




RESEARCH ARTICLE | APRIL 09 2024

Interpretation of autoencoder-learned collective variables using Morse–Smale complex and sublevelset persistent homology: An application on molecular trajectories

Shao-Chun Lee ; Y Z  



J. Chem. Phys. 160, 144104 (2024)

<https://doi.org/10.1063/5.0191446>



Articles You May Be Interested In

Representations of energy landscapes by sublevelset persistent homology: An example with *n*-alkanes

J. Chem. Phys. (March 2021)

Additive energy functions have predictable landscape topologies

J. Chem. Phys. (April 2023)

Variational embedding of protein folding simulations using Gaussian mixture variational autoencoders

J. Chem. Phys. (November 2021)



The Journal of Chemical Physics

Special Topics Open for Submissions

[Learn More](#)

Interpretation of autoencoder-learned collective variables using Morse–Smale complex and sublevelset persistent homology: An application on molecular trajectories

Cite as: J. Chem. Phys. 160, 144104 (2024); doi: 10.1063/5.0191446

Submitted: 13 December 2023 • Accepted: 22 March 2024 •

Published Online: 9 April 2024



Shao-Chun Lee^{1,2} and Y Z^{1,2,3,a)}

AFFILIATIONS

¹ Department of Nuclear, Plasma, and Radiological Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA

² Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA

³ Department of Nuclear Engineering and Radiological Sciences, Department of Materials Science and Engineering, Department of Robotics, and Applied Physics Program, University of Michigan, Ann Arbor, Michigan 48105, USA

^{a)} Author to whom correspondence should be addressed: yzyz@umich.edu

ABSTRACT

Dimensionality reduction often serves as the first step toward a minimalist understanding of physical systems as well as the accelerated simulations of them. In particular, neural network-based nonlinear dimensionality reduction methods, such as autoencoders, have shown promising outcomes in uncovering collective variables (CVs). However, the physical meaning of these CVs remains largely elusive. In this work, we constructed a framework that (1) determines the optimal number of CVs needed to capture the essential molecular motions using an ensemble of hierarchical autoencoders and (2) provides topology-based interpretations to the autoencoder-learned CVs with Morse–Smale complex and sublevelset persistent homology. This approach was exemplified using a series of n-alkanes and can be regarded as a general, explainable nonlinear dimensionality reduction method.

Published under an exclusive license by AIP Publishing. <https://doi.org/10.1063/5.0191446>

INTRODUCTION

Given a set of molecular trajectories, is it possible to automatically identify physically significant collective variables (CVs), which not only expedite computer simulations but also assist human comprehension? Dimensionality reduction (DR) methods are pivotal in addressing this question, yet major concerns regarding their explainability remain unresolved. The principle of DR methods is to retain specific relationships between data instances. For example, principal component analysis (PCA)¹ and multidimensional scaling (MDS) are the two most common linear DR methods. The former preserves the covariance matrix of the data, and the latter preserves the pair-wise Euclidian distances between the data. For nonlinear correlations, nonlinear DR techniques,^{2,3} such as locally linear embedding (LLE), Isomap, diffusion map,^{2,4,5} and autoencoder, can be employed. LLE can be thought of as iteratively applying

PCA to sets of neighbors, assuming the manifolds are linear locally. The Isomap preserves the geodesic distances, while the diffusion map preserves the diffusion distances between neighbors.⁶ Among these nonlinear DR methods, the autoencoder stands out as a distinct method because it consists of an encoder and a decoder that are artificial neural networks. The encoder maps the N-dimensional data to a reduced M-dimensional CV space, which is represented by a bottleneck layer with m nodes,

$$CV = h(x), \quad (1)$$

where h represents the encoder and x represents a N-dimensional vector. The decoder reconstructs the N-dimensional samples from the CV space,

$$\hat{x} = \hat{h}(CV), \quad (2)$$

where \hat{x} represents the reconstructed N-dimensional vector and \hat{h} represents the decoder. The training of the autoencoder essentially minimizes the reconstruction error, L ,

$$L = (\hat{x} - x)^2. \quad (3)$$

Various studies have focused on using DR methods to identify CVs from molecular conformations.^{1,7–15} Brown *et al.* and Martin *et al.* studied the topology of the conformational space of a cyclo-octane and compared both linear and nonlinear DR methods, including PCA, LLE, Isomap, and autoencoder.^{13,15} Chen and Ferguson proposed an autoencoder-based scheme that facilitates the exploration of molecular conformations.^{16,17} One advantage of an autoencoder over other nonlinear DR techniques is that it contains a decoder component that provides an approximated one-to-one mapping from the CV space back to the original space.¹⁶ This facilitates the exploration of the CV space, where samples are less populated.

While autoencoder seems to be a robust nonlinear DR tool, two issues regarding the lack of “explainability” are yet to be addressed: (1) The optimal number of reduced dimensions, M_{opt} , is unknown and, thus, needs to be prescribed. (2) The physical meaning of the autoencoder-learned CVs is mostly obscure. To resolve the first issue, Chen *et al.* computed the fraction of variance explained by the autoencoder as a function of the number of CVs. They determined the optimal number of CVs at the point where adding one more CV does not significantly increase the fraction of variance, which is referred to as the “L-method.”^{17,18} However, the method does not offer a detailed explanation of how it retains important variance while filtering out unwanted variance. On the other hand, Glielmo *et al.* proposed an alternative approach. They evaluated the ratio (μ_i) of the distances of the two nearest neighbors of a data point i . The optimal intrinsic dimensionality of the data was obtained by maximizing the likelihood of all μ_i .¹⁹

As for the second issue, one possible way of mapping CVs to molecular motions is to use a circular autoencoder that contains pairs of bottleneck nodes with circular activation functions,²⁰ providing a mapping between CVs and the circular motions of the molecules. Although circular motions are common in molecules, one usually needs *a priori* knowledge about the molecules before training a circular autoencoder. Scholz and Vigário used hierarchical autoencoders (HAEs)²¹ and yielded a set of CVs ranked by the explained variances, yet it is challenging for humans to make connections between these hierarchical CVs and molecular conformations. Furthermore, imposing the ranking of CVs compromises the reconstruction quality.

On the other hand, one could hypothesize that if two independently trained autoencoders possess the same architecture (input, bottleneck nodes, output, etc.), they would retain identical relationships within the data. Consequently, these two autoencoders are expected to capture free energy landscapes (FELs) that are isomorphic and unaffected by either the architectures of the autoencoders or the stochasticity imposed in the training process. In other words, the key to explainability may lie in the topology of the reduced CV space. We noticed that Manuchehrfar *et al.* highlighted the sensitivity of the topological structure in the reduced space due to the process of DR.²² However, our perspective suggests that this sensitivity primarily arises from the engineering of the feature

space and the distance measures. In essence, we contend that the topological structure of the autoencoder-learned CV space remains consistent when the feature space and distance measures are identical. In fact, Glielmo *et al.* presented a comparative framework for evaluating the information content between different distance measures.²³ This approach can be particularly useful for assessing the quality of autoencoder architectures, including feature engineering and loss function design. Motivated by a recent work that revealed such a topological representation of a real-valued function with sublevelset persistent homology,²⁴ in this paper, we adopted both sublevelset persistent homology and, additionally, Morse–Smale complex to give clear physical meanings to the autoencoder-learned CVs. Before presenting the framework proposed in this study, we first introduce the topological data analysis tools used in this study: Sublevelset persistent homology and the Morse–Smale complex.

Sublevelset persistent homology: Similar to a contour tree,²⁵ sublevelset persistent homology is a mathematical technique that helps us understand the topology of the data. For point data, persistent homology describes how the data points are connected to other points within a specific cutoff range, which is often referred to as the persistent level. By examining variations across different persistent levels, persistent homology reveals hidden topological structures, such as clusters and rings, within the data. Manuchehrfar *et al.* analyzed the topological changes of the high-dimensional dynamic probability surface and identified the locations of probability peaks, indicating stable states, and connecting ridges, implying reaction pathways, using persistent homology.²² Hiraoka *et al.* analyzed the atomic structures of amorphous solids and uncovered hierarchical atomic ring structures using persistent homology.²⁶ If the data are represented by a scalar field, a sublevelset is the subset of a scalar field where every point in this subset has a function value no more than a given real value, r . In the context of FEL, a sublevelset can be thought of as a set of configurations accessible by the system at a given temperature. The sublevelset persistent homology represents the topological variation of the sublevelset as r varies.

