lower potential energy than the third basin, C. The transition barrier is 1.0 eV from A/B to C and 1.2 eV from A to B.

Even though the basins for these two models are separable in the 2D $(\tilde{x}, \tilde{y})$ subspace, their structure in the 5D space is obscured by the nonlinear embedding and degeneracy. For example, a linear path in the $(\tilde{x}, \tilde{y})$ subspace (Figure 4(a) and 5(a)) is equivalent to a hypersurface in the 5D space (Figure 4(b) and 5(b)).

For each model, 300,000 configurations were collected from MD trajectories near the basin (without transitions) and trajectories from transition path sampling (Figure 4(a,b)), respectively. The simulation details are documented in Supplementary Section 4, and the data is uploaded to online database.[38,40] In order to test how well the model is able to learn and generalize the transition dynamics, the training and test sets are initialized with different $x_1/\tilde{x}$ and $x_3/\tilde{y}$ ratios, such that the two sets have no overlap in the 5D space but overlap significantly in the $(\tilde{x}, \tilde{y})$ subspace (Figure 4(a,b)). The intention is to test whether the algorithm correctly reduces dimensionality to discover the "true" CV which is given by the 1D minimum energy path connecting the basins in the first model and the 2D $(\tilde{x}, \tilde{y})$ subspace in the second model. The expectation is that if the low-dimensional manifold is identified correctly from the training set, the model will be able to generalize (achieve low prediction error) on the test set even in the presence of degeneracy.

For the Müller-Brown model, the latent space variable ξ_{mt} is chosen to be one-dimensional. After training, the classifier achieves an accuracy of 98% on the test set. As shown in Figure 4(d), the 2% misclassified configurations are located around the transition state ξ_{mt} = 0. This misclassification is hard to avoid, due to the slow variation of the potential energy landscape around the transition ridge. A slight velocity change can lead the system to a different basin, and thus, the configurations around that region can be labeled as both A and B. Figure 4(a,b) shows configurations with label A mixing with

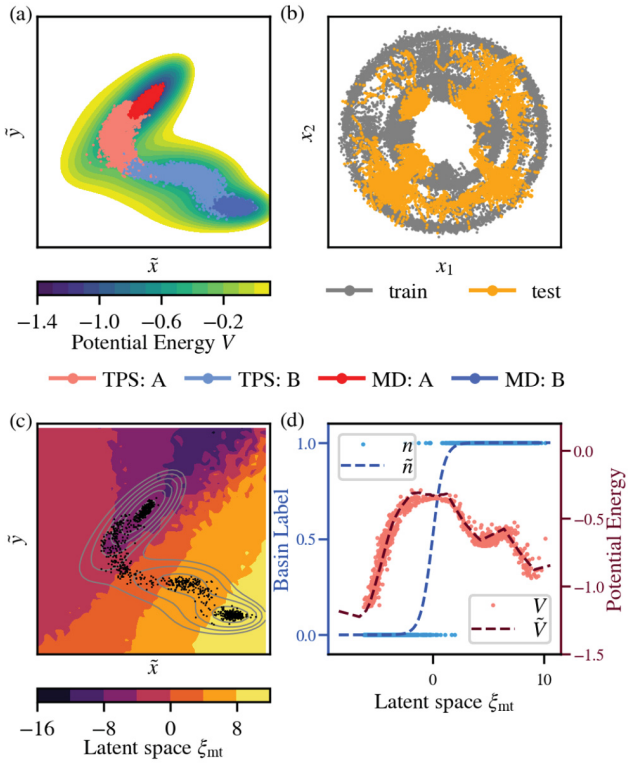


Figure 4. 5D Müller-Brown model. (a) Training data plotted in the $(\tilde{x}, \tilde{y})$ subspace. The configurations labeled with basins A and B are colored as red and blue, respectively. The dots with paler colors are obtained from TPS simulations, while the darker ones are from MD simulations at the basins. The background color contours depict the potential energy $V_{2d}(\tilde{x}, \tilde{y})$. (b) Training and test sets in the (x_1, x_2) subspace, colored gray and orange, respectively. (c) Contour of ξ_{mt} in the $(\tilde{x}, \tilde{y})$ subspace. The ξ_{mt} value of each $(\tilde{x}, \tilde{y})$ value is averaged from five sets of $(x_1, x_2, ..., x_5)$. The black dots are configurations from the test set. The gray lines are true potential energy contours. (d) The predicted/actual potential energy $(\tilde{V}/V)$ and basin label $(\tilde{n}/n)$ as a function of ξ_{mt}.

configurations with label B around the transition state in both the $(\tilde{x}, \tilde{y})$ and (x_3, x_4) subspaces. Thus, this vague separation boundary is kept in the latent space ξ_{mt}. The PEP predicted potential energy $\tilde{V}$ closely follows the ground truth values with a test set mean absolute error (MAE) of 0.04 eV (Figure 4)(d)). The change of $\tilde{V}$ along ξ_{mt} is similar to the actual potential energy V change along the path from A to B. In particular, V is maximized at the decision boundary of the classifier (ξ_{mt} = 0 in Figure 4(d)).

As shown in Figure 4(c), ξ_{mt} is relatively smooth, and more importantly, ξ_{mt} is seen to monotonically tracks to the reaction progress. Especially in the area covered by the test set, the ξ contours are perpendicular to the reaction path and tangential to the potential energy isosurface.

Umbrella sampling is employed to compute the free energy profile along ξ_{mt} and in the $(\tilde{x}, \tilde{y})$ subspace, which is summarized in Table 1, with detailed plots given in Supplementary Figure 1. Free energies computed with the latent space variable ξ_{mt} are reasonably close to the one estimated with the "true" CVs with an error of 0.05−0.08 eV.

For the three-well model, a 2-D latent space (ξ_1, ξ_2) is learned because a single dimension is not enough to differentiate three different transition paths (A ↔ B, A ↔ C,

https://doi.org/10.1021/acs.jctc.1c00143
J. Chem. Theory Comput. 2022, 18, 2341−2353

