JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Analysis of the Geometrical Evolution in On-the-Fly Surface-Hopping Nonadiabatic Dynamics with Machine Learning Dimensionality Reduction Approaches: Classical Multidimensional Scaling and Isometric Feature Mapping

Xusong Li,[†,§,⊥,‡] Yu Xie,[†] Deping Hu,[†,⊥] and Zhenggang Lan*[,†,§,⊥]

[†]CAS Key Laboratory of Biobased Materials, Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao 266101, China

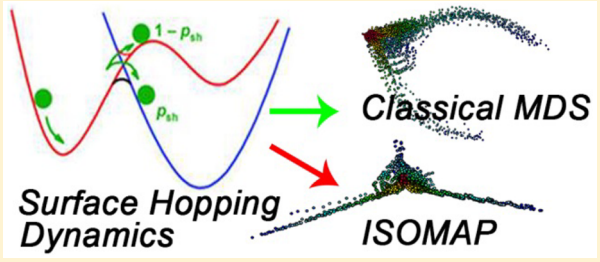
[§]Sino-Danish Center for Education and Research/Sino-Danish College, University of Chinese Academy of Sciences, Beijing 100049, China

[⊥]University of Chinese Academy of Sciences, Beijing 100049, China

[‡]Department of Chemistry, Technical University of Denmark, Kemitorvet 207, 2800, Kgs. Lyngby, Denmark

Supporting Information

ABSTRACT: On-the-fly trajectory-based nonadiabatic dynamics simulation has become an important approach to study ultrafast photochemical and photophysical processes in recent years. Because a large number of trajectories are generated from the dynamics simulation of polyatomic molecular systems with many degrees of freedom, the analysis of simulation results often suffers from the large amount of high-dimensional data. It is very challenging but meaningful to find dominating active coordinates from very complicated molecular motions. Dimensionality reduction techniques provide ideal tools to realize this purpose. We apply two dimensionality reduction approaches (classical multidimensional scaling and isometric feature mapping) to analyze the results of the on-the-fly surface-hopping nonadiabatic dynamics simulation. Two representative model systems, $CH_2NH_2^+$ and the phytochromobilin chromophore model, are chosen to examine the performance of these dimensionality reduction approaches. The results show that these approaches are very promising, because they can extract the major molecular motion from complicated time-dependent molecular evolution without preknown knowledge.



1. INTRODUCTION

Nonadiabatic dynamics plays an essential role in photophysics and photochemistry.[1−3] The simulation of nonadiabatic dynamics is not a trivial task because of the breakdown of the Born−Oppenheimer approximation.[4] Within the framework of quantum dynamics, several theoretical methods were proposed to simulate nonadiabatic dynamics, including standard quantum dynamics,[4] the multiconfiguration time-dependent Hartree method (MCTDH),[5] multilayer multiconfiguration time-dependent hartree method (ML-MCTDH),[6] and quantum dissipative dynamics[4,7] based on density operator. However, it is only possible to use these methods to treat the nonadiabatic dynamics of small molecular systems, model systems with reduced dimension, or the simplified Hamiltonian with specific forms, such as spin-boson or linear vibronic coupling models. To solve the nonadiabatic dynamics of complex systems, great efforts were also made to develop trajectory-based and Gaussian-wavepacket-based dynamical methods within the frameworks of mixed quantum-classical and semiclassical approaches, such as Ehrenfest dynamics and its variation,[8−13] ab initio multiconfigurational Ehrenfest (AI-MCE),[14] trajectory surface hopping (TSH),[15−29] quantum classical Liouville equation,[30−35] quasiclassical or semiclassical dynamics with mapping model,[36−41] ab initio multiple spawning (AIMS),[42] variational multiconfiguration Gaussian (VMCG),[43] and other approaches.[44−47] The direct combination of trajectory-based methods and electronic-structure calculations allows us to propagate trajectory in the manner of on-the-fly.[16,17,19−22,42,48−53] This opens a possible way to simulate the full-dimensional nonadiabatic dynamics of complex realistic systems at the atomic level.

Normally, the trajectory-based methods require the propagation of a large number of trajectories and the average of these trajectories gives the molecular evolution in the nonadiabatic dynamics. Beside the time-dependent electronic population, the characterization of nonadiabatic dynamics also requires the analysis of other simulation results, for instance the geometrical evolution with respect to time, geometrical features associated with nonadiabatic transitions near potential energy crossings and

Received: April 13, 2017
Published: September 1, 2017

DOI: 10.1021/acs.jctc.7b00394
J. Chem. Theory Comput. 2017, 13, 4611−4623

the evolution of electronic characters. Such analysis provides the direct understanding of active reaction coordinates responsible for nonadiabatic decay. In the analysis of on-the-fly TSH, a very simply way is to pick up a few "typical" trajectories representing the characteristic features in dynamics evolution. This widely used protocol provides a direct view of the time-dependent geometrical evolution by eye view. By comparing the initial geometries and hopping geometries, we easily identify reaction coordinates. It is also possible to plot several internal coordinates vs time and to find important ones. However, this approach faces two challenging issues. (1) How to select the "representative" trajectory and how to characterize the overall trajectory distribution in the high-dimensional coordinate space? (2) How to find a suitable way to characterize reactive coordinates if many internal molecular coordinates are involved in nonadiabatic dynamics? The answer to the first question seems not difficult, because we may simply plot the time-dependent distribution of involved internal coordinates based on many trajectories instead of only a few. However, the answer to the second problem is not straightforward, because many nuclear degrees of freedom may strongly mix in the nonadiabatic decay. Overall, as pointed out recently by Tully, the analysis of trajectory-based nonadiabatic molecular dynamics represents a great challenge.[54]

In the mathematical view, a single geometry snapshot is represented by a point in a high-dimensional coordinate space, which is characterized by a high-dimensional vector. The overall geometry evolution in the propagation of many trajectories corresponds to the motion of a swarm of points (or a density distribution) in this high-dimensional space. Thus, the identification of reactive coordinates responsible for the nonadiabatic dynamics is equivalent to the construction of the effective moving pathways in the high-dimensional space. Possibly, these pathways may lie within a low-dimensional reduced subspace embedded in the original high-dimensional full space. This view allows us to consider the identification of active coordinates in nonadiabatic dynamics as a so-called "dimensionality reduction" task.[55] In the last decades, many linear and nonlinear dimensionality reduction methods were proposed in the research field of machine learning (ML), which includes principal component analysis (PCA),[56−58] multidimensional scaling (MDS),[59−61] isometric feature mapping (ISOMAP),[62,63] diffusion map,[64,65] and so on. These methods show great impacts to several research fields, such as pattern recognitions, language, and figure recognitions and artificial intelligence. These dimensionality reduction methods and their modified versions were used to analyze the results of the ground-state molecular dynamics.[55,66−70] An overview of all technical details of these methods and their application in ground-state molecular dynamics can be found in a recent review paper.[55] However, much less work was done in the employment of these technologies to analyze the excited-state nonadiabatic dynamics. Only recently, some initial efforts were devoted in this field. For example, the diffusion map and PCA were used in the data analysis of nonadiabatic dynamics.[71−74] Thus, more studies should be performed to employ various dimensionality reduction approaches, in order to extract the effective reaction coordinate from nonadiabatic dynamics simulation. Although the employment of these dimensionality reduction tricks in principle does not require the preknowledge of the reaction mechanism, great caution should be paid in implementation and analysis. The brute force employment of these pure mathematical methods may result in artifacts in data analysis, because no physical insight is involved. Thus, it is also necessary to carefully check the performance of these ML-based dimensionality reduction methods as tools to analyze the nonadiabatic dynamics.

In this contribution, we try to use dimensionality reduction approaches in the investigation of the geometrical evolution in nonadiabatic dynamics. Two widely used dimensionality reduction technologies, classical MDS and ISOMAP, are employed to identify the geometrical evolution features from the TSH calculations, and to build the active reaction coordinates responsible for nonadiabatic dynamics. After the detailed discussion on implementation details, we use classical MDS and ISOMAP to examine the dynamical evolution of two typical model systems, $CH_2NH_2^+$ and phytochromobilin chromophore (PΦB) model. Although the current work is based on the TSH calculations, it is also possible to use these analysis protocols in other trajectory-based nonadiabatic dynamics methods. This treatment may significantly reduce the difficulties in the analysis of the geometrical evolution in the nonadiabatic dynamics of a polyatomic system, particular in the case that many degrees of freedom are involved.

This work is organized as follows. Section 2 focuses on the theoretical and computational details. After the basic theories of classical MDS and ISOMAP were introduced, various implementation details, from data prescreening to effective-coordinate construction, are discussed. Section 3 discusses the performance of classical MDS and ISOMAP in the analysis of nonadiabatic dynamics of two model systems, $CH_2NH_2^+$ and PΦB model. Section 4 gives the summary of the current work.