Morse–Smale complex: The Morse–Smale complex is a mathematical technique that partitions a scalar function—a Morse function—into regions in which the functional values vary monotonically based on the critical points.²⁷ A smooth function f that maps a smooth, compact, p -dimensional manifold onto a real number, i.e., $f: \mathcal{M} \mapsto \mathbb{R}$, is Morse if the Hessian matrix evaluated on any critical point x is not singular. Figure 1 shows an example of how Morse–Smale complexes partition a two-dimensional scalar field. Each Morse–Smale complex corresponds to a pair of local maximums and local minimums on its boundary, as shown in Fig. 1(b). The partition can also be performed in two other ways: (1) A set of Morse–Smale complexes that correspond to the same local maximum is referred to as an ascending manifold [Fig. 1(c)]. (2) A set of Morse–Smale complexes that correspond to the same local minimum is referred to as a descending manifold [Fig. 1(d)]. Monotonically increasing/decreasing functional value along the scalar field terminates at the same local maximum/minimum, starting from any point in the ascending/descending manifold. The Morse–Smale complex has been applied in various scientific fields. Canzals *et al.* analyzed the relevance of stable and unstable patches for docking sites on the surface of a molecule with Morse–Smale decomposi-

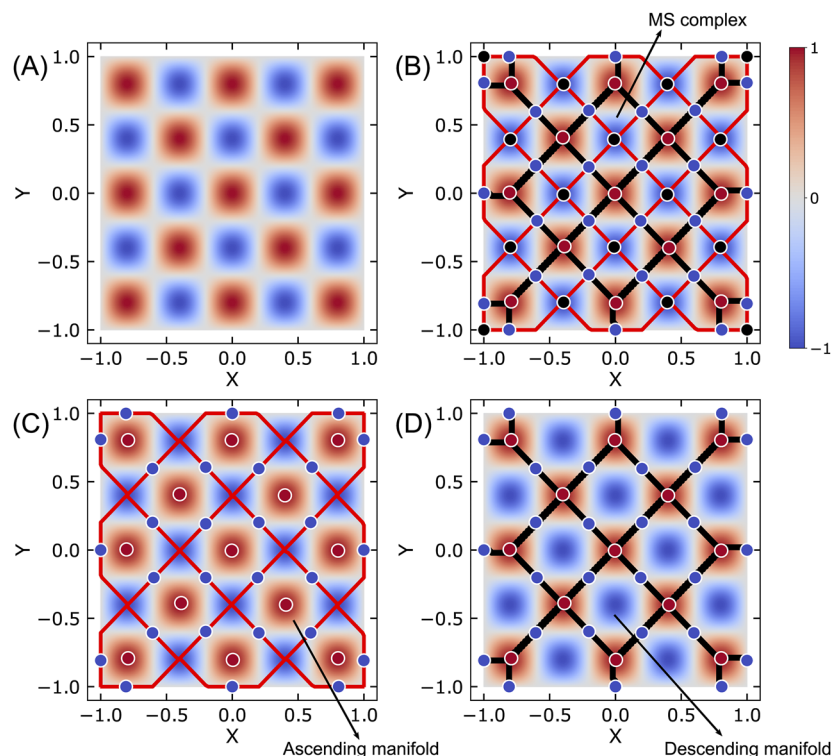


FIG. 1. Illustration of Morse–Smale complex decomposition. (a) $f(x, y) = \cos(2.5\pi x) \cos(2.5\pi y)$. A Morse–Smale complex decomposition shows (b) all the Morse–Smale complexes, (c) only the ascending manifolds, and (d) only the descending manifolds. [Red dots: local maxima, blue dots: saddle points, black dots: local minima, red lines: boundaries of the ascending manifolds, and black lines: boundaries of the descending manifolds. The local minima in both (c) and (d) are removed for clarity since they are irrelevant to the application of this study.]

tion;²⁸ Laney *et al.* studied the structure of turbulence by analyzing the evolution of topological features constructed by Morse–Smale complexes.²⁹ In principle, Morse–Smale complex decomposition can be applied to a scalar field of arbitrary dimension.

The objective of this study is to develop a framework that can automatically extract important CVs from molecular trajectories using neural-network-based autoencoders and analyze the topological features within the learned CV space using sublevelset persistent homology and the Morse–Smale complex as a form of explanation for the CVs. To focus on the development of the methodology, we selected three simple molecules: butane, pentane, and cyclohexane, whose CVs are known as the dihedral angles. Figure 2 shows the framework of this study. First, we performed dimensionality reduction on all-atom MD molecular trajectories using HAEs and estimated the contribution of the variance from each hierarchical CV. The optimal number of CVs, denoted as M_{optim} , was determined by removing CVs whose contributing variances are below a certain physically meaningful threshold or exhibit large noise due to the stochasticity imposed during the training of HAE. Next, we trained a traditional autoencoder with M_{optim} and estimated the FEL on the learned CV space. Finally, we acquired a topological representation of this autoencoder-learned FEL using sublevelset persistent homology and Morse–Smale complex decomposition, which could

be treated as a form of explanation for the CVs generated by autoencoders.³⁰

ALL-ATOM MD SIMULATIONS

We performed MD simulations of the gas phase alkanes (butane, pentane, and cyclohexane) using LAMMPS.³⁰ The OPLS-AA forcefield³¹ was used. We placed one molecule in the cubic simulation box with the box size set to 10 nm without a periodic boundary condition. The positions, momentum, and angular momentum were zeroed at every step. The time step was set to 1 fs. The temperature was maintained using a Nosé–Hoover thermostat (NVT).³² All cases are equilibrated for 0.1 ns before production. Other details are tabulated in Table I.

PREPROCESSING OF THE TRAINING DATA

In this study, we selected only the backbone degrees of freedom for simplification. Translational and rotational symmetries are removed by constructing a local reference coordination system from three selected atoms; thus, the dimension of each sample is

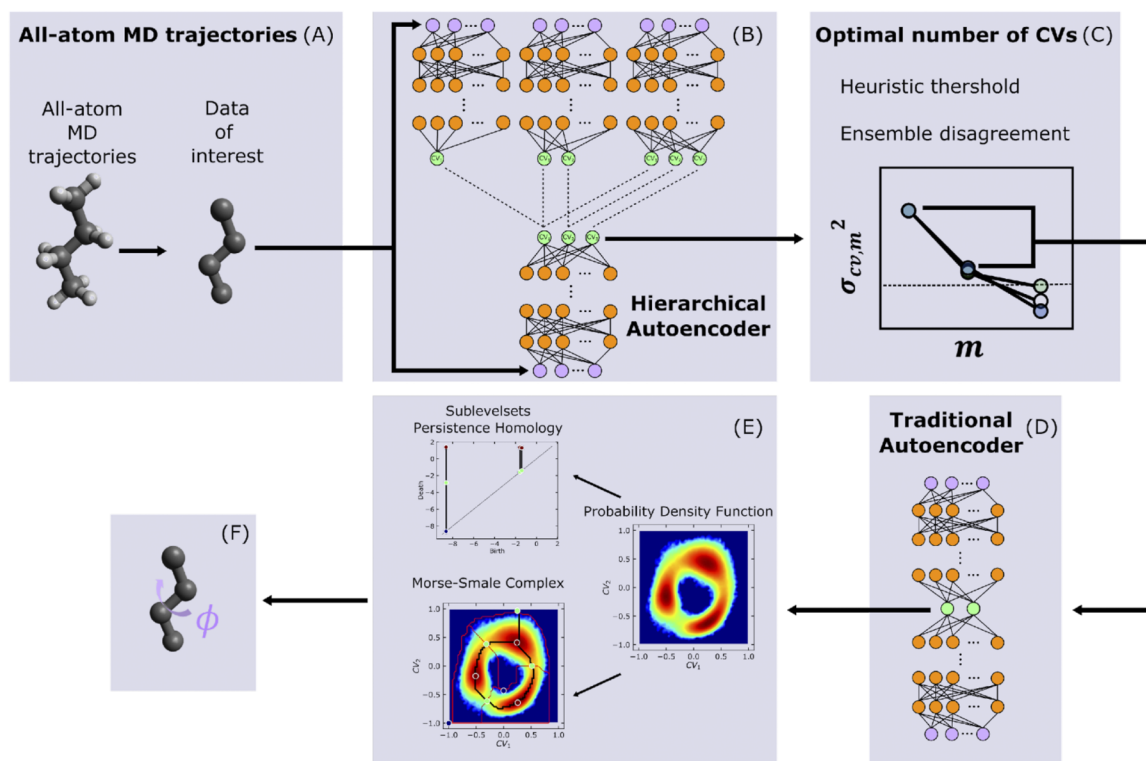


FIG. 2. Framework of topology-based interpretation of autoencoder-learned collective variables. (a) All-atom MD trajectories. (b) Architecture of a hierarchical autoencoder (HAE). (c) Determination of the optimal number of CVs. (d) Traditional autoencoder with the optimal number of CVs. (e) Sublevelset persistent homology and Morse–Smale complex decomposition of the autoencoder-learned CV space. (f) Automatically extracted transition pathways.