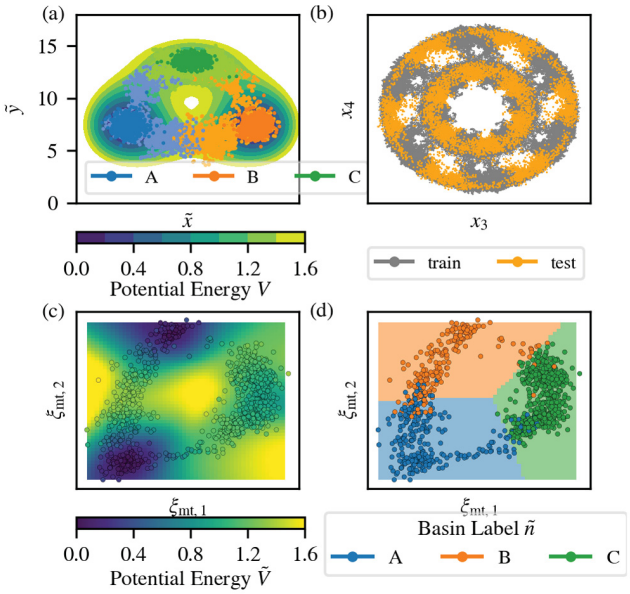


Figure 5. 5D three-well model. (a) Training data in the $(\tilde{x}, \tilde{y})$ subspace. The configurations labeled with basins A, B, and C are colored as blue, orange, and green, respectively. The dots with paler colors are obtained from TPS simulations, while the darker ones are from MD simulations. The background color contours depict the potential energy $V_{2d}(\tilde{x}, \tilde{y})$. (b) Training and test sets in the (x_3, x_4) subspace, colored gray and orange, respectively. (c, d) Potential energy and basin labels in the latent space (ξ_1, ξ_2). The background colors represent (c) the predicted potential energy $\tilde{V}$ and (d) the predicted basin $\tilde{n}$. The scatter points represent the test set, colored by (c) actual potential energies V and (d) actual basin labels n.

Table 1. Reaction Free Energy from Basin A to Basin B of the 5D Müller-Brown Model Computed with Umbrella Sampling along the "True" Collective Variables $\tilde{x}$,$\tilde{y}$ and the Latent Space Variable ξ_{mt}[a]

case	$\Delta F_{A\to B}$	$F_B - F_A$
$(\tilde{x}, \tilde{y})$	0.89 ± 0.02	0.51 ± 0.02
ξ_{mt}	0.94	0.43

[a]Units: eV.

B ↔ C). The trained classifier is able to divide the (ξ_1, ξ_2) subspace into three regions with an accuracy of 90% (Figure 5(d)). In this case, the boundaries between these regions are linear because a linear classifier is used. It is worth noting that this 90% accuracy is close to its theoretical limit because the potential energy landscape is relatively flat around basin C. This is because when using the "true" CVs $(\tilde{x}, \tilde{y})$ as the input, the classification accuracy is seen to be limited to 90%. The PEP prediction for the three-well model has a mean absolute error of 0.07 eV. As shown in Figure 5(c), the PEP reflects the energy rise and fall along all three transition paths.

3.2. Alanine Dipeptide. The multitask learning algorithm is also applied to a real molecular system, the alanine dipeptide in the vacuum. This 22-atom molecule is often used as a model system to demonstrate protein folding and to test dimension reduction algorithms.[16] Compared to the above toy models, the alanine dipeptide is more complex due to its higher input dimension (66 Cartesian coordinates). To train the multitask model, the configurations are shifted and rotated such that the center carbon atom locates at (0, 0, 0), the two connecting C atoms lie on the $x - y$ plane, and one of them lies on the x-axis. Thus, the input feature dimension is, in fact, 63.

This molecule has three metastable states in vacuum, $C7^{ax}$, $C7_{eq}$, and β states.[43] The bottom of these three metastable states can be identified in the 2D Ramachandran plot using two backbone dihedral angles ϕ and θ as coordinates, a well-known widely used set of good CVs for this system. At temperatures below 150 K, no transition between states occurs in a 1-ns long unbiased MD simulation (details in Supplementary Figure 2). $C7_{eq} \leftrightarrow \beta$ transitions are observed above 200 K. As the temperature increases, the $C7_{eq}$-β basin and the $C7_{ax}$ basin grow larger. Above 600 K, some transition events from and to the $C7_{ax}$ can occur within 1 ns. Because the multitask is particularly aimed at dealing with high-barrier transitions, we choose to study the $C7_{eq}$ to $C7_{ax}$ transition at a relatively low temperature, 50−100 K.

Using the 700 K transition events as seeds, transition path sampling can find two different transition paths connecting these three metastable states at 120 K. The transition path ensemble is visualized in the Ramachandran plot in Figure 6.

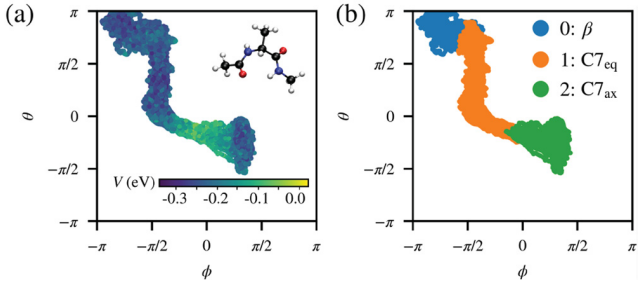


Figure 6. Illustration of training data of the alanine dipeptide in the torsion angles (ϕ, θ) subspace. The molecule structure of the alanine dipeptide is plotted as a subset in (a). ϕ and θ are two torsion angles of the C−N chain. Each point in the plot represents one atomic configuration, and they are colored by (a) the potential energy and (b) the basin label.

The $C7_{eq} \leftrightarrow \beta$ transition has a lower potential energy at the saddle point than that of the $C7_{eq} \leftrightarrow C7_{ax}$ transition. The training data set includes 2 ns MD simulations at the basins and 100 TPS trajectories, each with a 2-ps length. Including some warm-up runs, the total length of simulation used to generate the training set is around 2.5 ns. The test set is obtained in a separate TPS simulation at a slightly lower temperature of 100 K.

The multitask network learned a one-dimensional CV, ξ_{mt}, using the atomic Cartesian coordinates as input. The classifier achieves 86% accuracy, and the PEP network predicts potential energy with a mean absolute error of 0.1 eV. Similar to the 5D models, ξ_{mt} can separate the three basins and reflect the change of potential energy from the saddle points to the basin (Figure 7). Moreover, when trained using the multitask architecture, the encoder can learn the important structural features, the ϕ and θ dihedral torsion angles, from the Cartesian coordinates. This is reflected in Figure 7(a) which shows that the leaned ξ_{mt} is smoothly connected to the two torsion angles ϕ and θ. In contrast, we find that a variety of single-task learning procedures results in a much less smooth connection between ξ_{mt} and the (ϕ, θ) set (see Section 4 and Supplementary Section 5.2).

https://doi.org/10.1021/acs.jctc.1c00143
J. Chem. Theory Comput. 2022, 18, 2341−2353

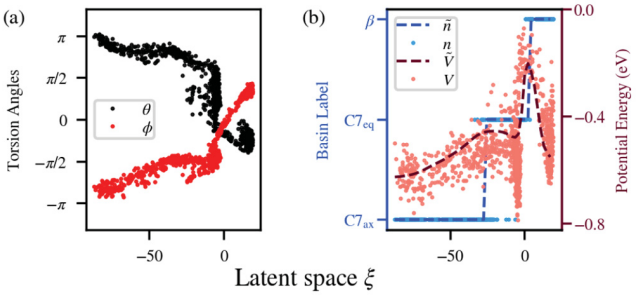


Figure 7. (a) Torsion angles ϕ and θ and (b) predicted basin label $\tilde{n}$, actual basin label n, predicted potential energy $\tilde{V}$, and actual potential energy V as a function of the latent space variable ξ_{mt}. The training data is obtained at 120 K, and the test data is obtained at 100 K.