2. THEORETICAL METHODS AND COMPUTATIONAL DETAILS

2.1. Theoretical Methods. *2.1.1. Classical Multidimensional Scaling.* If many data points exit in a high dimensional space, it is not possible to examine their distribution pattern by eye view directly. A possible treatment on this type of problems is to create a group of points with the same number in a low-dimensional space, which preserves the distribution pattern of original data set. This idea of dimensionality reduction provides us a straightforward way to investigate the distribution of data set under study, because it is much easier to "see" these points in the low-dimensional space.

The classical MDS is one of widely used dimensionality reduction methods,[56,60] which tries to construct the low-dimensional embedding pattern that preserves all pairwise dissimilarities between data points under study. The basic principle of classical MDS is outlined here and all theoretical details can be founded in previous works.[60]

We define the pairwise dissimilarity matrix $\mathbf{D}$, in which each element d_{ij} represents the so-called "distance" between two data points. Later on, we also call $\mathbf{D}$ as the "distance matrix" in this work for illustration.

The scalar product matrix $\mathbf{B}$ is obtained using the below equation

$$\mathbf{B} = -\frac{1}{2}\mathbf{J}\mathbf{D}^2\mathbf{J} \tag{1}$$

where $\mathbf{D}$ is a squared proximity matrix as

$$\mathbf{D} = [d_{ij}^{\ 2}] \tag{2}$$

and $\mathbf{J}$ is a center matrix as

$$\mathbf{J} = \mathbf{I} - n^{-1}\mathbf{1}\mathbf{1}^{\mathrm{T}} \tag{3}$$

DOI: 10.1021/acs.jctc.7b00394
J. Chem. Theory Comput. 2017, 13, 4611−4623

where **I** is unit matrix and n is the number of objects.

The diagonalization of the **B** matrix gives a set of eigenvalues. After sorting the eigenvalues from largest to smallest, the important dimension is identified. The larger eigenvalue corresponds to more important dimension. For instance, if a reduced space with m-dimension is considered, we keep the m largest positive eigenvalues λ_1 ...λ_m and the corresponding m eigenvectors e_1 ... e_m.

Embedding coordinates of all points in the low-dimensional space are obtained by

$$\mathbf{L} = (e_1 \cdots e_m)\begin{pmatrix} \sqrt{\lambda_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sqrt{\lambda_m} \end{pmatrix} \tag{4}$$

The relative embedding error in the dimensionality reduction is computed according to the stress function introduced in the MDS textbook.[60]

2.1.2. Isometric Feature Mapping. When a data set is distributed over a manifold, the distribution pattern of these data points is not well described by the pairwise Euclidean distance matrix. In this case, dimensionality reduction approaches within the framework of manifold learning may become useful. One of such approaches is the ISOMAP method,[62,63,66] which performs the classical MDS analysis based on the distance matrix represented by the so-called "geodesic distance" instead of the "Euclidean distance".

The ISOMAP idea is easily understood. In principle the pairwise geodesic distance matrix should provide a better representation of the distribution pattern of a data set over a manifold. The "geodesic distance" between two points can be approximated by the possible shortest hopping pathway through all interconnected neighbor points. After the construction of the pairwise geodesic distance matrix, the next step goes back to the classical MDS analysis.

2.1.3. Trajectory Surface Hopping Dynamics. In this work, we mainly discuss the dimensionality reduction analysis of the geometry evolution in on-the-fly Tully's TSH simulation, although it is trivial to use the similar analysis protocol in other trajectory-based methods. We outline the concept of Tully's TSH approaches, because all details can be found in many previous works.[15−17,21,22,27] In the Tully's TSH framework, the nuclear motion on the single potential energy surface is treated by numerical propagation of the classical trajectory. The electronic wave function was treated by quantum evolution. Nonadiabatic transitions between different electronic states were considered via Tully's fewest switches approach.[15] After the combination of the on-the-fly dynamics, the TSH approach becomes extremely powerful in the treatment of relatively large and realistic molecular systems at atomic level with full degree of freedom. The decoherence correction (with the parameter α = 0.1 hartree) proposed by Granucci et al. was employed.[75] The initial conditions (such as geometries and velocities) in the dynamic simulations were generated from the Wigner distribution of normal modes of the ground state. All dynamics calculations and data analyses were done within the developing version of the JADE package,[21,22] which includes a module to interface with several electronic-structure methods in different standard quantum chemistry packages.

2.2. Implementation Details. To employ the classical MDS and ISOMAP in the analysis of the nonadiabatic dynamics trajectories, we need to collect many geometry snapshots during the trajectory propagation. A pairwise distance matrix is built to describe the "dissimilarity" between two geometries. Following the idea of previous works,[71] we use the root-mean-square deviation (RMSD) of two geometries to represent their dissimilarity. The resulted distance matrix is directly used in the further classical MDS analysis. To perform the ISOMAP analysis, we need to build the pairwise geodesic distance matrix and then perform the classical MDS analysis. After the classical MDS or ISOMAP analysis, all molecular geometries in nonadiabatic dynamics evolution are represented by a group of data points in a low-dimensional space. The "distance" between two geometries is represented by the Euclidean distance between two corresponding points in this low-dimensional space. Overall, the distribution of data points in the low-dimensional space should provide a direct view of the trajectory propagation in the original high-dimensional coordinate space. Next we discussed implementation details.

2.2.1. Geometry Collection. Let us assume that many trajectories are generated from the TSH calculations. Several geometries are selected in each trajectory, which include the starting geometry at time zero, many snapshots in the trajectory propagation with certain time intervals and geometries at hops. We collected these selected geometries over all trajectories to define a preliminary geometry database. We also added the ground-state minimum geometry (or several minimum-energy geometries) into this database, because it naturally defines a starting reference in trajectory analysis.

After a trajectory hops back to the ground states, the excessive kinetic energy may simply result in a highly distorted molecular geometry or even cause the cleavage of the system in the lack of dissipation. To distinguish the nonadiabatic decay dynamics and the further "hot" molecular motion on the ground state, we may stop the geometries snapshot to avoid the system dissociation. Certainly, sometimes the system does not break after jumping back to the ground state and the photoproducts are formed. In this case, we do not need to stop the trajectory counting and the geometry evolution may give a full dynamical picture of photoinduced reactions, namely from reactants to products.

2.2.2. Calculation of the Pairwise Dissimilarity/Distance Matrix. The dissimilarity or distance between any two geometries is computed by measuring their RMSD. The calculation of the RMSD follows previous works,[76,77] which removes the contribution of translational and rotational motions. If necessary, symmetric operations (such as mirror reflection and permutation of identical atoms) should also be taken into account in the RMSD calculations. The pairwise distance matrix defines a fully connected network or a graph, in which each node corresponds to a selected geometry and the internode connection represents their dissimilarity.

A few of highly distorted geometries may appear, which are far from the starting points or ground-state minimum. The existence of these "strange" geometries may give strange numerical results, although they are very rare. We should carefully remove them and check the analysis result. In practices, we define the ground-state minimum as the reference point, because all trajectories start from some initial geometries near it. The "distances" between all geometries and ground-state minimum were computed. When the distance is larger than a cutoff value, we do not take the corresponding geometry into account. The setup of reasonable cutoff values was possible by carefully checking the final results. In this way, we sometimes disregard < 5% points in the final analysis.

DOI: 10.1021/acs.jctc.7b00394
J. Chem. Theory Comput. 2017, 13, 4611−4623

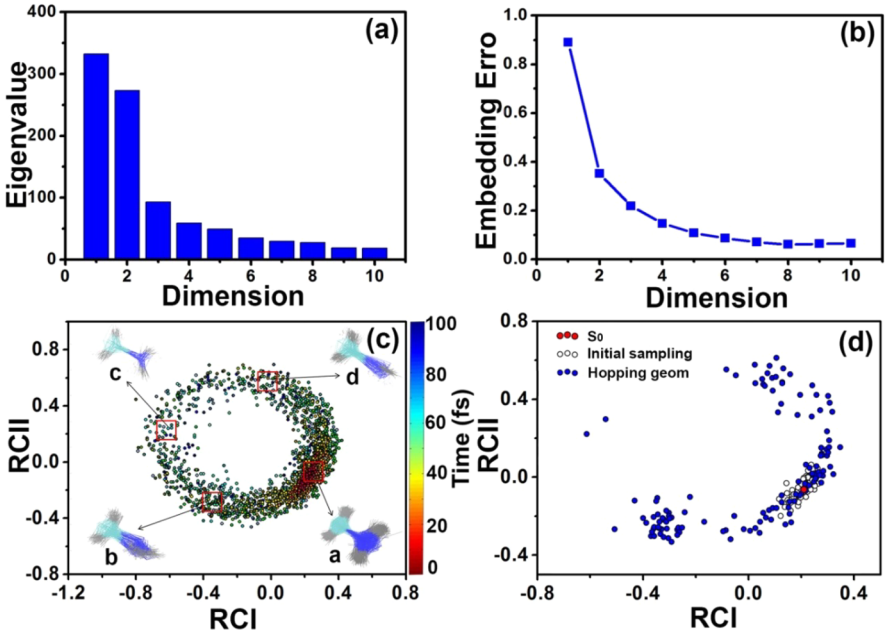


Figure 1. (a) Eigenvalue spectrum for the classical MDS analysis of the nonadiabatic dynamics of $CH_2NH_2^+$ without symmetry consideration. (b) The embedding error as the function of dimension. (c) Locations of sampled geometries in the low-dimensional space spanned by two leading reduced coordinates RCI and RCII. Color codes indicate the time dependent dynamics process. Geometrical aggregations in four representative locations are shown. (d) Locations of the initial geometries and the hopping geometries in the two-dimensional reduced space.