TABLE I. Details of MD simulations.

	Butane	Cyclohexane	Pentane
Temperature (K)	400	600	400
Simulation time (ns)	5	50	5
Trajectories are dumped every No. of time steps	1	10	1
Total number of frames	5 000 000		

$N = 3N_a - 6$, where N_a is the number of backbone atoms. Assuming the three atoms are A, B, and C, the local reference coordination system is defined as follows:

$$\hat{\mathbf{e}}_x = \frac{\mathbf{r}_{AB}}{|\mathbf{r}_{AB}|}, \quad \hat{\mathbf{e}}_z = \frac{\hat{\mathbf{e}}_x \times \mathbf{r}_{BC}}{|\hat{\mathbf{e}}_x \times \mathbf{r}_{BC}|}, \quad \hat{\mathbf{e}}_y = \frac{\hat{\mathbf{e}}_z \times \hat{\mathbf{e}}_x}{|\hat{\mathbf{e}}_z \times \hat{\mathbf{e}}_x|}, \quad (4)$$

$$\begin{bmatrix} \hat{\mathbf{e}}_x \\ \hat{\mathbf{e}}_y \\ \hat{\mathbf{e}}_z \end{bmatrix} \mathbf{R} = \mathbf{R}', \quad (5)$$

where $\mathbf{r}_{AB} = \mathbf{r}_B - \mathbf{r}_A$ is the position vector from A to B, and \mathbf{R} and \mathbf{R}' are the coordinates of the atoms before and after being

transformed into the local reference coordination system. The first three carbons of linear alkanes and three of the adjacent carbons in cyclo-hexane are selected for constructing the local reference coordination.

TRAINING OF THE AUTOENCODERS

Pytorch 1.10.0³³ with CUDA 10.2³⁴ support was used in this study. The architecture of our prototypical HAE consists of one encoder and M decoders, where M varies from 1 to N , as illustrated in Fig. 2(b). Both the encoder and decoder are five-layer, fully connected, feedforward neural networks. For the encoder part, the input layer contains N neurons, followed by three hidden layers with 64, 256, and 64 neurons, respectively. The fifth layer contains M nodes, which correspond to the M hierarchical CVs. As for the decoder, the HAE consists of M decoders, which have three hidden layers with 64, 256, and 64 neurons as well. The m th decoder reconstructs the input data from CV_1 to CV_m . The loss function is defined as the sum of reconstruction errors from all the decoders,

$$Loss_{HAE} = \sum_m^M \chi_{1 \sim m}^2, \quad (6)$$

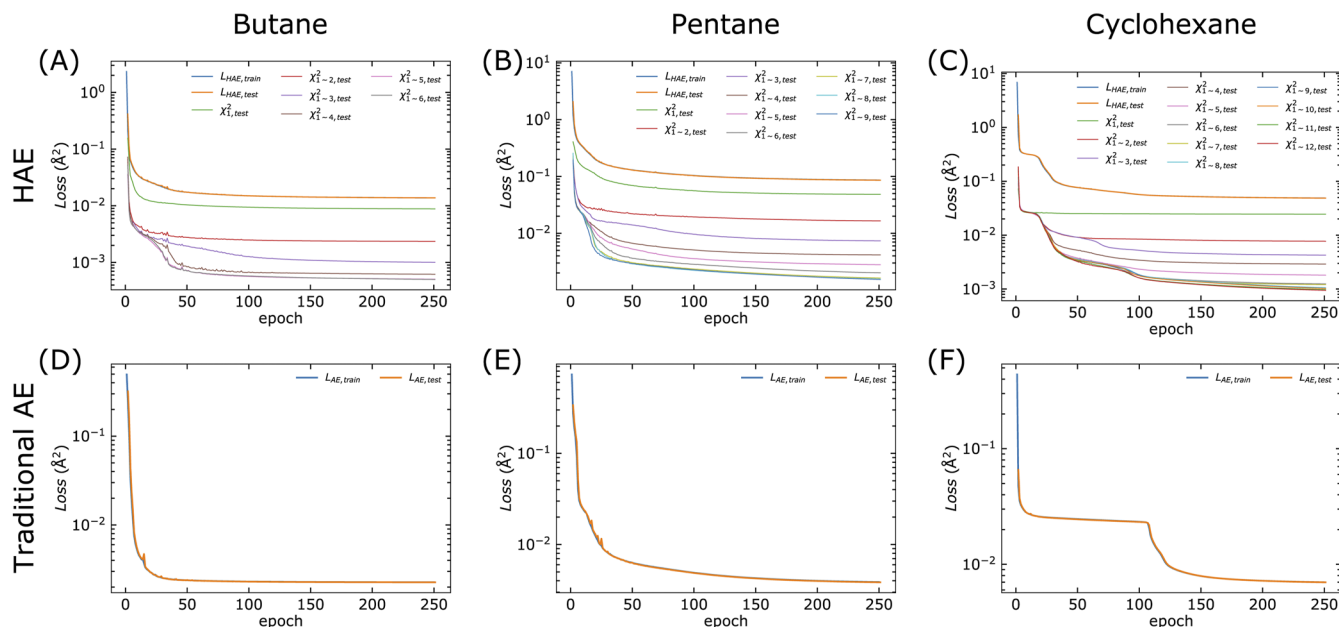


FIG. 3. Learning curves of training. (a) and (d) Butane. (b) and (e) Pentane. (c) and (f) Cyclohexane. (a)–(c) The loss function of HAEs as a function of training epoch. (d)–(f) The loss function of traditional autoencoders as a function of training epoch. The traditional autoencoders are trained with $M = M_{\text{optim}}$.

where $\chi^2_{1 \sim m}$ indicates the mean squared error between the input samples and the reconstructed samples using the first m CVs. The traditional autoencoder has the same architecture as HAE, except that there is only one decoder that reconstructs the input from all the CVs, as shown in Fig. 2(d). The loss function of the traditional autoencoder is thus defined as

$$\text{Loss}_{\text{AE}} = \chi^2_{1 \sim M}. \quad (7)$$

The Tanh activation function was used for all layers. We trained the network for a maximum of 250 epochs with the Adam optimizer³⁵ with an exponentially decaying learning rate of 10^{-3} to 10^{-5} . The resulting learning curves are shown in Fig. 3. Note that the number of CVs for the traditional autoencoders is set to M_{optim} , which was obtained in the subsequent section.

DETERMINATION OF THE OPTIMAL NUMBER OF CVS

To determine the optimal number of CVs, denoted as M_{optim} , we formulated an estimation of the variance contributed by each hierarchical CV. The total variance, σ^2_{total} , of a set of samples, x , can be expressed as

$$\sigma^2_{\text{total}} = \sum_{i=1}^N \text{var}(x_i) + \sum_{i < j}^N 2\text{cov}(x_i, x_j), \quad (8)$$

where $\text{var}(x_i)$ is the variance of the i th variable of x and $\text{cov}(x_i, x_j)$ is the covariance of the i th and j th variables. Assum-

ing the CVs learned by HAE are independent, σ^2_{total} can also be written as

$$\sigma^2_{\text{total}} = \left(\sum_{m'=1}^m \sigma^2_{\text{cv},m'} \right) + \chi^2_{1 \sim m}, \quad (9)$$

where m' is a dummy variable, $\sigma^2_{\text{cv},m'}$ is the variance contributed by each CV, and $\chi^2_{1 \sim m}$. Hence, each σ^2_{cv} can be computed through an iterative process,

$$\sigma^2_{\text{cv},1} = \sigma^2_{\text{total}} - \chi^2_1, \quad (10)$$

$$\sigma^2_{\text{cv},2} = \sigma^2_{\text{total}} - \sigma^2_{\text{cv},1} - \chi^2_{1 \sim 2}, \quad (11)$$

⋮

$$\sigma^2_{\text{cv},M} = \sigma^2_{\text{total}} - \left(\sum_{m=1}^{M-1} \sigma^2_{\text{cv},m} \right) - \chi^2_{1 \sim M}. \quad (12)$$

Herein, we used two approaches to estimate M_{optim} . The first approach is to set a physically meaningful cutoff variance σ^2_{cutoff} and so that $\sigma^2_{\text{cv},M_{\text{optim}}} > \sigma^2_{\text{cutoff}} > \sigma^2_{\text{cv},M_{\text{optim}}+1}$. If we consider the vibration of the bonds as thermal noise, it is reasonable to define σ^2_{cutoff} by some pre-factor times the variance of the bond length from the sample, i.e., $\sigma^2_{\text{cutoff}} = f \sigma^2_{\text{bond}}$. Note that this heuristically set threshold is human understandable and can be defined differently depending on the level of resolution of the molecule one wants to probe.