Compared to the previous 5D models, the classification and potential energy accuracy is lower for the alanine dipeptide due to a larger ratio between thermal fluctuations and the reaction barrier. As a result, many configurations around the $C7_{eq}$ basin are mapped to a small range of ξ_{mt} (Figure 7(a)). However, the $\xi_{mt} - \phi$ and $\xi_{mt} - \theta$ correlation is still relatively smooth in this region. We also note that the training can be much easier with training data obtained using a lower temperature of 50 K since thermal fluctuations are smaller and potential energy is described better with the reaction progress (see Supplementary Table 6).

The learned latent space variable ξ_{mt} from the 50 K training data is then used as a reaction coordinate for the transition between $C7_{eq}$ and $C7_{ax}$. Because the torsion angle ϕ is enough to describe this reaction progress, only ϕ is used as the conventional CV reference. To compare ξ_{mt} and ϕ, two umbrella sampling simulations are employed to estimate the free energy landscape: using either ξ_{mt} as the CV (Figure 8(a,c,e)) at 50 K or ϕ as the CV (Figure 8(b,d,f)) at 300 K. We use LAMMPS[44] and PLUMED[45] codes with an additional interface to load the TensorFlow neural network CV model.[37] Details of the umbrella sampling settings can be found in Supplementary Section 5.3. The resulting umbrella sampling trajectories are analyzed with the bin-less multistate free energy estimation method[46] and UWHAM.[47] For each set of trajectories, two free energy profiles at 50 K are obtained using ξ_{mt} and ϕ (Figure 8(a,b)) for integration. Because CVs are used twice in this procedure, we denote the CV used in umbrella samplings as the sampling CV (SCV) and the one used in reweighting analysis to reconstruct the free energy landscape as the reweighting CV (RCV).

As a reference, ϕ is used as the SCV to sample configurations at 300 K (Figure 8(b,d,f)); the free energy profiles are computed at 50 K. Because the sampling temperature is higher, configurations occupy a larger region. The free energy barriers of the forward/backward transitions between $C7_{eq}$ and $C7_{ax}$ are 0.42/0.33 using ξ_{mt} as the RCV and 0.41/0.32 using ϕ as the RCV.

When ξ_{mt} is used as the SCV at 50 K, the sampled configurations form a narrow path in the $\phi - \theta$ subspace. As expected, the potential energy in this path increases from $C7_{eq}$ to the transition states and then decreases, while ξ_{mt} monotonically increases from $C7_{eq}$ to $C7_{ax}$ (Figure 8(c,e)). The two free energy profiles with the RCV = ξ_{mt} and RCV = ϕ are almost the same with a forward/backward reaction free energy of 0.46/0.35 eV (Figure 8(a) and Table 2). The free energy is only 0.07/0.04 eV different from the one from the

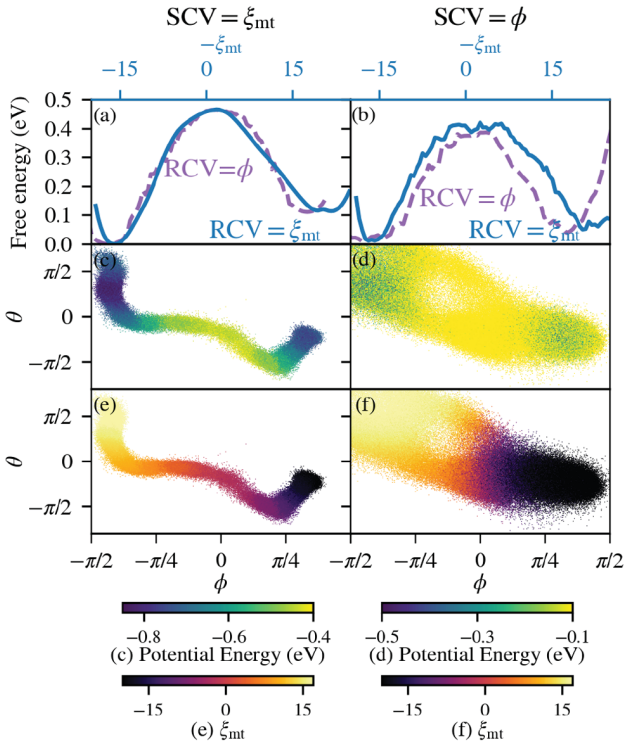


Figure 8. Umbrella sampling for the $C7_{eq}$-$C7_{ax}$ transition. The multitask model is trained with 50 K data. (a, c, e) use ξ_{mt} at 50 K as the CV to define the bias, while (b, d, f) use torsion angle ϕ at 300 K. (a, b) The free energy profile at 50 K along ξ_{mt} (blue line) and ϕ (purple dashed line). (c–f) Sampled atomic configurations plotted in the torsion angles (ϕ, θ) subspace. The configurations are colored by (c, d) potential energy and (e, f) the latent space variable ξ_{mt}.

Table 2. Reaction Free Energy from $C7_{eq}$ to $C7_{ax}$ ($F_{forward}$) and from $C7_{ax}$ to $C7_{ax}$ ($F_{backward}$) of the Alanine Dipeptide and the Free Energy Difference between the Two States ΔF along the Reweighting CV (RCV) Direction by Analyzing Trajectories from Umbrella Samplings Biased along the Sampling CV (SCV)[a]

SCV	RCV	$F_{forward}$	$F_{backward}$	ΔF
ϕ	ϕ	0.39	0.31	0.08
ϕ	ξ_{mt}	0.42	0.36	0.06
ξ_{mt}	ϕ	0.46	0.35	0.11
ξ_{mt}	ξ_{mt}	0.46	0.35	0.11

[a]The simulation temperature is 50 K for the umbrella sampling using ξ_{mt} as the SCV and 300 K for the ones using ϕ as the SCV. The free energy is always estimated at 50 K. Units: eV.

SCV/RCV = ϕ (Figure 8(b)), which can be a result of low sampling efficiency at 50 K and numerical errors. We conclude that in evaluation of free energy barriers and sampling of reaction paths, the learned ξ_{mt} performs as effectively as the reference ϕ coordinate and thus can serve as a good SCV and RCV for estimating reaction rates. Finally, it also worth noting that the amount of data required to train this multitask neural network is 2 orders of magnitude lower than that used by Ma and Dinner[29] for the same system.

3.3. Au (110) Surface. As an example of an application to a realistic extended system, the learning framework is applied to a unit restructuring reaction on the Au(110) surface under vacuum, which is known to exhibit a missing-row reconstruc-

https://doi.org/10.1021/acs.jctc.1c00143
J. Chem. Theory Comput. 2022, 18, 2341−2353

tion.[48] The deconstruction of this missing-row reconstruction can happen at elevated temperatures,[49] upon charge modulation,[50] or with surface molecule adsorption.[51] The intermediate metastable states of the reconstruction/deconstruction process are thought to control the surface reactivity of Au catalysts but are hard to detect experimentally due to their transient nature and the limited time resolution of microscopic techniques.