2.2.3. Classical Multidimensional Scaling (MDS). Starting from the distance matrix, we performed the dimensionality reduction by using classical MDS. The leading eigenvalues are given and the embedding coordinate of each point in the low-dimensional space is computed. In the current analysis, the total number of geometries is not extremely huge. Thus, we do not use additional tricks, such as the selection of landmark points[61,66] and clustering technologies,[66] to reduce the number of geometries. These tricks should in principle become very helpful if an extremely huge amount of geometries are involved.

2.2.4. Isometric Feature Mapping (ISOMAP). In ISOMAP, the classical MDS analysis should be performed based on the pairwise geodesic distance matrix. As discussed previously, the pairwise distance (or RMSD) matrix corresponds to a fully connected graph, in which each node represents a selected geometry and the associating connecting distance reflects the distance between two nodes. We cut the connection between two nodes when their distance is larger than a threshold. Two ways were normally used to set up this cutoff threshold, either by giving the largest cutoff distance (ε-ball approach) or by setting the maximum number of connecting nodes for each node (k-point approach).[62,63] This step gives us a partial interconnected graph.

Next, the pairwise geodesic distances for all nodes were constructed by using Dijkstra's algorithm[78,79] or the Floyd–Warshall algorithm.[79−81] In principle we wish to get a new fully connected graph, in which all nodes are connected and the distance between two nodes reflect their geodesic distance. In practices, only the k-point approach gives us a fully connected graph, while the ε-ball approach may not. When ε is too small, the resulted graph may not be fully connected. For example, the whole graph may be divided into a few subgraphs that are not connected to each other, or a few of the nodes may be isolated. This results in the numerical problem in the further analysis because the "residual" infinity distances become dominant in the dimensionality reduction. In this case, we removed all points that are not connected to the point representing ground-state minimum and then defined a smaller fully connected graph. Certainly, we need to avoid that too many points are removed. In both ε-ball and k-point approaches, the dependences of embedding errors on dimension, cutoff ε or k values, are carefully checked.

2.2.5. Further Analysis of Geometry Evolution. We plotted the embedding coordinates as scattering points in the low-dimensional space spanned by reduced coordinates. Each point represents a geometry snapshot. After grouping geometries with similar reduced coordinates, we obtain a clear view on geometrical feature. The correlation between reduced coordinates and internal coordinates is constructed by averaging all geometries within a grid. Then the leading reaction coordinates are easily identified by following the pathway from starting points to hopping points, as well the time evolution of all points.

2.2.6. Implementation Issues. The RMSD between two geometries is computed with a simple homemade FORTRAN code and the algorithm was taken from previous work.[76,77] Most analysis codes were written with Python language. A few of the Python routines, including data prescreening and data analysis, were developed based on the Scikit-learn toolkit.[82,83] For example, the pairwise geodesic distance graph was constructed by calling the inner module of Scikit-learn Python toolkit. Because the MDS module in Scikit-learn Python toolkit is not implemented in the way of standard classical MDS, we developed the python routine for this purpose.

3. RESULTS

3.1. $CH_2NH_2^+$ Results. The nonadiabatic dynamics of $CH_2NH_2^+$ was investigated by the TSH method at the SA3-CASSCF(6,4)/6-31G* level (three states in the state average calculation with six electrons in four orbitals: two σ, one π, and one π^*), the JADE code[21,22] on the basis of the MOLPRO program[84] was used to simulate the nonadiabatic dynamics. The initial state is S_2 and we totally consider 100 trajectories. For each

DOI: 10.1021/acs.jctc.7b00394
J. Chem. Theory Comput. 2017, 13, 4611−4623

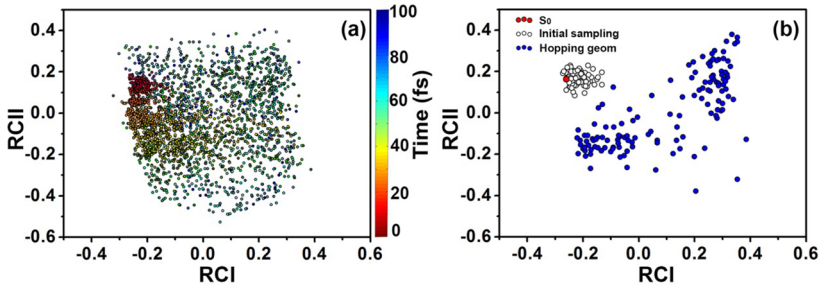


Figure 2. Classical MDS analysis of the nonadiabatic dynamics of $CH_2NH_2^+$ with symmetry consideration. (a) Locations of the sampled geometries as a function of RCI and RCII. Colors indicate the evolution over time. (b) Locations of the initial geometries and the hopping geometries in the map.

trajectory, the snapshots are taken at every 2 fs and the maximum time to stop the geometry record is 100 fs. When a trajectory stays on the ground state for more than 20 fs after the internal conversion, we stop count geometries. The ground-state minimum-energy geometry and the first $S_1 \rightarrow S_0$ hopping geometries were also included. Totally 4458 geometries were collected to form a database. Then the pairwise dissimilarity matrix was computed by measuring the RMSD between all geometry pairs. Next we removed the geometries far from the ground-state minimum; ∼5% geometries were discarded.

3.1.1. Classical MDS Analysis without Symmetry Consideration. To get a preliminary understanding of the nonadiabatic dynamics of $CH_2NH_2^+$, we first consider the classical MDS analysis based on the RMSD matrix without symmetry consideration. Both eigenvalues and embedding errors decrease with the dimension increasing, see Figure 1a,b. Because the first and second dimensions are dominant as shown in Figure 1a, we plot the representing points in the two-dimensional space, spanned by two leading coordinates, RCI and RCII in Figure 1c. They can be treated as the major reactive coordinates. The geometrical motion in the nonadiabatic dynamics vs time is reflected by the evolution of the scattering points in Figure 1c. It is clear that the data points representing the trajectory propagation clearly form a circle, more precisely an ellipse. The major axis of this ellipse is basically parallel with RCI, which is slightly longer than the minor axis parallel with RCII. Because we did not take symmetry into account, this ellipse in fact directly reflects the internal rotation of the CN bond in the nonadiabatic dynamics of $CH_2NH_2^+$. The starting geometrical aggregation (labeled as **a**) is not far from the crossing point between the major axis and the ellipse itself. In the geometrical aggregates **b** and **d**, two moieties at the C and N sides are nearly perpendicular to each other. We have two such geometrical aggregates with a similar number of structures because the internal rotation of the central CN bond may take place along the clockwise or anticlockwise pathway. The structure aggregate **d** collects all structures after the 180° rotation of the C−N bond. The current ellipse perfectly explains the internal rotation of the C−N bond, and the hop may take place at different torsional angles. However, the simple two reduced coordinates cannot provide more insight on other important molecular motion; for example, we easily see that some hopping geometries overlap with the starting geometries as shown in the Figure 1d. In one word, when no symmetry is considered, the limited information in the two-dimensional space may not provide detailed understanding on the geometry evolution.

3.1.2. Classical MDS Analysis with Symmetry Consideration. Next the symmetry properties (permutation of identical atoms and mirror reflection) were taken into account in the construction of RMSD distance matrix, and then the classical MDS was performed. Two largest eigenvalues are dominant and the embedding errors rapidly decrease with the dimension increasing, as shown in Figure S1 of Supporting Information. The geometry evolution in the two-dimensional space is given in Figure 2a. Under symmetry considerations, the clockwise and anticlockwise internal rotations along the CN bond are equivalent, and thus the distribution of data points does not give a circle. The starting geometries are located near RCI ≈ −0.2 and RCII ≈ 0.15. In the early stage of dynamics, these points mainly move more along the RCII direction. Afterward, they move along the RCI direction, and a large area of two-dimensional space is covered. Most importantly, initial geometries and hopping geometries are clearly separated as shown in Figure 2b. The inclusion of symmetry consideration displays more details of molecular evolution.