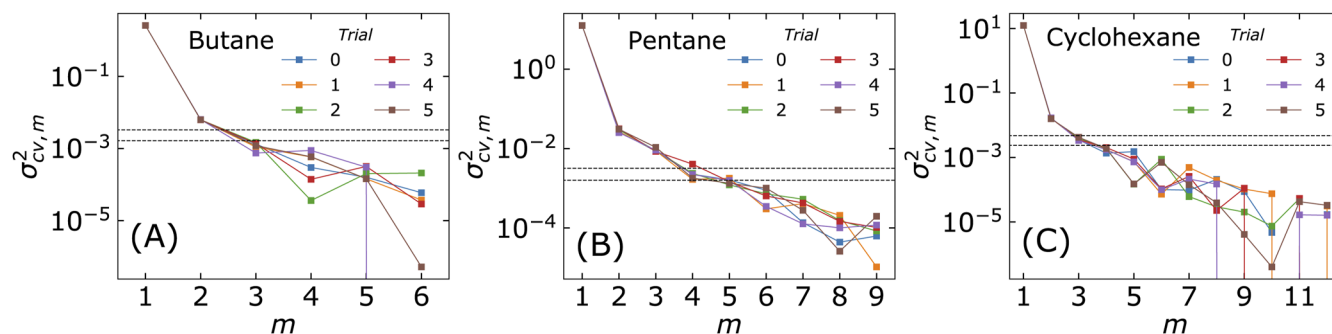


FIG. 4. Determination of the optimal number of CVs through hierarchical variances from the first HAE ensemble, containing six trials, all of which had $M = N$. (a) Butane, (b) pentane, and (c) cyclohexane. Dotted horizontal lines indicate $f\sigma_{bond}^2$, where $f = 1$ and 2.

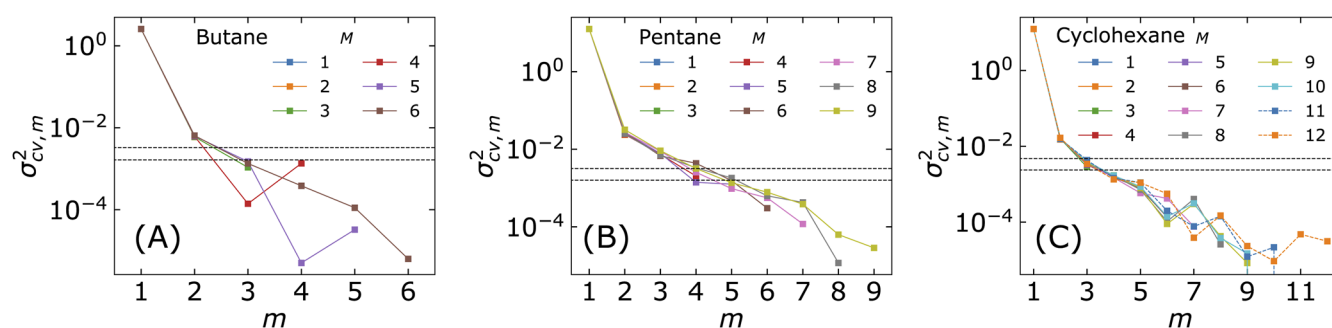


FIG. 5. Determination of the optimal number of CVs through hierarchical variances from the first HAE ensemble, containing N trials, each with $M = 1 \sim N$. (a) Butane, (b) pentane, and (c) cyclohexane. Dotted horizontal lines indicate $f\sigma_{bond}^2$, where $f = 1$ and 2.

The second approach is to examine the reproducibility of each σ_{cv}^2 by repeating the training of HAE multiple times. Each training session is a “trial” that randomly initializes the weights and biases of the neural network and randomly shuffles the training batches. For all three cases in this study, we trained two ensembles.

The first ensemble consists of six trials, all of which had $M = N$; the second ensemble consists of N trials, each with $M = 1 \sim N$. The resulting σ_{cv}^2 from the two ensemble are shown in Figs. 4 and 5, respectively. We also plotted $f\sigma_{bond}^2$ with f ranging from 1 to 2, indicating which hierarchical CV started to explain the bond vibrations. We found that σ_{cv}^2 showed high reproducibility and a strictly descending order for the first several CVs before they started to disagree from different trials. We used “ensemble uncertainty” to describe this disagreement in σ_{cv}^2 , and we attributed its abrupt growth to the fitting of thermal noises by the HAEs during learning. For the three cases in this study, we found that setting f to 2 is a good choice as it avoids those deviations. The M_{optim} for the cases of butane, pentane, and cyclohexane are selected as 2, 3, and 2 dimensions, respectively. We then trained traditional autoencoders with the obtained M_{optim} . The CVs learned by these traditional autoencoders are the ones we attempted to give an explanation to. Reasons for giving explanations to traditional autoencoders instead of HAEs were discussed in the following section.

COMPARISON BETWEEN HAES AND TRADITIONAL AUTOENCODERS

To understand the differences between how a HAE and a traditional autoencoder behave, we first compared their sample distribution in the CV space. Figure 6 compares the sample distribution in the CV space ($M = 2$) by a HAE and a traditional autoencoder in the case of gas phase butane. We found that HAE shows a triangular distribution of the sample, while the traditional autoencoder shows a circular distribution. It is no surprise that both autoencoders showed a distribution of a one-dimensional loop that corresponds to the dihedral angle of butane. This isomorphism indicates that both autoencoders have learned identical relationships between the samples. However, the HAE distorted the one-dimensional loop where the highly elongated part corresponds to the cis conformation where the dihedral angle is close to 0. This is because imposing the importance ranking on a set of non-periodic CVs forces CV_1 to explain as much variance as possible. The consequence is that the HAE gave up explaining less populated samples with CV_1 because those high energy states contributed less to the reconstruction error solely from CV_1 . This can be verified by examining the χ_1^2 , as shown in Fig. 6(e). On the other hand, the traditional autoencoder gave a rather undistorted distribution of the samples as the dihedral angle is evenly projected on CV_1 and CV_2 .

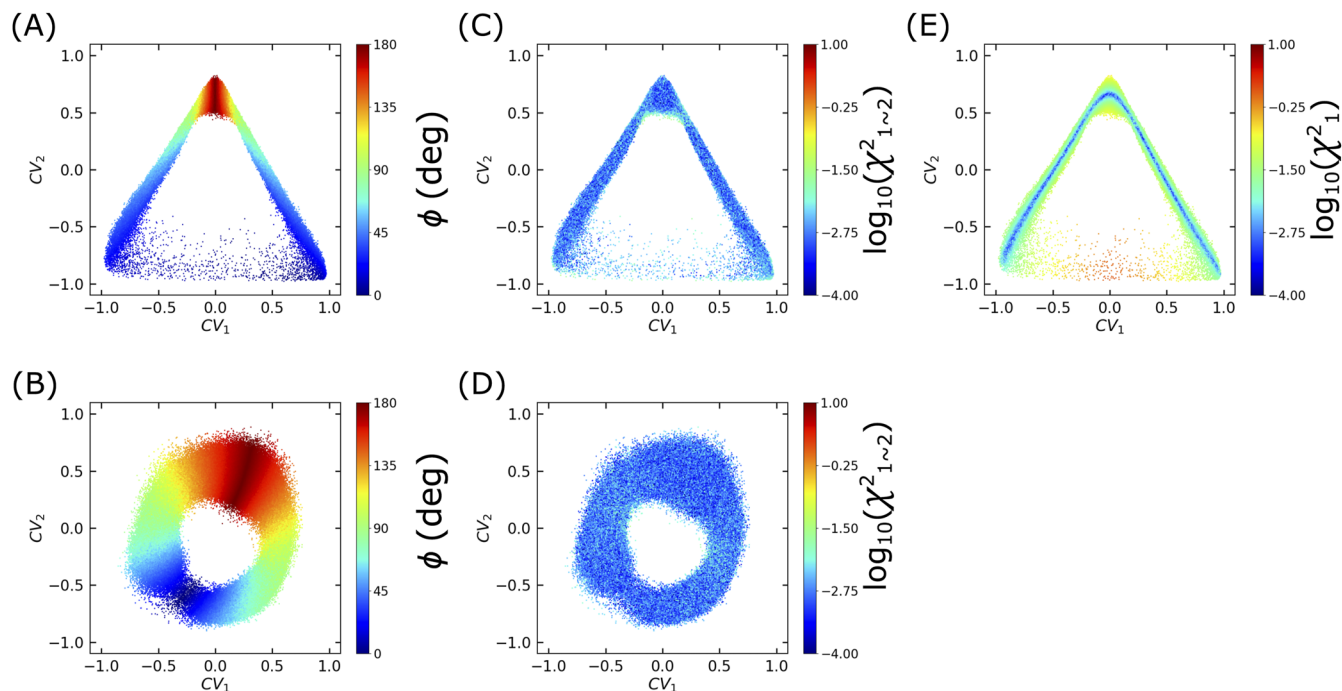


FIG. 6. Comparison of a HAE and a traditional autoencoder in the case of butane in the gas phase. Samples encoded by (a), (c), and (e) the HAE and by (b) and (d) the traditional autoencoder. The samples are colored by their (a) and (b) dihedral angles, (c) and (d) $\log_{10}(\chi^2_{1\sim 2})$, and (e) $\log_{10}(\chi^2_1)$.