In this section, we focus on studying the free energy landscape at 300 K for a two-atom reconstruction event, shown in Figure 9, an important unit reaction for the transition

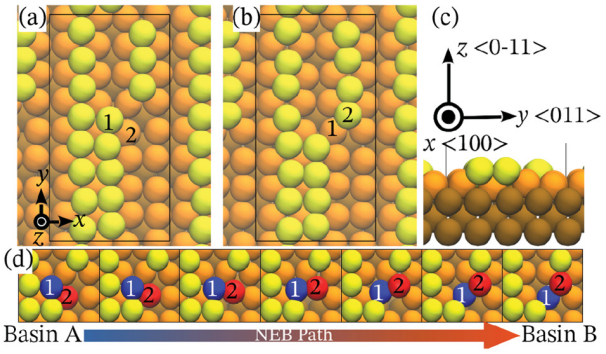


Figure 9. A model Au(110) surface. (a, b) The top view and (c) side view of the two metastable states and (d) the connecting reaction pathway found by NEB. The atoms with the biggest displacement in reaction (d) are colored as blue and red, while in (a, b) they are marked as 1 and 2, respectively. The crystal orientations are noted in (a) and (c). The top layer, subsurface, and bulk Au atoms are represented by yellow, orange, and brown spheres, respectively. The supercell boundary is depicted as the black frame.

between missing-row reconstructed surface and the flat (110) surface with terraces. During the transition from basins A (Figure 9(a)) to B (Figure 9(b)), a top surface atom (marked as 1) moves into the subsurface layer, popping a subsurface atom (marked as 2) up to the top surface and increasing the missing-row reconstructed area.

First, the nudged-elastic-band (NEB) method is used to search for the transition path at 0 K (Figure 9(c)). The initial training data includes short MD trapped at both basins A and B, as well as TPS at 300 K, using the NEB path images to initialize the starting trajectories. A two-dimensional latent space $\xi_{mt} = (\xi_{mt,1}, \xi_{mt,2})$ is used in the multitask network. In order for the model input to satisfy periodic boundary conditions, we choose to transform the x- and y-coordinates of each atom to $(l_x \sin(x/l_x), l_x \cos(x/l_x))$ and $(l_y \sin(y/l_y), l_y \cos(y/l_y))$, where l_x and l_y are periodic supercell dimensions. The transformed x- and y-coordinates, as well as the z-coordinates of all 144 atoms, are used as the encoder input.

We perform 6 training iterations to converge the CV latent space. As shown in Figure 11, the resulting two-dimensional latent space $(\xi_{mt,1}, \xi_{mt,2})$ is able to successfully separate basins A and B and other "unknown" states (see below), with a classification accuracy of 91.3% (Figure 11(c) and potential energy mean absolute error of 3.3 meV/atom (Figure 11(d)). The free energy barrier is computed to be 0.4 eV using $(\xi_{mt,1}, \xi_{mt,2})$ as the SCV and RCV. The free energy landscape is depicted at Figure 10.

We emphasize that unlike the other two systems, for which obvious or exactly known CV references exist, it is hard to find

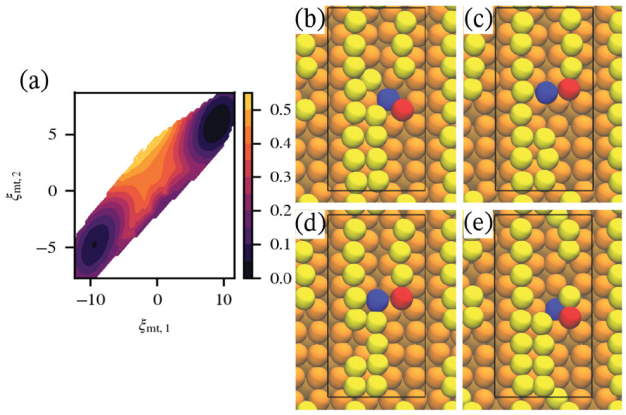


Figure 10. (a) The free energy landscape of the Au(110) missing row desconstruction event. (b–e) Examples of "other" basins, obtained at the last CV training iteration. The color code is the same as in Figure 9.

simple CVs in this realistic system for comparison, because multiple metastable states are closely connected to this reaction path. Because we are only interested in the free energy barrier between basins A and B, the other metastable states are all marked as the "unknown" basins. These "unknown" basins post a challenge in the free energy barrier estimation. Due to their adjacency to the basins A and B in the real space, umbrella sampling simulation can easily cross the barriers to these "unknown" basins. However, the classification accuracy on these basins is not ideal because they are not included in the training set. Therefore, the encoder can map them to basin A or basin B, which affects the free energy calculations, either lowering the free energy in the basins or lowering the free energy at the transition states. This problem can be solved by using the iterative training framework. As shown in Supplementary Figures 3 and 4 and Supplementary Tables 7–9, the iterative training framework gradually improves the multitask network's accuracy on classification and potential energy for these "unknown" basins. As a result, the free energy barrier between basins A and B converges to a constant value.

In order to visualize these "unknown" basins, the z-coordinate of atoms 1 and 2 (written as z_1 and z_2) are used here, only to demonstrate how label (Figure 11(a)) and potential energy (Figure 11(b)) vary in the Cartesian coordinate space. In this (z_1, z_2) subspace, basins A and B are well separated, and the potential energy increases in the proximity of the transition states. However, the other basins appear during TPS and umbrella sampling simulations, visualized in Figure 10 and the Supporting Information. Some of these "unknown" basins share the same (z_1, z_2) coordinate combination as basins A and B, and thus, (z_1, z_2) is not a good collective variable due to the overlap of basins. In fact, using (z_1, z_2) as the RCV, the free energy barrier is estimated as 0.25 eV, significantly lower than the value using $(\xi_{mt,1}, \xi_{mt,2})$ as the RCV.