By checking geometry aggregates at different grid regions in Figure 3, we know that RCI mainly refers to the central CN

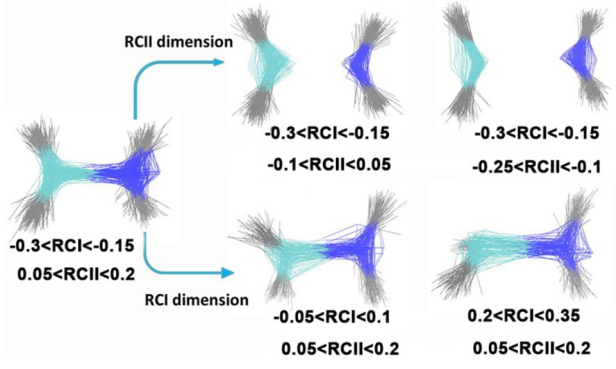


Figure 3. Geometry aggregates at a few important grid regions with different values of RCI and RCII.

torsional motion and RCII is governed by the CN stretching motion. This relationship between the reduced coordinates and internal coordinates (average over geometries within a small grid) is further confirmed by Figure 4. It is clear that the RCI is relevant to the torsional motion along the CN bond because one of HCNH dihedral angle changes from ∼10° (RCI ≈ −0.25) to ∼70° (RCI ≈ 0.25). However, the RCI almost does not depend on the CN stretching motion. As the contrast, the RCII is not relevant to the torsional motion but highly correlated to the CN stretching motion. Within the two-dimensional space, the geometrical evolution in the nonadiabatic dynamics of $CH_2NH_2^+$ is easily analyzed. In the early stage of dynamics, the system mainly experiences the CN stretching motion on the excited states. Later on the torsional motion starts. The hops

DOI: 10.1021/acs.jctc.7b00394
J. Chem. Theory Comput. 2017, 13, 4611−4623

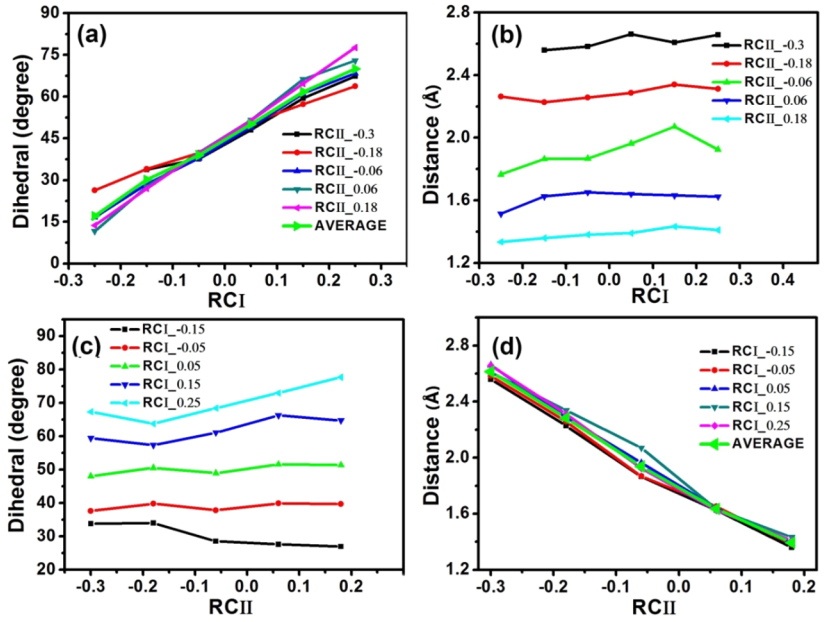


Figure 4. Averaged values internal coordinates vs reduced coordinates. (a) CN bond torsion angle vs RCI. (b) CN bond length vs RCI. (c) CN bond torsion angle vs RCII. (d) CN bond length vs RCII.

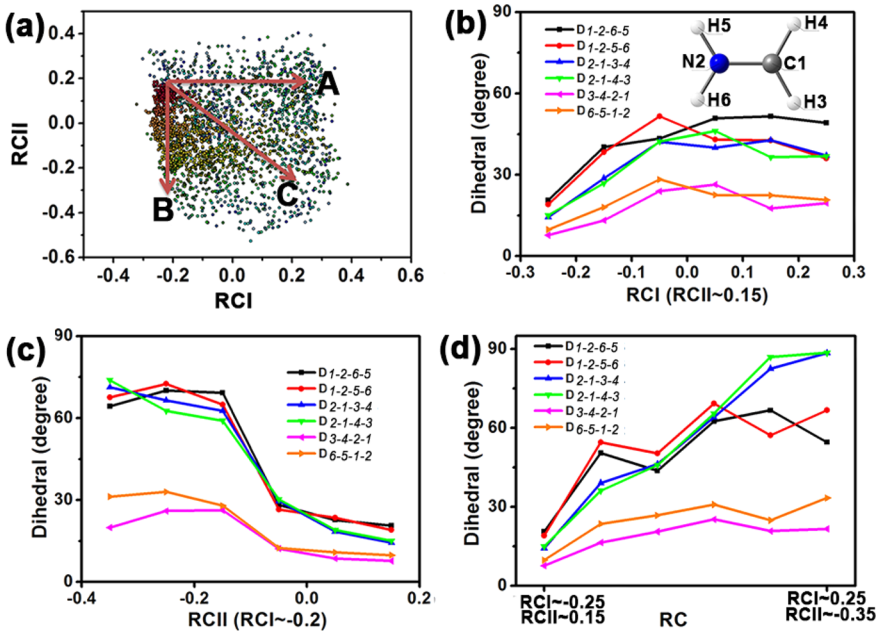


Figure 5. Six dihedral angles (relevant to the pyramidalization motion) vs reduced coordinates in the Classical MDS results: (a) three selected pathways (**A**, **B**, and **C**); (b) along the **A** pathway; the inserting panel gives the $CH_2NH_2^+$ geometry and atomic labels; (c) along the **B** pathway; (d) along the **C** pathway.

back to the ground state may take place when the CN stretching motion is dominant or torsional motion is dominant. Some hops may also happen with the mixture of two internal motions. This picture is consistent with the previous studies results on the dynamics of $CH_2NH_2^+$.[16,17,21,26,85] More interestingly, the previous study figured out that the S_1/S_0 conical intersection seam is relevant to the torsional coordinate along the CN bond and the CN stretching coordinate. With more torsional angle, less stretching motion is required to access the conical intersection seam. This feature is exactly consistent with the hopping distribution in the reduced space (see Figure 2b).

It is well-known that the pyrimidalization plays an essential role in the nonadiabatic dynamics of $CH_2NH_2^+$,[16,17,21,26,85] that is, the minimum protonated-Schiff-base (PSB) model. In fact, such a type of hydrogen-out-of-plane (HOOP) motion is also very important in the nonadiabatic dynamics of other PSB systems.[86−88] Thus, it is very interesting to examine whether the current analysis protocol can identify such a critical feature.

Starting from Franck−Condon region (RCI ≈ −0.2 and RCII ≈ 0.15), we considered three different paths, **A**, **B**, and **C** as shown in Figure 5a. Six representative dihedral angles ($D_{1-2-6-5}$, $D_{1-2-5-6}$, $D_{2-1-3-4}$, $D_{2-1-4-3}$, $D_{3-4-2-1}$ and $D_{6-5-1-2}$) were chosen to characterize the pyramidalization, and the atomic

DOI: 10.1021/acs.jctc.7b00394
J. Chem. Theory Comput. 2017, 13, 4611−4623

labels were shown in Figure 5 (b). The **A** pathway describes the pure twisting motion in which the RCI changes and RCII remains unchanged. We clearly see that the pyramidalization angles increase along the **A** pathway (Figure 5b). Along the **B** pathway, the RCII changes, while RCI remains unchanged. This pathway describes the pure CN stretching motion and the pyramidalization is also relevant (Figure 5c). The **C** pathway describes that both CN twisting and stretching motions are involved (Figure 5d), which starts from the FC region to the ending point near RCI ≈ 0.25 and RCII ≈ −0.35, Along it, we also see the contribution of pyramidalization. This shows that pyramidalization contributes to two dominant reduced coordinates (RCI and RCII). Also because of this reason, it is not suitable to check the dependence of pyramidalization vs only one reduced coordinate by averaging of relevant dihedral angles over other reduced coordinates. Overall, the identification of pyramidalization as the reactive coordinate is possible within a two-dimensional reduced space spanned by RCI and RCII.

As a short summary, RCI is mainly correlated to the torsional motion along the CN bond while RCII is relevant to the stretching motion along the CN bond. The pyramidalization motions, on the other hand, contribute to both RCI and RCII.

3.1.3. ISOMAP Analysis with Symmetry Consideration. Next, we try to perform the ISOMAP analysis, in which both ε-ball and k-point cutoff approaches were examined. As shown in Figure 6, when the cutoff threshold is large than 0.4 in the ε-ball approach, the distribution of data points in the low dimensional space is very similar to that obtained by classical MDS. When this value becomes smaller, the different distribution patterns appear and the embedding error rises (Figure S2). Only when the cutoff threshold is less than 0.15 does the embedding error start to decrease again. Although different distribution patterns appear with small ε, too many data (>50% in the ε < 0.15 case) were disregarded due to their nonconnectivity. Thus, we do not discuss the results from the ε-ball ISOMAP analysis.