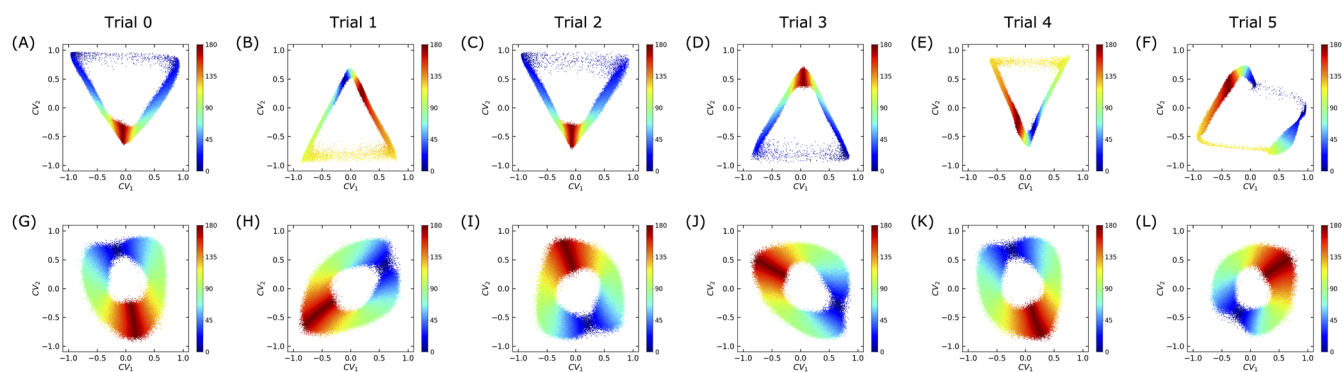


FIG. 7. Distribution of the samples embedded in the CV space from six trials. (a)–(f) Six trials of HAEs. (g)–(l) Six trials of traditional autoencoders. The samples are color coded with the dihedral angle.

Figure 7 shows the embedding of the samples with six trials of HAEs and traditional autoencoders. We found that some of the trial HAEs elongated the distribution of the cis conformation while others elongated the gauche conformations (the last trial HAE elongated both conformations). We attributed this instability to the stochasticity originated in the initial weights and biases of the neural networks and the training batches during training. As for traditional autoencoders, the CVs corresponded to different projections of the dihedral angles among different trails, but the distribution of the samples

remained undistorted and unaffected by the stochasticity during training.

Table II shows the comparison of the reconstruction quality between HAE and traditional autoencoders by examining their $\chi^2_{1\sim M_{\text{optim}}}$. It is shown that the penalty imposed by the HAE resulted in an increase in the reconstruction error by $\sim 4\%$, $\sim 91\%$, and $\sim 10\%$ in the case of butane, pentane, and cyclohexane, respectively. The degraded reconstruction quality is primarily contributed by the samples in the highly distorted regions.

TABLE II. Comparison of the reconstruction quality between HAE and traditional autoencoders.

	$\chi^2_{1 \sim M_{\text{optim}}} (\text{\AA}^2)$		$\frac{H-T}{T}$
	HAE (H)	Traditional autoencoder (T)	
Butane	2.4×10^{-3}	2.3×10^{-3}	0.04
Pentane	7.3×10^{-3}	3.8×10^{-3}	0.91
Cyclohexane	7.7×10^{-3}	7.0×10^{-3}	0.10

As mentioned in the previous section, HAEs and traditional autoencoders learned the same topology about the samples, but they exhibited a systematic distinction in the distribution of the samples due to the restrictions imposed by HAE. The distortion within the autoencoder-learned FELs can be quantified by the Jacobian determinant of the decoder,

$$J = \begin{bmatrix} \frac{\partial \hat{h}_1}{\partial CV_1} & \cdots & \frac{\partial \hat{h}_1}{\partial CV_M} \\ \vdots & \ddots & \vdots \\ \frac{\partial \hat{h}_N}{\partial CV_1} & \cdots & \frac{\partial \hat{h}_N}{\partial CV_M} \end{bmatrix}, \quad (13)$$

$$J = \sqrt{\det[J^T J]}, \quad (14)$$

where \hat{h}_i is the i th output of the decoder. Figure 8 shows the value of J as a function of CVs. We found that HAE gives a very small J at the cis conformation, indicating that a unit length in the CV space corresponds to a small variation near the cis conformation in the reconstructed space.

Combining the comparisons of reconstruction error and the distribution of the samples, the traditional autoencoder seems to learn a rather undistorted CV space. Therefore, we focus on explaining the traditional autoencoders instead of HAEs using sublevelset persistent homology analysis and Morse–Smale complex decomposition.

SUBLEVELSET PERSISTENT HOMOLOGY AND MORSE–SMALE COMPLEX DECOMPOSITION

The Topology ToolKit (TTK)^{36,37} was used to compute the sublevelset persistent homology and Morse–Smale complex decomposition. We defined the probability density (ρ) using the following equation:

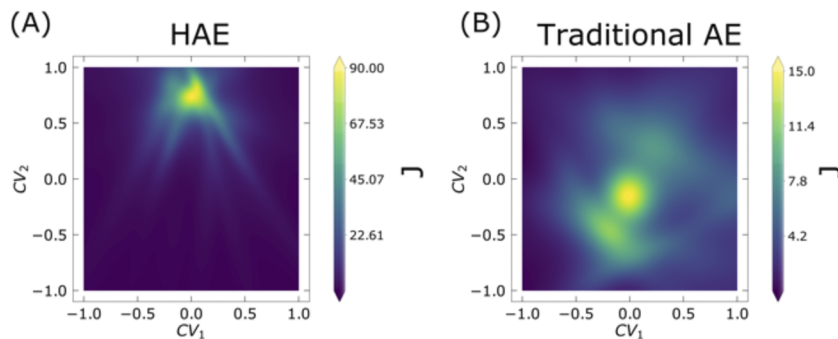
$$\rho(\xi'_1, \dots, \xi'_{M_{\text{optim}}}) d\xi'_1 \dots d\xi'_{M_{\text{optim}}} = \frac{S}{S_{\text{total}}}, \quad (15)$$

where ξ_i is the variable along the i th CV, S is the number of samples within a high-dimensional cube ranging from $(\xi_1, \dots, \xi_{M_{\text{optim}}})$ to $(\xi_1 + \delta\xi_1, \dots, \xi_{M_{\text{optim}}} + \delta\xi_{M_{\text{optim}}})$, and S_{total} is the total number of samples. This is essentially creating a histogram for the samples in the CV space. This is a naive density estimation approach and can be sensitive to noise in higher dimensions, yet robust density estimation techniques can be found in the literature, such as point-adaptive k-nearest neighbors.¹⁹ The probability density is proportional to the Boltzmann factor at constant temperature, volume, and number of particles,

$$\rho(\xi_1, \dots, \xi_{M_{\text{optim}}}) \propto e^{-F(\xi_1, \dots, \xi_{M_{\text{optim}}})/k_B T}, \quad (16)$$

where F is the free energy, k_B is the Boltzmann constant, and T is the temperature. We then performed sublevelset persistent homology analysis and Morse–Smale complex decomposition on the logarithm of the probability density, which are proportional to the autoencoder-learned FELs.

Figure 9 displays the results of topological representation in the case of butane. Figure 9(a) displays the Morse–Smale complex decomposition where critical points and ascending/descending manifolds are visualized, and Fig. 9(b) shows the birth–death plot representation of the persistent diagram. The birth–death plot visualizes the emergence and disappearance of the topological features in the sublevelset persistent homology and Morse–Smale complex. To associate this birth–death plot with the topological features of the FEL, we can imagine a sea level rising and gradually flooding the FEL. A critical point of index n , denoted as C_n , with a functional value of V_0 will be presented at (birth, death) = (V_0, V_0) ,

**FIG. 8.** Comparison of the Jacobian determinant spanned by the CV space in the case of butane. (a) HAE. (b) Traditional autoencoder.