4. DISCUSSION

The key to efficient learning of collective variables is to combine dimensionality reduction using the encoder with the downstream parts handling tasks T2 and T3 discussed in Section 2, which infuse physical information about the system into the learning process. To compare the performance of all

https://doi.org/10.1021/acs.jctc.1c00143
J. Chem. Theory Comput. 2022, 18, 2341−2353

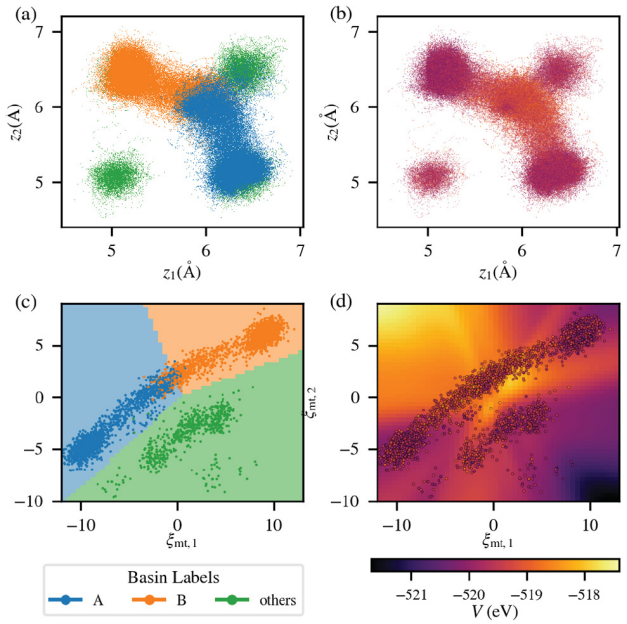


Figure 11. Sixth iteration training data in (a, b) the (z_1, z_2) space and (c, d) the latent $(\xi_{mt,1}, \xi_{mt,2})$ space, where z_1 and z_2 are the z-coordinates of atoms 1 and 2 marked in Figure 9, respectively. Each data point is colored by (a, c) basin labels and (b, d) potential energy.

three models discussed in Section 3, two single-task neural networks are trained: with only the classifier or the PEP downstream parts.

For the sake of brevity, only the case of the 5D Müller-Brown model is presented in the main text, and the remaining data sets are left to Supplementary Sections 4 and 5 (Supplementary Figures 5−7). Similar to Sections 2 and 3, where the latent space variable ξ learned with the multitask framework is denoted as ξ_{mt}, the latent space variables learned with alternative architectures will also be noted with a subscript.

In the first single task learning setup, the neural network has an encoder of the latent space ξ_{clf} and a classifier trained with only the L_{clf} loss function. The latent space ξ_{clf} in this case can still identify and separate the two basins with 95% accuracy (Figure 12(a, e)). Around the transition states, the potential energy V is sharply concentrated around its conditional mean on ξ_{clf}, and its contour is perpendicular to the reaction path around the saddle point region. However, outside of the transition region, the contour is tangential to the reaction path and different energy states mixed at the same ξ_{clf} value. The free energy derived from this ξ_{clf} is 0.52 eV, prominently lower than the ground truth value 0.89 eV. The free energy difference between basin A and B is also greatly underestimated.

The mixing of high energy and low energy states and underestimated reaction free energy indicate that the single task network fails to learn the reaction path in the underlying $(\tilde{x}, \tilde{y})$ subspace, and ξ_{clf} cannot preserve the reaction progress. It misses the nuance of the reaction progress exactly because of the numerical accuracy issue affecting estimation of the committor mentioned in Section 2.3, i.e., the classifier function has very small variation close to the basins. This problem is less severe for the region around the transition state because most of the training data around that area are generated from the TPS shooting point configurations. That is why the potential energies around the transition state ($\xi_{clf} \approx 0$) in Figure 12(e) are closely correlated with ξ_{clf}.

Next, we consider a single-task learning framework where the network has an encoder and a PEP, trained with only the

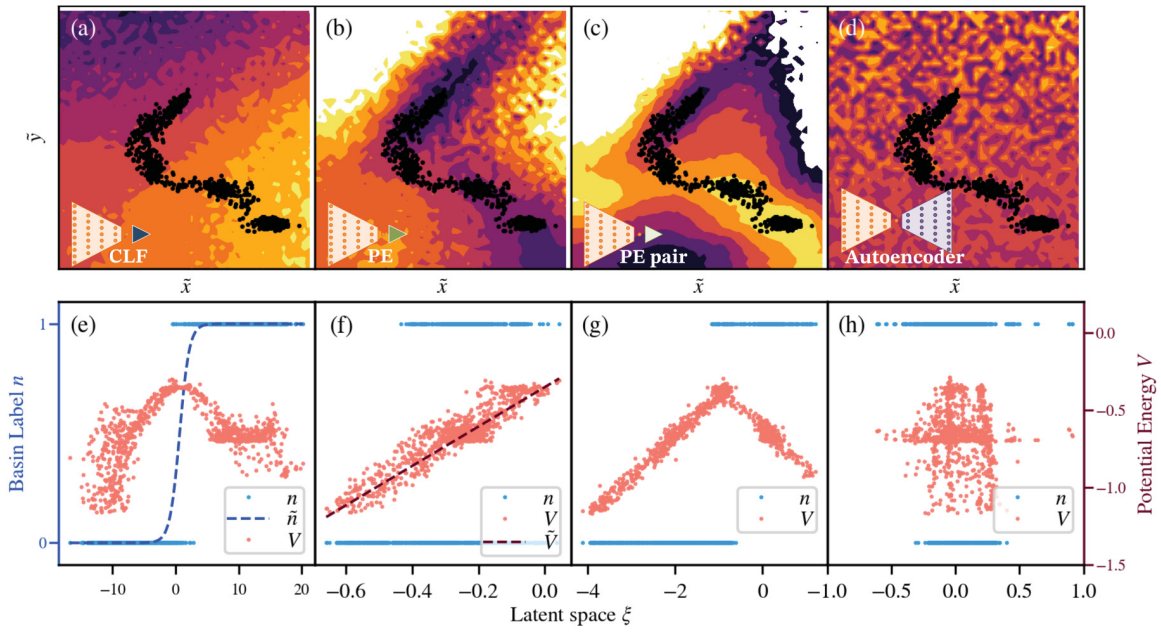


Figure 12. Comparison among different single-task architectures consisting of (a, e) an encoder and a classifier, trained with L_{clf}; (b, f) an encoder and a PEP, trained with L_{pe}; (c, g) an encoder, trained with L_{pe}^{pair}; (d, h) an encoder and a decoder, trained with $L_{reconst}$. (a−d) The spatial distribution of the latent space variable ξ in the $(\tilde{x}, \tilde{y})$ subspace. The black dots represent the location of the training and test sets, and the background contours are colored by the ξ value. We note that a single $(\tilde{x}, \tilde{y})$ value can correspond to many $(x_1, ..., x_5)$ values. For each $(\tilde{x}, \tilde{y})$ in these plots, only one set of $(x_1, ..., x_5)$ is randomly chosen to satisfy eqs 8 and 9. (e−h) Predicted potential energy $\tilde{V}$, predicted label $\tilde{n}$ as a function of ξ.

https://doi.org/10.1021/acs.jctc.1c00143
J. Chem. Theory Comput. 2022, 18, 2341−2353

L_{pe} loss. The contours of ξ_{pe} in Figure 12(b) are close to the true potential energy. However, Figure 12(f) shows that while the encoder clearly orders configurations by their potential energy in the latent space, it assigned many of the configurations from two different basins onto the same values of ξ_{pe}.

The success of multitask learning originates from the synergistic effect among all parts of the loss function. The effect of the L_{clf} loss dominates around the transition state hypersurfaces ($\xi \approx 0$) and removes the energy degeneracy across basins by separating them in the latent space. Simultaneously, minimizing the L_{pe} loss tends to order the configurations by potential energy so that the reaction progress from the bottom of the basin toward the transition states is preserved.