When the k-point cutoff approach is employed, no data point is disregarded. As shown in Figure S3, the embedding error is large with k = 2, while similar distribution patterns are observed when k = 4−8. Thus, we discuss the results with the k = 8 case, in which the RCI coordinate is relevant to both of CN bond stretching and twisting motion (Figure 7). Starting from the FC region with RCI ≈ 0, the RCI is dominated by the CN stretching motion for RCI > 0, while it is highly relevant to the CN twisting motion for RCI < 0. It is also clear that the hops take place either in the region with either large CN stretching or strong torsional motion.

We also examine the correlations between pyramidalization motions and reduced coordinates. Similar to the MDS analysis, two pathways (**A** and **B** in the Figure 8a) are observed, whose directions are determined by the connection pathways from the Franck−Condon region to two different hopping areas. As shown in Figure 8b,c, it is clear that the pyrimidalization motions take place from starting points to the hopping points. Although such pyrimidalization motions appear in both **A** and **B** pathways, they are more significant in the later one involving the CN twisting motion. We also try to check the correlation between these dihedral angels and two reduced coordinates directly (**C** and **D** in the Figure a). As shown in Figure 8d,e, the pyramidalization motions seem highly correlated to the second reduced coordinate RCII, although it is weakly relevant to RCI. Thus, the inclusion of the second dimensionality (or the second reduced coordinate RCII) should give more detailed insight of molecular motion, although the first leading reduced coordinate RCI has already capture major geometry evolution.

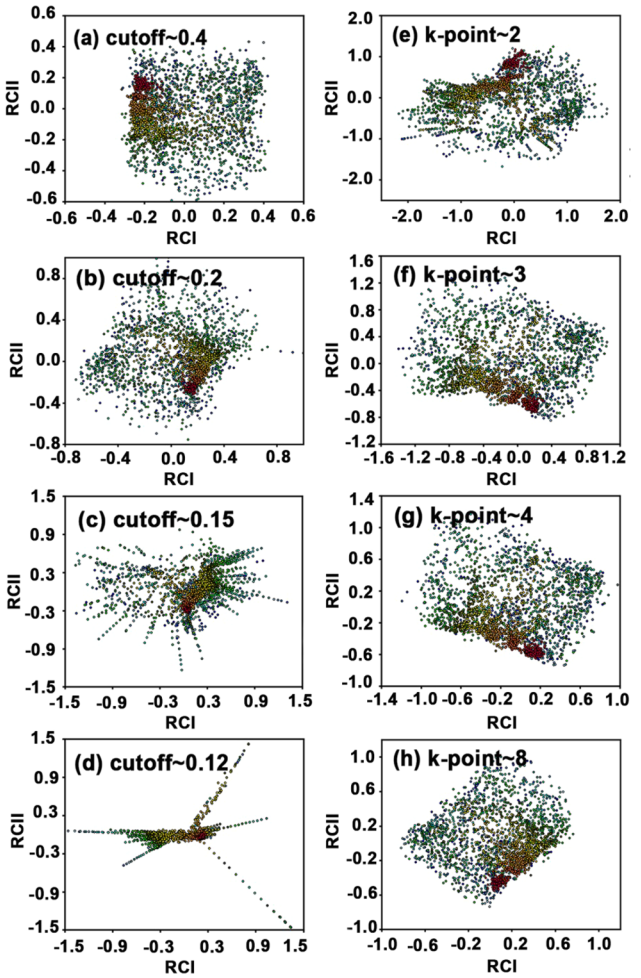


Figure 6. ISOMAP analysis of the nonadiabatic dynamics of $CH_2NH_2^+$ with symmetry consideration. (a−d) ε-ball ISOMAP results with different cutoff thresholds and (e−h) k-point ISOMAP results with different k values.

The results above implied that the ISOMAP analysis with suitable setup may provide more information on important molecular motion. The nonlinear feature of ISOMAP may become extremely attractive when the data points are distributed in a manifold with high dimension.

3.2. The ZaZsZa Isomer of PΦB Model. The photoinduced dynamics of phytochromes have received considerable attention,[89−91] because of its relevance to the bioclock regulation of the plants and other autotrophs, such as seed germination, growth, phototaxis, pigmentation, and so on. Our previous works[92] on the phytochromobilin chromphore (PΦB) model gave physical insight to its nonadiabatic dynamics. Here we try to reinvestigate the nonadiabatic dynamics of the PΦB model compound, particular the ZaZsZa isomer that was identified as the main isomer of phytochromes by Raman experiments.[93,94] Because the current system is not very small, it is interesting to examine its photoinduced nonadiabatic dynamics in the basis of the classical MDS and ISOMAP analyses. The nonadiabatic dynamics of the PΦB model was investigated by TSH method at the semiempirical OM2/MRCI level[17,95−97] including 16 electrons in 12 orbitals: six π, two n, and four π^*. To unify the format of output data for classical MDS and ISOMAP analyses,

DOI: 10.1021/acs.jctc.7b00394
J. Chem. Theory Comput. 2017, 13, 4611−4623

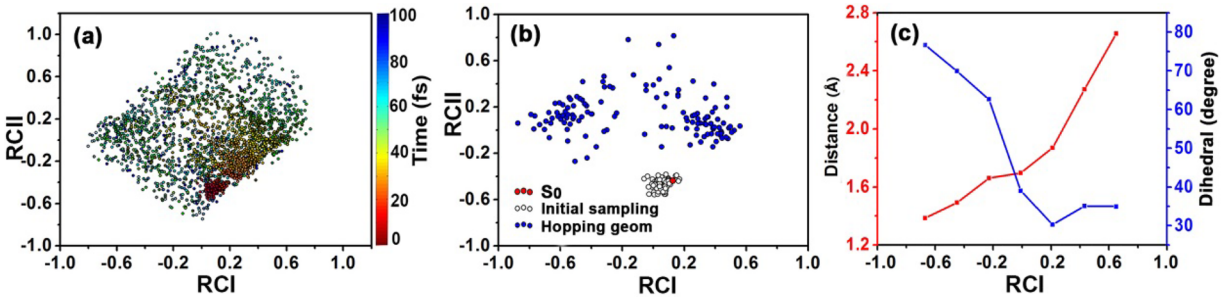


Figure 7. ISOMAP analysis of the nonadiabatic dynamics of $CH_2NH_2^+$ with symmetry consideration. The *k*-point approach was taken with $k = 8$: (a) locations of the sampled geometries as a function of RCI and RCII; colors indicate the time following dynamics process; (b) locations of the initial geometries and the hopping geometries in the map; (c) two key coordinates vs RCI.

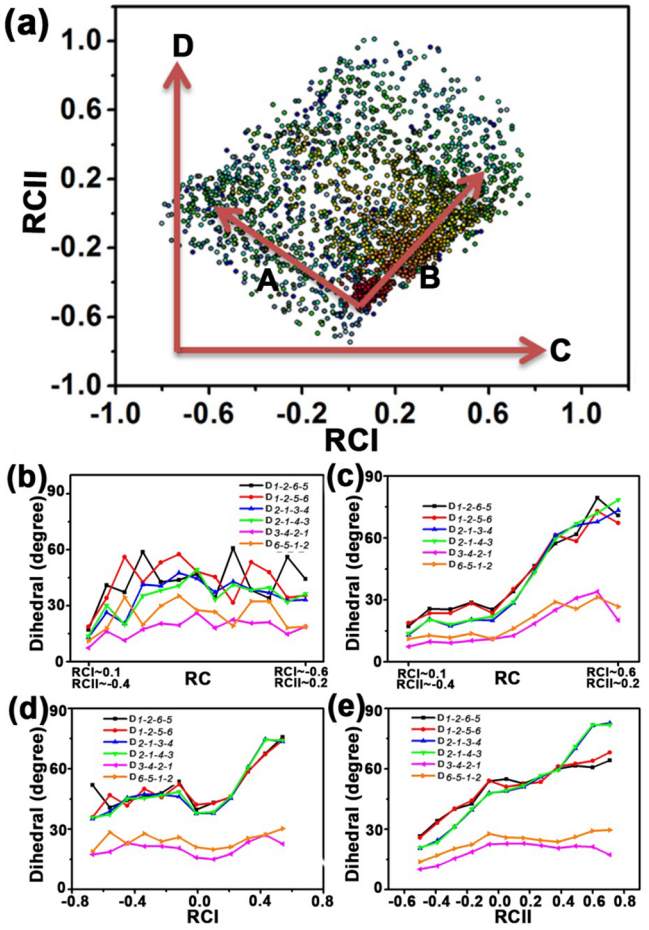


Figure 8. Six dihedral angles (relevant to the pyramidalization motion) vs reduced coordinates in the ISOMAP results; (a) four selected pathways; (b) along with the **A** pathway; (c) along with the **B** pathway; (d) averaged values of these dihedral angles vs RCI; (e) averaged values of these dihedral angles vs RCII.

the JADE code[21,22] with the interface to the MNDO program[98] was used to simulate the nonadiabatic dynamics. The current calculations employed the same setup as reported in our previous work.[92] For each trajectory, the snapshots were taken at every 20 fs and the maximum time to record geometry is 600 fs. Each successful trajectory gives 30 geometries. A few trajectories do not go to the end and in these cases less snapshots are taken. After including the first $S_1 \rightarrow S_0$ hopping geometries and the ground-state minimum, 3867 geometries were collected to form a data set. The RMSD calculations do not include all hydrogen atoms.