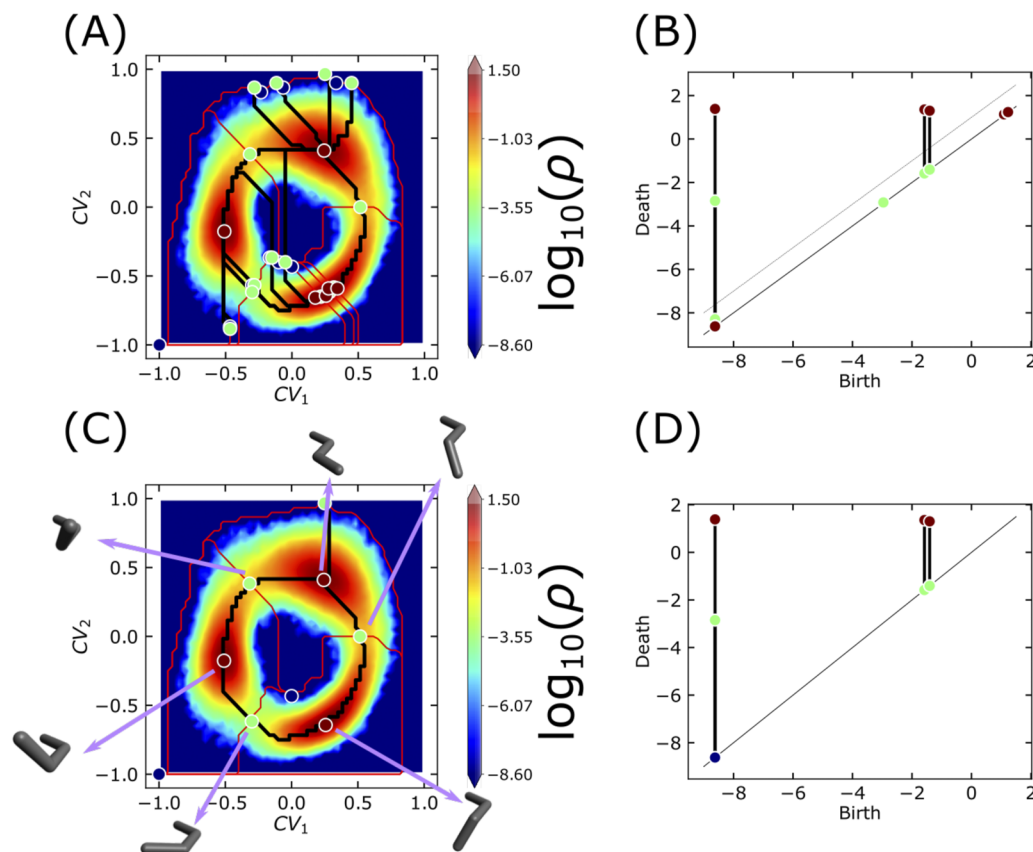


FIG. 9. Topological representation and filtration of autoencoder-learned FELs for butane. (a) Morse–Smale complex decomposition and (b) persistent diagram before topological filtration. (c) Morse–Smale complex decomposition and (d) persistent diagram after topological filtration. The dotted line in (b) indicates the filtration threshold. Critical points with indices 0, 1, and 2 are marked as blue, green, and red circles. The boundaries of the ascending manifolds are marked as red solid lines. The boundaries of the descending manifolds are marked as black solid lines.

indicating the “birth” of C_n . A vertical bar will be attached on top of C_n and ends at another critical point of index $n + 1$, denoted as C_{n+1} , with a functional value of V_1 at (birth, death) = (V_0, V_1) , indicating the “death” of C_n . The length of the vertical bar indicates the “persistence” of the critical point pair between C_n and C_{n+1} . This critical pair implies a pairing between two neighboring transitions or stable states that share the same Morse–Smale complex. In Fig. 9(a), we saw multiple critical points identified within the same transition/stable state, even though butane should only have three transitions and three stable states. This phenomenon arises due to the noise induced during probability density estimation, which generates multiple low persistence critical pairs. We applied topological filtration to filter out low persistence critical pairs. Figures 9(c) and 9(d) show the Morse–Smale complex and persistent diagram after the removal of low persistence critical pairs as they are caused by statistical noise. Critical points of index 2 (red circles) and 1 (green circles) correspond to the stable states and transition states, respectively. The boundaries of the ascending manifolds (red solid lines) partition the CV space into

regions of inherent structures, while the boundaries of the descending manifolds (black solid lines) are associated with the minimum energy paths between stable states. Topological filtration of the Morse–Smale complexes allows us to eliminate physically insignificant thermal noise while retaining crucial topological states of the FELs and their connections that are subsequently used for the reconstruction of real-space molecular motions along the CVs (see Movie S1). However, we acknowledge that the persistence threshold for topological filtration is a parameter that necessitates manual tuning, and a rigorous selection protocol has not been established in this study. Automating this process could be a valuable area for future research.

To further demonstrate the success of this method, we automatically extracted the CVs and reconstructed the real-space molecular motions along the CVs for gas phase pentane and cyclohexane. The Morse–Smale complex and persistent diagram before and after topological filtration are displayed in Figs. 10 and 11, respectively. Without topological filtration, multiple critical points and Morse–Smale complexes were identified due to noise induced in density estima-

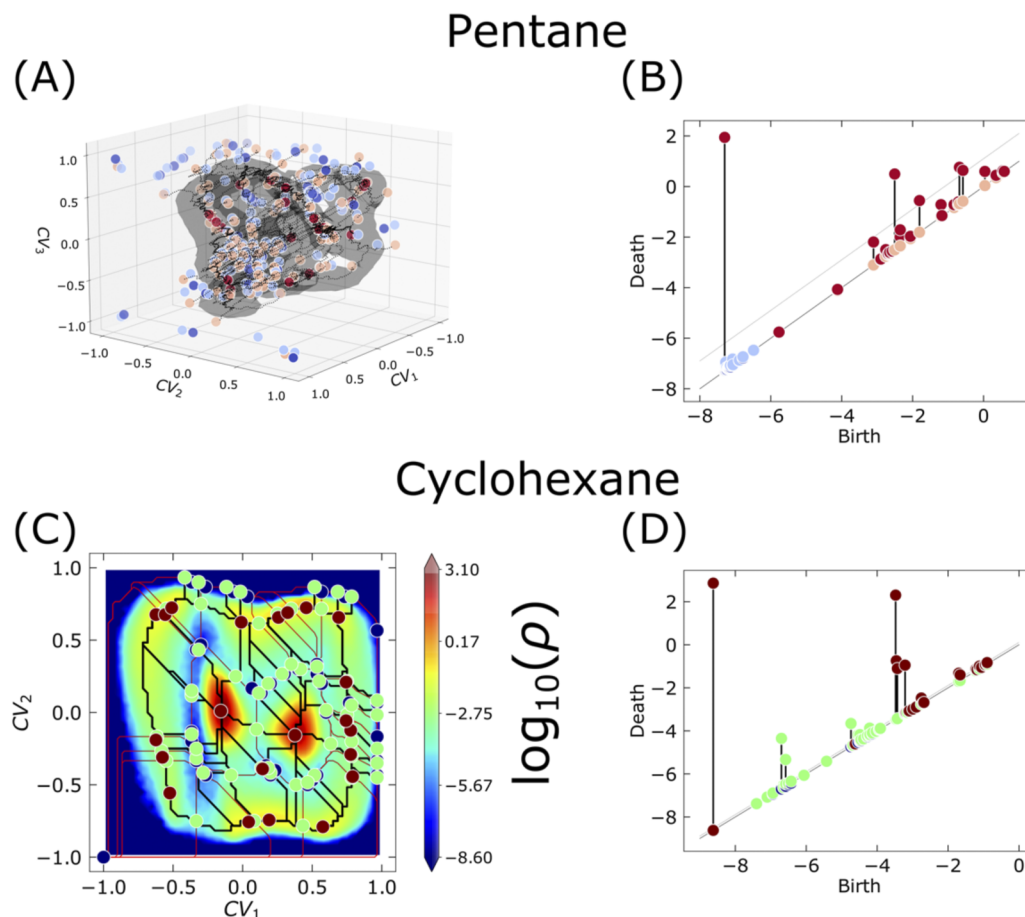


FIG. 10. Topological representation of the autoencoder-learned FELs without topological filtration for (a) and (b) pentane and (c) and (d) cyclohexane. The results after the topological filtration are shown in Fig. 11. In the case of pentane, critical points with indices 0, 1, 2, and 3 are marked as blue, light blue, light red, and red circles. The boundaries of the descending manifolds are marked as black dotted lines. Irrelevant boundaries are removed for clarity. The volume inside the iso-surface in (a) contains 98% of the samples. In the case of cyclohexane, the color-codes of the circles and lines are the same as in Fig. 9.