There is the freedom to choose the exact expressions for L_{pe} and L_{clf} as long as they accomplish tasks T2 and T3 listed in Section 2.1. In fact, T3 (preserving atomic structural evolution) can be achieved by any loss function and architecture that captures information about the proximity between configurations along the reaction path. We demonstrate below a successful example with an alternative form of the loss depending on the potential energy ${L_{pe}}^{pair}$ and an unsuccessful example of an autoencoder reconstruction loss $L_{reconst}$.

In the first example, potential energy is used to measure the proximity of configurations, whereby the pairwise distance between two configurations i and j in the latent space is trained to match their potential energy difference. Thus, ${L_{pe}}^{pair}$ is taken as the L2 norm for differences between $V_i - V_j$ and ${d_{ij}}^{(l)}$, written as

$$L_{pe}^{pair} = \sum_{\{i,j|n_i=n_j\}} v_{ij}[(d_{ij}^{(l)})^2 - (V_i - V_j)^2]^2 \tag{11}$$

Here $V_i - V_j$ is the potential energy difference between two data points i and j, and ${d_{ij}}^{(l)} = \|\xi_i - \xi_j\|$ is the pairwise Euclidean distance in the reduced l-dimensional latent space. Only (i, j) pairs where both points are within the same basin class are considered. Unlike u_i in eq 3, which is defined for a single data point, the weight v_{ij} is defined for a pair of data points

$$v_{ij} = 1 - s\left(\frac{|V_i - V_j| - V_0}{\Delta V}\right) \tag{12}$$

where s is a sigmoid function. This weight drops to zero when the potential energy difference of the pair is much greater than V_0 ($|V_i - V_j| \gg V_0$). The parameter ΔV is used to control how fast the weight drops to zero around V_0.

In the second example, an autoencoder[52] is tested, which maps atomic configurations to the latent space variable $\xi_{reconst}$ with an encoder and then maps $\xi_{reconst}$ onto a reconstructed atomic configuration $\tilde{x}_i$ with a decoder. Autoencoders are often used for dimension reduction and manifold learning. It is generally believed that the latent space can preserve the proximity of configurations by minimizing the Euclidean distances ${d_{ij}}^{(3N_a)}$ between the original and reconstructed atomic configurations. The reconstruction loss $L_{reconst}$ is defined as follows

$$L_{reconst} = \sum_i |\mathbf{x}_i - \tilde{\mathbf{x}}_i|^2 \tag{13}$$

where N_a is the number of atoms in the atomic configurations.

For $L = {L_{pe}}^{pair}$, Figure 12(g) shows that the encoder still orders the configurations by potential energy in the latent space, with the additional feature that the two basins are separated, thanks to the fact that only same-basin pairs are used in eq 11. The resulting ${\xi_{pe}}^{pair}$ is very closely correlated to the actual potential energy landscape in $(\tilde{x}, \tilde{y})$ (Figure 12(c)). More interestingly, such correlation extends beyond the training and test set region in the $(\tilde{x}, \tilde{y})$ space. In this sense, it is more robust than the multitask latent space.

However, this observed robustness is purely fortuitous; it does not work well on other models. For the 5D three-well model, ${L_{pe}}^{pair}$ mixes basins A and B as seen in Supplementary Figures 6, 7(g). For the alanine dipeptide, also it mixes the class labels for all three basins (Supplementary Figure 5g), but this problem can be remedied by introducing additional terms in the loss function to separate pairs of configurations belonging to different basins. Eq 11 is just an example.

For the autoencoder with $L_{reconst}$, the basin labels and potential energies are entangled in the latent space $\xi_{reconst}$. From Figure 12(d), it appears that $\xi_{reconst}$ has negligible correlation with the "true" CVs $\tilde{x}$ or $\tilde{y}$.

A successful loss function must guide the latent space to preserve the proximity between configurations in the reaction progress. Because the autoencoder directly uses Euclidean distances $|x_i - \tilde{x}_i|$, it can have difficulty capturing reaction progress in the presence of large displacements near the free energy landscape basins. In the two 5D models, points that are close in the 2D $(\tilde{x}, \tilde{y})$ subspace can be far away from each other in the 5D space. In particular, x_5 has little impact on $\tilde{x}$ and the potential energy, but its fluctuation can dominate the 5D Euclidean distance between configurations. Therefore, the $L_{reconst}$ loss is likely to fail due to the interference of x_5. In fact, for the two 5D models, almost all standard dimension reduction techniques that are based solely on Euclidean distances in the 5D space are expected to struggle. Using potential energy as the distance function can help the model assign correct reaction progress to these large displacement configurations. It is especially useful for energy-activated transitions, because potential energy is closely related to the reaction progress in the configuration space around the reaction path. That is why including L_{pe} or ${L_{pe}}^{pair}$ in the multitask joint loss along with the basin label information tends to work well.

In addition, the limited dimensionality of the latent space may inhibit the performance of autoencoders. Autoencoders may have a better chance if the dimensionality of $\xi_{reconst}$ exceeds the intrinsic dimension of the data manifold.[53] For example, Chen et al.[13] used an autoencoder with a 2D latent space for exploring the energy landscape of the alanine dipeptide, while Wang et al.[54] used a 9-dimensional latent space to coarse grain the same molecule.

The multitask framework introduced in this work can take training configurations from different types of simulations, as long as the basin class and potential energy labels are available. The training data do not need to follow the Boltzmann distribution, as required by methods for finding the committor function, or be Markovian, as required by the methods for finding slow eigenmodes. As noted in Section 3.2 for the alanine dipeptide, this flexibility increases the data efficiency of the training framework. It has 100% data utilization and needs orders of magnitude lower amount of training data compared to conventional committor learning frameworks. Including the

https://doi.org/10.1021/acs.jctc.1c00143
J. Chem. Theory Comput. 2022, 18, 2341−2353

potential energy label can also improve the convergence in the iterative training framework, which further enhances its data efficiency. This is exemplified with the Au surface reconstruction reaction discussed in Section 3.3. As shown in Supplementary Table 9, the multitask case converges faster than the single-task case in reaction free energies. Furthermore, the multitask learning framework can also be generalized to accommodate additional downstream parts with other manifold learning loss functions that utilize time correlations of configurations.[55]

We note that our learning objective is limited to reactions that involve a substantial change in potential energy. It assumes the data is distributed around reaction tubes whose potential energy correlates well with reaction progress or around the bottom of basins whose potential energy does not vary significantly. For diffusion-dominated processes or reactions with entropy bottlenecks, potential energy is not a good distance metric, and our learning procedure may not have advantages over existing methods.