The classical MDS analysis of the snapshots in trajectory propagation gives a very interesting distribution pattern in the low-dimensional space spanned by two reduced coordinates. Starting from the Franck−Condon region labeled by the "**A**" area in Figure 9 (RCI ≈ −0.2 and RCII ≈ 0.1), two different propagation pathways are clearly identified. One branch follows the **A**−**B**−**C** pathway and the second branch is distributed along the **A**−**D**−**E** pathway.

Along the **A**−**B**−**C** pathway, the dihedral angle of the C_9-C_{10} bond changes from ∼0° at the **A** area, ∼ 90° at the **B** area to ∼180° at the **C** area. We also clearly see that many hops in this channel take place near the B area. Thus, it is very clear that the torsional motion of the C_9-C_{10} bond is responsible for the **A**−**B**−**C** pathway. As a contrast, it is very clear that the torsional motion of the $C_{10}-C_{11}$ bond is responsible for the **A**−**D**−**E** pathway. Most hops in this pathway take place near the **D** area with the dihedral angle of the $C_{10}-C_{11}$ bond close to 90°. The ratio between the numbers of hops in two channels is around 3:1.

Overall, the two-dimensional data map clearly explains the nonadiabatic dynamics. The major reaction channel in the nonadiabatic decay follows the C_9-C_{10} torsion, while the secondary channel follows the $C_{10}-C_{11}$ torsional motion. All main features are very consistent with our previous results based on the analysis of the time evolution of a large number of internal coordinates over many trajectories.[92]

Although the classical MDS analysis provides a clear insight on molecular motion in the nonadiabatic dynamical evolution, at least two reduced coordinates are needed for a good representation. This feature is also reflected by the fact that the second eigenvalue give an important contribution (see Figure S4 in SI)

Next we performed the ISOMAP analysis with both ε-ball and *k*-point approaches. When the very small cutoff threshold $\varepsilon = 0.2$ was used, the embedding error becomes significant (Figure S5). When such value falls into the range of 0.3−0.6, the embedding errors become stable. Similar features are observed in the *k*-point cases. The very small *k* value ($k = 2$) gives very large embedding error and such error becomes stable with $k > 5$ (Figure S6). When the suitable cutoff values were employed in the *k*-point case ($k > 5$) and the ε-ball case ($\varepsilon = 0.4-0.6$), our analyses found that important features are rather similar in data distribution. Thus, we only discuss the ε-ball case with $\varepsilon = 0.4$.

When the suitable cutoff threshold, such as 0.4, was employed in the ε-ball approach, the ISOMAP analysis compresses the data distribution into the tribranched shape in the two-dimensional space (Figure 10). The second eigenvalue gives much less

DOI: 10.1021/acs.jctc.7b00394
J. Chem. Theory Comput. 2017, 13, 4611−4623

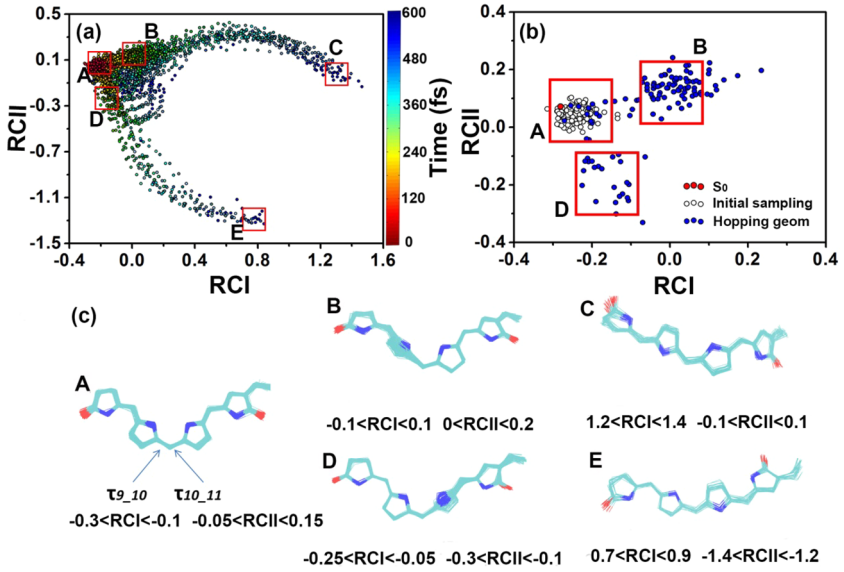


Figure 9. Classical MDS analysis of the nonadiabatic dynamics of PΦB model with symmetry consideration. (a) Locations of sampled geometries in the low-dimensional space spanned by two leading reduced coordinates RCI and RCII. Color codes indicate the time following dynamics process. (d) Locations of the initial geometries and the hopping geometries in the two-dimensional reduced space. (c) Geometrical aggregations in five representative locations are shown.

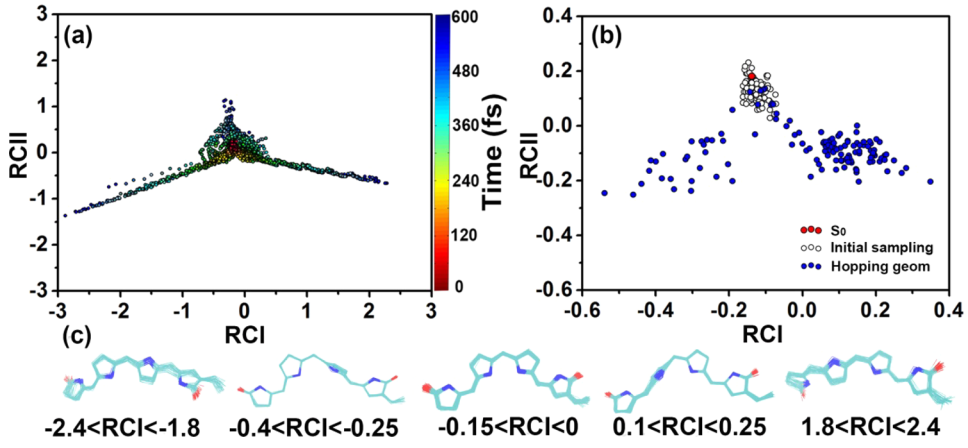


Figure 10. ISOMAP analysis of the nonadiabatic dynamics of the PΦB model with symmetry consideration. The ε-ball approach is used with cutoff = 0.4. (a) Locations of sampled geometries in the low-dimensional space spanned by two leading reduced coordinates RCI and RCII. Color codes indicate the time following the dynamics process. (d) Locations of the initial geometries and the hopping geometries in the two-dimensional reduced space. (c) Geometrical aggregations in five representative locations are shown.

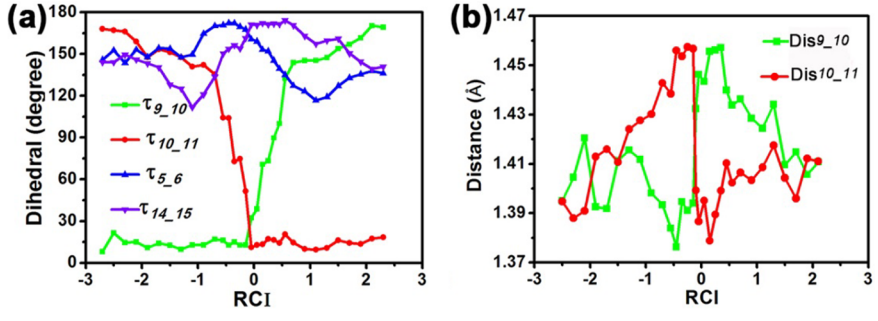


Figure 11. (a) The evolution of the four key coordinates vs RCI. (b) The evolution of the two bond distances vs RCI.

contribution (shown in Figure S4) and thus the RCII contribution is minor. It may be enough to use the most leading reduced coordinate RCI to describe major dynamics behavior, and thus we can assign it as the main reaction coordinate.

The Franck−Condon region is located near RCI ∼ −0.1. When the RCI becomes larger (RCI > 0), Figure 11a shows that the dihedral angle of the C_9−C_{10} bond increases dramatically from 0° (RCI ≈ 0) to 90° (RCI ≈ 0.2) and even to 150° (RCI ≈

DOI: 10.1021/acs.jctc.7b00394
J. Chem. Theory Comput. 2017, 13, 4611−4623

1). The hops with RCI > 0 take place near the region with dihedral angle of the C_9-C_{10} bond near 90° (RCI ≈ 0.2). After RCI > 1.5, the torsion becomes less important. On the other hand, the significant $C_{10}-C_{11}$ torsion is responsible for the region with −1 < RCI < 0 and hops following this channel take place in the region with the dihedral angle of the $C_{10}-C_{11}$ bond near 90° (RCI ≈ −0.3).