tion. Figures 11(a) and 11(b) show the Morse–Smale complex and persistent diagram in the case of pentane after topological filtration. Movie S2 provides a movie for viewing Fig. 11(a) from different angles. Two primary 1-dimensional pathways were identified: (1) The rotation of the first dihedral angle while the second dihedral angle is in trans state (see Movie S3). (2) The rotation of the second dihedral angle while the first dihedral angle is in trans state (see Movie S4). The two 1-dimensional pathways correspond to the boundary of the Morse–Smale complexes associated with the critical points of index 1 (1-saddles) and 2 (local maxima). Figures 11(c) and 11(d) show the results for cyclohexane; the two mirror symmetric transition pathways between chair, half-chair, twisted boat, and boat conformations are well identified (see Movies S5 and S6). Furthermore, we observed a rare transition pathway connecting two chair states via a co-planer state near the center of the CV space (see Movie S7). These movies intuitively disclose the physical meaning of the otherwise elusive CVs.

To further illustrate the effect of the penalty imposed by the HAE, we compared the resulting Morse–Smale complex decomposition in gas phase butane, as illustrated in Fig. 12. The transition pathways identified in the CV space from the HAEs were highly distorted. Nearly 1/3 of the transition pathway was associated with the cis conformation with trivial conformational change. On the other hand, the transition pathways identified in the CV space of the traditional autoencoder distribute the conformational variation evenly. Note that Fig. 12(b) is identical to Fig. 9(c). We also visualized the sample distribution in the case of pentane with traditional autoencoders (Movie S8: color coded with the first dihedral angle, and Movie S9: color coded with the second dihedral angle) and HAE (Movie S10: color coded with the first dihedral angle, and Movie S11: color coded with the second dihedral angle). From the sample distribution, we found that HAE exhibited a highly distorted CV space, resulting in a large variance in the probability density estimation and failure in Morse–Smale complex decomposition.

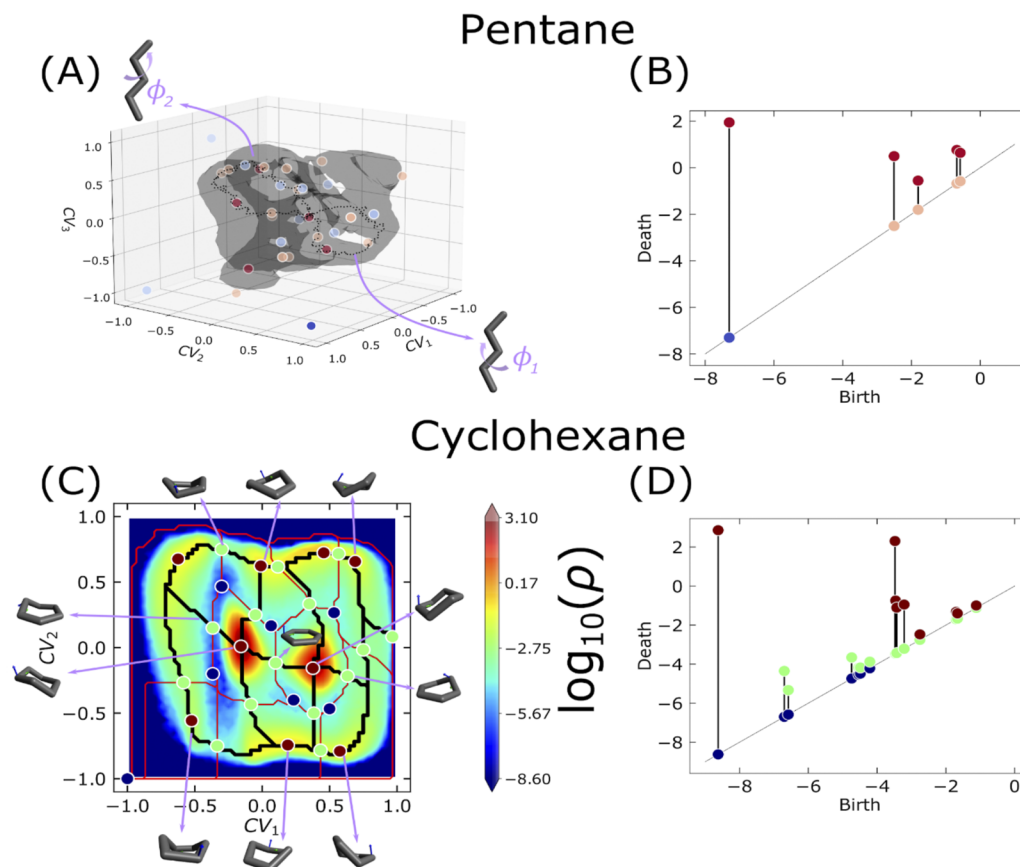


FIG. 11. Topologically filtered representation of the autoencoder-learned FELs for (a) and (b) pentane and (c) and (d) cyclohexane. In the case of pentane, critical points with indices 0, 1, 2, and 3 are marked as blue, light blue, light red, and red circles. The boundaries of the descending manifolds are marked as black dotted lines. Irrelevant boundaries are removed for clarity. The volume inside the iso-surface in (a) contains 98% of the samples. In the case of cyclohexane, the color-codes of the circles and lines are the same as in Fig. 9.

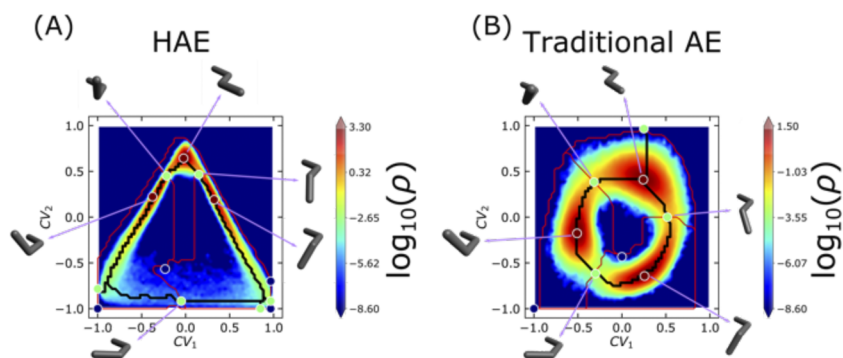


FIG. 12. Comparison of the resulting Morse–Smale complex decomposition in the case of gas phase butane. (a) HAE. (b) Traditional autoencoder.

CONCLUSION

In this paper, we present a general topology-based, explainable nonlinear DR method applicable to molecular trajectories. The optimal number of CVs was found via two approaches: (1) Apply the heuristic threshold, which holds physical significance, to hierarchical variances, and (2) identify the abrupt rise in the ensemble uncertainty of HAEs, which provides insights into the upper limit of dimensions not associated with thermal noise. The autoencoder-learned FELs are converted into topological representations, including sublevelset persistent homology and the Morse–Smale complex, which are invariant to the stochasticity imposed during the training of the neural networks. The former describes how the topology of the accessible states varies with free energy and allows the topological filtration to eliminate insignificant local motions while preserving the global topology, while the latter provides a connection between autoencoder-learned CVs and real-space molecular motions. The computations of both sublevelset persistent homology and the Morse–Smale complex are expensive; nevertheless, the method presented here can be applied to other molecular trajectories in principle.

SUPPLEMENTARY MATERIAL

The supplementary material includes high resolution figures and movies mentioned in the paper (Movies S1–S11).

ACKNOWLEDGMENTS

This work was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Materials Sciences and Engineering Division, under Award No. DE-SC0024447. It is also motivated by a previous National Science Foundation project under Grant No. 1934725, which provided the support for some initial attempts of this work. We are in debt to Professor Hendry Adams, Markus Pflaum, and Aurora Clark for their introduction of the topology tools to us and inspiring discussions with them throughout the NSF project.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Shao-Chun Lee: Conceptualization (lead); Data curation (lead); Formal analysis (lead); Investigation (lead); Methodology (lead); Project administration (equal); Software (lead); Validation (lead); Visualization (lead); Writing – original draft (lead); Writing – review & editing (equal). **Y Z:** Funding acquisition (lead); Project administration (equal); Resources (lead); Supervision (lead); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available within the article and its supplementary material.