5. SUMMARY

In summary, we propose to use a multitask training algorithm to learn collective variables from configurations labeled by the basin class and potential energy. These can be obtained, for instance, from MD trajectories and transition path sampling trajectories. The neural network architecture contains an upstream encoder that maps atomic configurations onto a low-dimensional latent space and two other downstream networks that predict the basin labels and potential energy from the latent space value, which is optimized for the classification of configurations among the basins and the prediction of the potential energy. The resulting free energy barrier can be a useful input for kinetics Monte Carlo modeling, where the transition rate for each reaction in the reaction network needs to be explicitedly listed.

The algorithm is applied to study a 5D Müller-Brown model, a 5D three-well model, the alanine dipeptide, and a Au(110) surface reconstruction step. We show that due to the synergy in the multiple learning objectives, the multitask model can perform nonlinear dimensionality reduction and identify collective variables that represent well the reaction progress between the basins. The multitask model requires significantly less training data compared to conventional methods. Finally, we demonstrate that the learned collective variables can be used in enhanced sampling methods, such as umbrella sampling, to obtain accurate free energy barriers. This approach opens the possibilities for automated discovery of low-dimensional coordinates for describing a variety of chemical reactions and computing their rates.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.1c00143.

> Proof showing connection between maximum likelihood methods and cross-entropy classification loss, implementation of neural networks, transition path sampling algorithm, and details of three model systems (5D Müller-Brown, 5D three-well model, alanine dipeptide, and reconstruction on Au(110) surface) (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Lixin Sun − *John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, United States;* orcid.org/0000-0002-7971-5222; Email: lixinsun@fas.harvard.edu

Boris Kozinsky − *John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, United States;* orcid.org/0000-0002-0638-539X; Email: bkoz@seas.harvard.edu

Authors

Jonathan Vandermause − *John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, United States*

Simon Batzner − *John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, United States*

Yu Xie − *John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, United States*

David Clark − *John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, United States*

Wei Chen − *John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, United States;* orcid.org/0000-0003-3598-2369

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.1c00143

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors are thankful for the helpful discussions with Patrick Riley, Ekin Dogus Cubuk, and Kai Kohlhoff. L.S., S.B., and W.C. are supported by the Integrated Mesoscale Architectures for Sustainable Catalysis (IMASC), an Energy Frontier Research Center funded by the US Department of Energy (DOE), Office of Science, Office of Basic Energy Sciences under Award No. DE-SC0012573; S.B., J.V., and B.K. acknowledge partial support from Bosch Research. J.V. is partially supported by the National Science Foundation (NSF), Office of Advanced Cyberinfrastructure, Award No. 2003725. Y.X. is supported by the US Department of Energy (DOE) Office of Basic Energy Sciences under Award No. DE-SC0020128. The training data is generated with Cori at the National Energy Research Scientific Computing Center (NERSC), a DOE Office of Science User Facility supported under Contract No. DE-AC02-05CH11231, through allocation m3275. The models are trained on the FASRC Cannon cluster supported by the FAS Division of Science Research Computing Group at Harvard University and Longhorn in the Frontera project supported by Texas Advanced Computing Center (TACC) of the University of Texas at Austin, through allocation DMR20013.

■ REFERENCES

(1) Torrie, G. M.; Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **1977**, *23*, 187−199.