The C_9-C_{10} twisting motion brings the system from the Franck−Condon region to the hopping points, and Figure 11b shows that the elongation of the C_9-C_{10} bond and the shortening of $C_{10}-C_{11}$ bond also take place. After hops, the system moves to photoproducts and this process is characterized by the C_9-C_{10} twisting angle accessing >150°. During the photoproduct formation on the ground state, the shortening of C_9-C_{10} and the elongation of the $C_{10}-C_{11}$ are observed. A similar bond alternation tendency is observed in the RCI < 0.1 case of the $C_{10}-C_{11}$ bond twisting channel.

Besides the main active coordinates, C_9-C_{10} or $C_{10}-C_{11}$, it is also possible to know whether other coordinates are involved in RCI. In Figure 11a, the C_9-C_{10} torsion is always accompanied by the small C_5-C_6 torsion. Concerning with the hopping distribution, it is very clear that the CI seam in this channel is relevant to these two coordinates. At the same time, the $C_{10}-C_{11}$ and $C_{14}-C_{15}$ torsional motions are also correlated at the corresponding CI seam. These features are consistent with our previous studies.[92]

As a short summary, the ISOMAP is a very powerful approach to perform the dimensionality reduction analysis of the nonadiabatic dynamics of the PΦB model. Although the classical MDS method is good enough to represent the dynamical feature, at least a two-dimensional space is required. The ISOMAP with the suitable setups on the other hand, compress all necessary information into a one-dimensional space. In this case all dynamical features can be well described by the reaction along a single reduced coordinate RCI, and thus we can safely assign it as the reaction coordinate. This provides a direct way to examine the essential molecular motion responsible for nonadiabatic decay.

All dimensionality reduction approaches try to extract the "main" or "leading" pattern based on the statistical analysis of a large number of data. Although it shows great promise in capturing the leading features, some minor details are missing. At the same time, the results by these approaches are only meaningful when a large number of data are available. Particularly, if we wish to identify a pattern, the amount of data to capture this pattern should not only be large enough but also have a sufficient proportion in all used data points. Thus, we expect that the current dimensionality reduction approaches cannot capture the very minor reaction channels, because the number of trajectories passing them is very low.

4. CONCLUSION

In this work, we employ two dimensionality reduction analysis tools, classical MDS and ISOMAP, to examine the molecular motion in nonadiabatic dynamics. Particularly, our purpose is to construct the reactive coordinates responsible for nonadiabatic decay by these dimensionality reduction algorithms. Two representative systems were employed to check the performance of these dimensionality reduction algorithms.

The classical MDS and ISOMAP analyses follow the below protocol. The dissimilarity/"distance" between any two arbitrary geometries obtained in the trajectory propagation is measured by their RMSD. After the RMSD distance matrix is set up based on many snapshots over all trajectories, we can perform the classical MDS and ISOMAP analyses. Then the geometrical data distribution pattern is constructed in the low-dimensional space spanned by reduced coordinates. By checking the geometry aggregates at different areas in the low-dimensional space, we can directly "see" the essential molecular motion. This provides a straightforward way to build the major reaction coordinates responsible for the nonadiabatic dynamics.

Although the classical MDS is a rather simple dimensional reduction approach, the employment of it already provides many essential evolution features of nonadiabatic dynamics, for both two-model systems under study. The symmetry properties in the dynamical evolution can be well displayed in the classical MDS analysis. The consideration of the symmetry (permutation, reflection) normally enhances the contribution of the leading dimension. Although the classical MDS is very powerful, we also realize that maybe several reduced coordinates, not only one, are equally important to illustrate the nonadiabatic dynamics. For example, at least two dimensions are required for the correct analyses of two representative examples $CH_2NH_2^+$ and PΦB in this work.

As a powerful nonlinear manifold dimensionality reduction method, ISOMAP may largely compress the information pattern of molecular motion in the nonadiabatic dynamics evolution. In some cases, even a single reduced coordinate RCI may provide many essential features of nonadiabatic dynamics. However, as pointed out by a previous work,[55,62] the setup of the correct cutoff parameter in the ISOMAP is not a trivial task.

Although these dimensionality reduction methods have been used to analyze the conformation evolution of the ground-state molecular dynamics, only recently have some works started to use these ideas to examine nonadiabatic dynamics.[71−74] The current work represents our initial efforts to use linear and nonlinear dimensionality reduction approaches to analyze the geometry evolution of nonadiabatic dynamics. Our current results display that it is possible to extract the major molecular motion responsible for the nonadiabatic dynamics by using a few leading reduced coordinates. These leading coordinates are a mixture of different key molecular motions. For example, the dimensionality reduction analysis of the nonadiabatic dynamics of $CH_2NH_2^+$ shows that several molecular motions (stretching, twisting, and pyramidalization) contribute to the leading reduced coordinates. However, the employment of these analysis tools is not trivial, which sometimes requires the careful setup of correct parameters and performing the suitable data prescreening. In more real situations, geometry evolution may become much more complicated in nonadiabatic dynamics. For example, several nuclear degrees of freedom may be highly mixed, very complicated molecular motions (such as ring opening and ring puckering) may be involved in the nonadiabatic decay, trajectory evolution may follow many distinguished reaction pathways, and the system size may become extremely huge. In these cases, the identification of the active coordinates by dimensionality reduction is more challenging. This should be a very interesting topic in future study.

Here we wish to give some comments on a few important issues that may be useful to address these challenges. The current dimensionality reduction approach is based on the statistical analysis of a large amount of data depicting the geometry evolution. Only when these data can capture the major geometry evolution, may we extract a few leading coordinates properly. Thus, the selection of the suitable molecular descriptors[99−106] (for the proper representation of many input geometries) in

DOI: 10.1021/acs.jctc.7b00394
J. Chem. Theory Comput. 2017, 13, 4611−4623

dimensionality reduction analyses should become very critical in the description of the complicated molecular motion in the nonadiabatic dynamics of more realistic situations. In this sense, other molecular descriptors,[99−106] such as Bag of Bonds,[104] Coulomb matrix,[99,105] etc., may become useful in the analysis of more complicated nonadiabatic dynamics. To deal with very complicated molecular motions, we may need some additional tricks (such as landmark points[61]) in the MDS and ISOMAP analysis, or even employ other dimensionality reduction approaches.[55,64,65,67,69,70] We also need to point out that sometimes it may not easy to only use one or two reduced coordinates to represent the highly mixed active motion of the molecular systems possible due to the strange topologic structure of data distribution. This problem is "detected" from the large embedding errors. In this case, more reduced coordinates should be used to describe the molecular motion until the embedding errors get small enough. In addition, it is very challenging to include the evolution of electronic degrees in the dimensionality reduction analysis of nonadiabatic dynamics, because it may shed light on the interesting correlation between electronic and nuclear degrees of freedom. The proper way to include the electronic motion in the molecular descriptor is also an interesting topic in the future. At the same time, these dimensionality reduction approaches can also be used in other trajectory-based or Gaussian-wavepacket-based methods. All problems mentioned above open many research opportunities in the future.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.7b00394.

Some useful illustrative figures of eigenvalues, embedding errors, and the data distributions in classical MDS and ISOMAP (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*Fax: +86-532-80662778. Tel.: +86-532-80662630. E-mail: lanzg@qibebt.ac.cn, zhenggang.lan@gmail.com.

ORCID

Xusong Li: 0000-0001-6586-7746

Yu Xie: 0000-0001-8925-6958

Funding

This work is supported by NSFC Project (Nos. 21673266 and 21503248) and the Natural Science Foundation of Shandong Province for Distinguished Young Scholars (JQ201504).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank Supercomputing Centre, Computer Network Information Center, CAS, National Supercomputing Center in Shenzhen, National Supercomputing Center in Guangzhou, and the Super Computational Centre of CAS-QIBEBT for providing computational resources.

■ REFERENCES

(1) Domcke, W.; Yarkony, D. R. Role of Conical Intersections in Molecular Spectroscopy and Photoinduced Chemical Dynamics. *Annu. Rev. Phys. Chem.* **2012**, *63*, 325−352.

(2) Matsika, S.; Krause, P. Nonadiabatic Events and Conical Intersections. *Annu. Rev. Phys. Chem.* **2011**, *62*, 621−643.

(3) González, L.; Escudero, D.; Serrano-Andrés, L. Progress and Challenges in the Calculation of Electronic Excited States. *ChemPhysChem* **2012**, *13*, 28−51.

(4) Domcke, W.; Yarkony, D. R.; Köppel, H. *Conical Intersections: Electronic Structure, Dynamics & Spectroscopy*; World Scientific, 2004.

(5) Meyer, H. D.; Manthe, U.; Cederbaum, L. S. The Multi-Configurational Time-Dependent Hartree Approach. *Chem. Phys. Lett.* **1990**, *165*, 73−78.

(6) Wang, H. B.; Thoss, M. Multilayer Formulation of the Multiconfiguration Time-Dependent Hartree Theory. *J. Chem. Phys.* **2003**, *119*, 1289−1299.