REFERENCES

- ¹A. Altis, P. H. Nguyen, R. Hegger, and G. Stock, “Dihedral angle principal component analysis of molecular dynamics simulations,” *J. Chem. Phys.* **126**, 244111 (24) (2007).
- ²A. L. Ferguson, A. Z. Panagiotopoulos, I. G. Kevrekidis, and P. G. Debenedetti, “Nonlinear dimensionality reduction in molecular simulation: The diffusion map approach,” *Chem. Phys. Lett.* **509**(1–3), 1–11 (2011).
- ³P. Das, M. Moll, H. Stamati, L. E. Kavrakli, and C. Clementi, “Low-dimensional, free-energy landscapes of protein-folding reactions by nonlinear dimensionality reduction,” *Proc. Natl. Acad. Sci. U. S. A.* **103**(26), 9885–9890 (2006).
- ⁴R. R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner, and S. W. Zucker, “Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps,” *Proc. Natl. Acad. Sci. U. S. A.* **102**(21), 7426–7431 (2005).
- ⁵A. L. Ferguson, A. Z. Panagiotopoulos, P. G. Debenedetti, and I. G. Kevrekidis, “Systematic determination of order parameters for chain dynamics using diffusion maps,” *Proc. Natl. Acad. Sci. U. S. A.* **107**(31), 13597–13602 (2010).
- ⁶M. A. Rohrdanz, W. Zheng, and C. Clementi, “Discovering mountain passes via torchlight: Methods for the definition of reaction coordinates and pathways in complex macromolecular reactions,” *Annu. Rev. Phys. Chem.* **64**, 295–316 (2013).
- ⁷X. Cao and P. Tian, “Dividing and conquering’ and ‘caching’ in molecular modeling,” *Int. J. Mol. Sci.* **22**(9), 5053 (2021).
- ⁸M. D. Ward, M. I. Zimmerman, A. Meller, M. Chung, S. J. Swamidass, and G. R. Bowman, “Deep learning the structural determinants of protein biochemical properties by comparing structural ensembles with DiffNets,” *Nat. Commun.* **12**(1), 3023 (2021).
- ⁹E. O. Salawu, “DESP: Deep enhanced sampling of proteins’ conformation spaces using AI-inspired biasing forces,” *Front. Mol. Biosci.* **8**, 587151 (2021).
- ¹⁰A. W. Senior, R. Evans, J. Jumper, J. Kirkpatrick, L. Sifre, T. Green, C. Qin, A. Židek, A. W. R. Nelson, A. Bridgland, H. Penedones, S. Petersen, K. Simonyan, S. Crossan, P. Kohli, D. T. Jones, D. Silver, K. Kavukcuoglu, and D. Hassabis, “Improved protein structure prediction using potentials from deep learning,” *Nature* **577**(7792), 706–710 (2020).
- ¹¹P. G. Bolhuis, “Two-state protein folding kinetics through all-atom molecular dynamics based sampling,” *Front. Biosci.* **14**, 2801–2828 (2009).
- ¹²B. Hashemian, D. Millán, and M. Arroyo, “Modeling and enhanced sampling of molecular systems with smooth and nonlinear data-driven collective variables,” *J. Chem. Phys.* **139**, 214101 (21) (2013).
- ¹³S. Martin, A. Thompson, E. A. Coutsiias, and J. P. Watson, “Topology of cyclo-octane energy landscape,” *J. Chem. Phys.* **132**(23), 234115 (2010).
- ¹⁴W. Chen, H. Sidky, and A. L. Ferguson, “Nonlinear discovery of slow molecular modes using state-free reversible VAMPnets,” *J. Chem. Phys.* **150**, 214114 (21) (2019).
- ¹⁵W. M. Brown, S. Martin, S. N. Pollock, E. A. Coutsiias, and J. P. Watson, “Algorithmic dimensionality reduction for molecular structure analysis,” *J. Chem. Phys.* **129**, 064118 (6) (2008).
- ¹⁶W. Chen and A. L. Ferguson, “Molecular enhanced sampling with autoencoders: On-the-fly collective variable discovery and accelerated free energy landscape exploration,” *J. Comput. Chem.* **39**(25), 2079–2102 (2018).
- ¹⁷W. Chen, A. R. Tan, and A. L. Ferguson, “Collective variable discovery and enhanced sampling using autoencoders: Innovations in network architecture and error function design,” *J. Chem. Phys.* **149**, 072312 (7) (2018).
- ¹⁸S. Salvador and P. Chan, in *Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI* (IEEE, Piscataway, NJ, 2004), pp. 576–584.
- ¹⁹A. Glielmo, I. Macocco, D. Doimo, M. Carli, C. Zeni, R. Wild, M. d’Errico, A. Rodriguez, and A. Laio, “DADapy: Distance-based analysis of data-manifolds in Python,” *Patterns* **3**(10), 100589 (2022).

- ²⁰M. J. Kirby and R. Miranda, "Circular nodes in neural networks," *Neural Comput.* **8**, 390–402 (1996).
- ²¹M. Scholz and R. Vigário, in *The 10th Euroean Symposium on Artificial Neural Networks (ESANN)* (i6doc, Bruges, 2002), pp. 439–444.
- ²²F. Manuchehrfar, H. Li, W. Tian, A. Ma, and J. Liang, "Exact topology of the dynamic probability surface of an activated process by persistent homology," *J. Phys. Chem. B* **125**(18), 4667–4680 (2021).
- ²³A. Glielmo, C. Zeni, B. Cheng, G. Csányi, and A. Laio, "Ranking the information content of distance measures," *PNAS Nexus* **1**, pgac039 (2) (2022).
- ²⁴J. Mirth, Y. Zhai, J. Bush, E. G. Alvarado, H. Jordan, M. Heim, B. Krishnamoorthy, M. Pflaum, A. Clark, Y. Z, and H. Adams, "Representations of energy landscapes by sublevelset persistent homology: An example with *n*-alkanes," *J. Chem. Phys.* **154**, 114114 (11) (2021).
- ²⁵G. H. Weber, P. T. Bremer, and V. Pascucci, "Topological landscapes: A terrain metaphor for scientific data," *IEEE Trans. Visualization Comput. Graphics* **13**(6), 1416–1423 (2007).
- ²⁶Y. Hiraoka, T. Nakamura, A. Hirata, E. G. Escobar, K. Matsue, and Y. Nishiura, "Hierarchical structures of amorphous solids characterized by persistent homology," *Proc. Natl. Acad. Sci. U. S. A.* **113**(26), 7035–7040 (2016).
- ²⁷S. Gerber and K. Potter, "Data analysis with the Morse–Smale complex: The msc Package for R," *J. Stat. Software* **50**(2), 1–22 (2012).
- ²⁸F. Cazals, F. Chazal, and T. Lewiner, "Molecular shape analysis based upon the Morse–Smale complex and the Connolly function," in *Proceedings of the Annual Symposium on Computational Geometry* (ACM, 2003), pp. 351–360.
- ²⁹D. Laney, P. T. Bremer, A. Mascarenhas, P. Miller, and V. Pascucci, "Understanding the structure of the turbulent mixing layer in hydrodynamic instabilities," *IEEE Trans. Visualization Comput. Graphics* **12**(5), 1053–1060 (2006).
- ³⁰S. Plimpton, "Fast parallel algorithms for short-range molecular dynamics," *J. Comput. Phys.* **117**(1), 1–19 (1995).
- ³¹A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimeshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, *Advances in Neural Information Processing Systems* (Curran Associates, Inc., 2019), Vol. 32, pp. 8024–8035.
- ³²J. Tierny, G. Favelier, J. A. Levine, C. Gueunet, and M. Michaux, "The topology ToolKit," *IEEE Trans. Visualization Comput. Graphics* **24**(1), 832–842 (2018).
- ³³T. Masood, J. Budin, M. Falk, G. Favelier, C. Garth, C. Gueunet, P. Guillou, L. Hofmann, P. Hristov, A. Kamakshidasan, C. P. Kappe, P. Klacansky, P. Laurin, J. A. Levine, J. Lukasczyk, D. Sakurai, M. Soler, P. Steneteg, J. Tierny, W. Usher, J. Vidal, and M. Wozniak, *An Overview of the Topology ToolKit* (Springer, 2019), Vol. 1–15.
- ³⁴W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, "Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids," *J. Am. Chem. Soc.* **118**(45), 11225–11236 (1996).
- ³⁵S. Nosé, "A unified formulation of the constant temperature molecular dynamics methods," *J. Chem. Phys.* **81**(1), 511–519 (1984).
- ³⁶NVIDIA, P. Vingelmann, and F. H. P. Fitzek, *CUDA, release: 10.2.89*, NVIDIA, 2020.
- ³⁷D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings* (ICLR, 2015), pp. 1–15.