(2) Souaille, M.; Roux, B. Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Computer physics communications* **2001**, *135*, 40−57.
(3) Barducci, A.; Bonomi, M.; Parrinello, M. Metadynamics. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2011**, *1*, 826−843.
(4) Bonomi, M.; Barducci, A.; Parrinello, M. Reconstructing the equilibrium Boltzmann distribution from well-tempered metadynamics. *J. Comput. Chem.* **2009**, *30*, 1615−1621.
(5) Ensing, B.; Klein, M. L. Perspective on the reactions between F− and CH3CH2F: The free energy landscape of the E2 and SN2 reaction channels. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6755−6759.
(6) Baftizadeh, F.; Cossio, P.; Pietrucci, F.; Laio, A. Protein Folding and Ligand-Enzyme Binding from Bias-Exchange Metadynamics Simulations. *Current Physical Chemistry* **2012**, *2*, 79−91.
(7) Rydzewski, J.; Nowak, W. Machine Learning Based Dimensionality Reduction Facilitates Ligand Diffusion Paths Assessment: A Case of Cytochrome P450cam. *J. Chem. Theory Comput.* **2016**, *12*, 2110−2120.
(8) Boninsegna, L.; Gobbo, G.; Noé, F.; Clementi, C. Investigating Molecular Kinetics by Variationally Optimized Diffusion Maps. *J. Chem. Theory Comput.* **2015**, *11*, 5947−5960.
(9) Ferguson, A. L.; Panagiotopoulos, A. Z.; Kevrekidis, I. G.; Debenedetti, P. G. Nonlinear dimensionality reduction in molecular simulation: The diffusion map approach. *Chem. Phys. Lett.* **2011**, *509*, 1−11.
(10) Ceriotti, M.; Tribello, G. A.; Parrinello, M. Simplifying the representation of complex free-energy landscapes using sketch-map. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 13023−13028.
(11) Ardevol, A.; Tribello, G. A.; Ceriotti, M.; Parrinello, M. Probing the Unfolded Configurations of a β-Hairpin Using Sketch-Map. *J. Chem. Theory Comput.* **2015**, *11*, 1086−1093.
(12) Ardevol, A.; Palazzesi, F.; Tribello, G. A.; Parrinello, M. General Protein Data Bank-Based Collective Variables for Protein Folding. *J. Chem. Theory Comput.* **2016**, *12*, 29−35.
(13) Chen, W.; Tan, A. R.; Ferguson, A. L. Collective variable discovery and enhanced sampling using autoencoders: Innovations in network architecture and error function design. *J. Chem. Phys.* **2018**, *149*, 072312.
(14) Wehmeyer, C.; Noé, F. Time-lagged autoencoders: Deep learning of slow collective variables for molecular kinetics. *J. Chem. Phys.* **2018**, *148*, 241703.
(15) Tiwary, P.; Berne, B. J. Spectral gap optimization of order parameters for sampling complex molecular systems. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 2839−2844.
(16) Branduardi, D.; Gervasio, F. L.; Parrinello, M. From A to B in free energy space. *J. Chem. Phys.* **2007**, *126*, 054103.
(17) Hovan, L.; Comitani, F.; Gervasio, F. L. Defining an Optimal Metric for the Path Collective Variables. *J. Chem. Theory Comput.* **2019**, *15*, 25−32.
(18) Jung, H.; Covino, R.; Hummer, G. Artificial Intelligence Assists Discovery of Reaction Coordinates and Mechanisms from Molecular Dynamics Simulations. 2019, *arXiv:1901.04595 [cond-mat, physics:physics]*. https://arxiv.org/abs/1901.04595 (accessed 2022-02-22).
(19) Li, Q.; Lin, B.; Ren, W. Computing committor functions for the study of rare events using deep learning. *J. Chem. Phys.* **2019**, *151*, 054112.
(20) Mendels, D.; Piccini, G.; Parrinello, M. Collective Variables from Local Fluctuations. *J. Phys. Chem. Lett.* **2018**, *9*, 2776−2781.
(21) Sultan, M. M.; Pande, V. S. Automated design of collective variables using supervised machine learning. *J. Chem. Phys.* **2018**, *149*, 094106.
(22) Noé, F.; Clementi, C. Collective variables for the study of long-time kinetics from molecular trajectories: Theory and methods. *Curr. Opin. Struct. Biol.* **2017**, *43*, 141−147.
(23) Chen, W.; Sidky, H.; Ferguson, A. L. Capabilities and limitations of time-lagged autoencoders for slow mode discovery in dynamical systems. *J. Chem. Phys.* **2019**, *151*, 064123.
(24) Dellago, C.; Bolhuis, P. G.; Geissler, P. L. *Advances in Chemical Physics*; John Wiley & Sons, Ltd.: 2003; pp 1−78.
(25) Dellago, C.; Bolhuis, P. G. In *Advanced Computer Simulation Approaches for Soft Matter Sciences III*; Holm, C., Kremer, K., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2009; pp 167−233, DOI: 10.1007/978-3-540-87706-6_3.
(26) Bolhuis, P. G.; Dellago, C. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Ed.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2010; pp 111−210, DOI: 10.1002/9780470890905.ch3.
(27) 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.
(28) Best, R. B.; Hummer, G. Reaction coordinates and rates from transition paths. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6732−6737.
(29) Ma, A.; Dinner, A. R. Automatic Method for Identifying Reaction Coordinates in Complex Systems. *J. Phys. Chem. B* **2005**, *109*, 6769.
(30) Peters, B.; Trout, B. L. Obtaining reaction coordinates by likelihood maximization. *J. Chem. Phys.* **2006**, *125*, 054108.
(31) Peters, B. Using the histogram test to quantify reaction coordinate error. *J. Chem. Phys.* **2006**, *125*, 241101.
(32) Zhang, L.; Qi, G.-J.; Wang, L.; Luo, J. Aet vs. aed: Unsupervised representation learning by auto-encoding transformations rather than data. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* **2019**, 2542−2555.
(33) Komodakis, N.; Gidaris, S. Unsupervised representation learning by predicting image rotations. *International Conference on Learning Representations*; ICLR: 2018.
(34) Collobert, R.; Weston, J. A unified architecture for natural language processing: Deep neural networks with multitask learning. *Proceedings of the 25th international conference on Machine learning*; 2008; pp 160−167, DOI: 10.1145/1390156.1390177.
(35) Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G. S.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Goodfellow, I.; Harp, A.; Irving, G.; Isard, M.; Jia, Y.; Jozefowicz, R.; Kaiser, L.; Kudlur, M.; Levenberg, J.; Mané, D.; Monga, R.; Moore, S.; Murray, D.; Olah, C.; Schuster, M.; Shlens, J.; Steiner, B.; Sutskever, I.; Talwar, K.; Tucker, P.; Vanhoucke, V.; Vasudevan, V.; Viégas, F.; Vinyals, O.; Warden, P.; Wattenberg, M.; Wicke, M.; Yu, Y.; Zheng, X. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. 2015. https://www.tensorflow.org/ (accessed 2022-02-22). Software available from https://www.tensorflow.org/
(36) Kingma, D. P.; Ba, J. Adam: A Method for Stochastic Optimization. 2014, arXiv:1412.6980. *arXiv e-prints*. https://arxiv.org/abs/1412.6980 (accessed 2022-02-22).
(37) MLmtCV, Multitask machine learning framework for collective variables. 2022. https://github.com/mir-group/MLmtCV (accessed 2022-02-16).
(38) NDSimulator - a python molecular dynamic engine for N-dimensional energy landscapes. 2022. https://github.com/mir-group/NDSimulator (accessed 2022-02-16).
(39) MLmtCV, PLUMED Plugin. 2022. https://github.com/mir-group/MLmtCV-PLUMED-Plugin (accessed 2022-02-16).
(40) 5D Muller-Brown and three-hole dataset. 2022. https://doi.org/10.7910/DVN/XLD7VD (accessed 2022-02-16).
(41) Müller, K.; Brown, L. D. Location of saddle points and minimum energy paths by a constrained simplex optimization procedure. *Theoret. Chim. Acta* **1979**, *53*, 75−93.
(42) Metzner, P.; Schütte, C.; Vanden-Eijnden, E. Illustration of transition path theory on a collection of simple examples. *J. Chem. Phys.* **2006**, *125*, 084110.
(43) Strodel, B.; Wales, D. J. Free energy surfaces from an extended harmonic superposition approach and kinetics for alanine dipeptide. *Chem. Phys. Lett.* **2008**, *466*, 105−115.
(44) Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **1995**, *117*, 1−19.
(45) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.;

Parrinello, M. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comput. Phys. Commun.* **2009**, *180*, 1961−1972.
(46) Tan, Z.; Gallicchio, E.; Lapelosa, M.; Levy, R. M. Theory of binless multi-state free energy estimation with applications to protein-ligand binding. *J. Chem. Phys.* **2012**, *136*, 144102.
(47) Zhang, B. W.; Arasteh, S.; Levy, R. M. The UWHAM and SWHAM Software Package. *Sci. Rep.* **2019**, *9*, 2803.
(48) Marks, L. D. Direct Imaging of Carbon-Covered and Clean Gold (110) Surfaces. *Phys. Rev. Lett.* **1983**, *51*, 1000−1002.
(49) Sturmat, M.; Koch, R.; Rieder, K. H. Real Space Investigation of the Roughening and Deconstruction Transitions of Au(110). *Phys. Rev. Lett.* **1996**, *77*, 5071−5074.
(50) Lozovoi, A. Y.; Alavi, A. Reconstruction of charged surfaces: General trends and a case study of Pt(110) and Au(110). *Phys. Rev. B* **2003**, *68*, 245416.
(51) Hiebel, F.; Shong, B.; Chen, W.; Madix, R. J.; Kaxiras, E.; Friend, C. M. Self-assembly of acetate adsorbates drives atomic rearrangement on the Au(110) surface. *Nat. Commun.* **2016**, *7*, 13139.
(52) Baldi, P. Autoencoders, unsupervised learning, and deep architectures. *Proceedings of ICML workshop on unsupervised and transfer learning* **2012**, 37−49.
(53) Dai, B.; Wang, Y.; Aston, J.; Hua, G.; Wipf, D. Connections with robust PCA and the role of emergent sparsity in variational autoencoder models. *Journal of Machine Learning Research* **2018**, *19*, 1573−1614.
(54) Wang, W.; Gómez-Bombarelli, R. Coarse-graining auto-encoders for molecular dynamics. *npj Computational Materials* **2019**, *5*, 125.
(55) Wang, D.; Tiwary, P. State predictive information bottleneck. *J. Chem. Phys.* **2021**, *154*, 134111.

https://doi.org/10.1021/acs.jctc.1c00143
J. Chem. Theory Comput. 2022, 18, 2341−2353