(7) Schröter, M.; Ivanov, S. D.; Schulze, J.; Polyutov, S. P.; Yan, Y.; Pullerits, T.; Kühn, O. Exciton-Vibrational Coupling in the Dynamics and Spectroscopy of Frenkel Excitons in Molecular Aggregates. *Phys. Rep.* **2015**, *567*, 1−78.

(8) Li, X. S.; Tully, J. C.; Schlegel, H. B.; Frisch, M. J. Ab Initio Ehrenfest Dynamics. *J. Chem. Phys.* **2005**, *123*, 084106.

(9) Isborn, C. M.; Li, X. S.; Tully, J. C. Time-Dependent Density Functional Theory Ehrenfest Dynamics: Collisions between Atomic Oxygen and Graphite Clusters. *J. Chem. Phys.* **2007**, *126*, 134307.

(10) Cheng, S. C.; Zhu, C. Y.; Liang, K. K.; Lin, S. H.; Truhlar, D. G. Algorithmic Decoherence Time for Decay-of-Mixing Non-Born-Oppenheimer Dynamics. *J. Chem. Phys.* **2008**, *129*, 024112.

(11) Meng, S.; Kaxiras, E. Real-Time, Local Basis-Set Implementation of Time-Dependent Density Functional Theory for Excited State Dynamics Simulations. *J. Chem. Phys.* **2008**, *129*, 054110.

(12) Bedard-Hearn, M. J.; Larsen, R. E.; Schwartz, B. J. Mean-field Dynamics with Stochastic Decoherence (MF-SD): A New Algorithm for Nonadiabatic Mixed Quantum/Classical Molecular-Dynamics Simulations with Nuclear-Induced Decoherence. *J. Chem. Phys.* **2005**, *123*, 234106.

(13) Zhu, C. Y.; Jasper, A. W.; Truhlar, D. G. Non-Born-Oppenheimer Liouville-Von Neumann Dynamics. Evolution of a Subsystem Controlled by Linear and Population-Driven Decay of Mixing with Decoherent and Coherent Switching. *J. Chem. Theory Comput.* **2005**, *1*, 527−540.

(14) Saita, K.; Shalashilin, D. V. On-the-Fly Ab Initio Molecular Dynamics with Multiconfigurational Ehrenfest Method. *J. Chem. Phys.* **2012**, *137*, 22A506.

(15) Tully, J. C. Molecular-Dynamics with Electronic-Transitions. *J. Chem. Phys.* **1990**, *93*, 1061−1071.

(16) Barbatti, M.; Granucci, G.; Persico, M.; Ruckenbauer, M.; Vazdar, M.; Eckert-Maksić, M.; Lischka, H. The on-the-Fly Surface-Hopping Program System NEWTON-X: Application to Ab Initio Simulation of the Nonadiabatic Photodynamics of Benchmark Systems. *J. Photochem. Photobiol., A* **2007**, *190*, 228−240.

(17) Fabiano, E.; Keal, T. W.; Thiel, W. Implementation of Surface Hopping Molecular Dynamics Using Semiempirical Methods. *Chem. Phys.* **2008**, *349*, 334−347.

(18) Webster, F.; Wang, E. T.; Rossky, P. J.; Friesner, R. A. Stationary-Phase Surface Hopping for Nonadiabatic Dynamics: Two-State Systems. *J. Chem. Phys.* **1994**, *100*, 4835−4847.

(19) Richter, M.; Marquetand, P.; González-Vázquez, J.; Sola, I.; González, L. SHARC: Ab Initio Molecular Dynamics with Surface Hopping in the Adiabatic Representation Including Arbitrary Couplings. *J. Chem. Theory Comput.* **2011**, *7*, 1253−1258.

(20) Doltsinis, N. L.; Marx, D. Nonadiabatic Car-Parrinello Molecular Dynamics. *Phys. Rev. Lett.* **2002**, *88*, 166402.

(21) Du, L. K.; Lan, Z. G. An on-the-Fly Surface-Hopping Program JADE for Nonadiabatic Molecular Dynamics of Polyatomic Systems: Implementation and Applications. *J. Chem. Theory Comput.* **2015**, *11*, 1360−1374.

(22) Du, L. K.; Lan, Z. G. Correction to an on-the-Fly Surface-Hopping Program JADE for Nonadiabatic Molecular Dynamics of Polyatomic Systems: Implementation and Applications. *J. Chem. Theory Comput.* **2015**, *11*, 4522−4523.

DOI: 10.1021/acs.jctc.7b00394
J. Chem. Theory Comput. 2017, 13, 4611−4623

(23) Zhu, C. Y.; Nobusada, K.; Nakamura, H. New Implementation of the Trajectory Surface Hopping Method with Use of the Zhu-Nakamura Theory. *J. Chem. Phys.* **2001**, *115*, 3031−3044.
(24) Belyaev, A. K.; Lasser, C.; Trigila, G. Landau-Zener Type Surface Hopping Algorithms. *J. Chem. Phys.* **2014**, *140*, 224108.
(25) Shenvi, N.; Subotnik, J. E.; Yang, W. T. Simultaneous-Trajectory Surface Hopping: A Parameter-Free Algorithm for Implementing Decoherence in Nonadiabatic Dynamics. *J. Chem. Phys.* **2011**, *134*, 144102.
(26) Tapavicza, E.; Tavernelli, I.; Rothlisberger, U. Trajectory Surface Hopping within Linear Response Time-Dependent Density-Functional Theory. *Phys. Rev. Lett.* **2007**, *98*, 023001.
(27) Werner, U.; Mitrić, R.; Suzuki, T.; Bonačić-Koutecký, V. Nonadiabatic Dynamics Within the Time Dependent Density Functional Theory: Ultrafast Photodynamics in Pyrazine. *Chem. Phys.* **2008**, *349*, 319−324.
(28) Fang, J. Y.; Hammes-Schiffer, S. Improvement of the Internal Consistency in Trajectory Surface Hopping. *J. Phys. Chem. A* **1999**, *103*, 9399−9407.
(29) Akimov, A. V.; Neukirch, A. J.; Prezhdo, O. V. Theoretical Insights into Photoinduced Charge Transfer and Catalysis at Oxide Interfaces. *Chem. Rev.* **2013**, *113*, 4496−4565.
(30) Horenko, I.; Salzmann, C.; Schmidt, B.; Schütte, C. Quantum-Classical Liouville Approach to Molecular Dynamics: Surface Hopping Gaussian Phase-Space Packets. *J. Chem. Phys.* **2002**, *117*, 11075−11088.
(31) Shi, Q.; Geva, E. A Derivation of the Mixed Quantum-Classical Liouville Equation from the Influence Functional Formalism. *J. Chem. Phys.* **2004**, *121*, 3393−3404.
(32) Martens, C. C.; Fang, J. Y. Semiclassical-Limit Molecular Dynamics on Multiple Electronic Surfaces. *J. Chem. Phys.* **1997**, *106*, 4918−4930.
(33) Ando, K.; Santer, M. Mixed Quantum-Classical Liouville Molecular Dynamics without Momentum Jump. *J. Chem. Phys.* **2003**, *118*, 10399−10406.
(34) Ryabinkin, I. G.; Hsieh, C. Y.; Kapral, R.; Izmaylov, A. F. Analysis of Geometric Phase Effects in the Quantum-Classical Liouville Formalism. *J. Chem. Phys.* **2014**, *140*, 084104.
(35) Kapral, R.; Ciccotti, G. Mixed Quantum-Classical Dynamics. *J. Chem. Phys.* **1999**, *110*, 8919−8929.
(36) Liu, J. A Unified Theoretical Framework for Mapping Models for the Multi-State Hamiltonian. *J. Chem. Phys.* **2016**, *145*, 204105.
(37) Tao, G. H. Coherence-Controlled Nonadiabatic Dynamics via State-Space Decomposition: A Consistent Way to Incorporate Ehrenfest and Born-Oppenheimer-Like Treatments of Nuclear Motion. *J. Phys. Chem. Lett.* **2016**, *7*, 4335−4339.
(38) Meyer, H. D.; Miller, W. H. Classical Analog for Electronic Degrees of Freedom in Non-Adiabatic Collision Processes. *J. Chem. Phys.* **1979**, *70*, 3214−3223.
(39) Stock, G.; Thoss, M. Semiclassical Description of Nonadiabatic Quantum Dynamics. *Phys. Rev. Lett.* **1997**, *78*, 578−581.
(40) Cotton, S. J.; Miller, W. H. Symmetrical Windowing for Quantum States in Quasi-Classical Trajectory Simulations: Application to Electronically Non-Adiabatic Processes. *J. Phys. Chem. A* **2013**, *117*, 7190−7194.
(41) Sun, X.; Wang, H. B.; Miller, W. H. Semiclassical Theory of Electronically Nonadiabatic Dynamics: Results of a Linearized Approximation to the Initial Value Representation. *J. Chem. Phys.* **1998**, *109*, 7064−7074.
(42) Ben-Nun, M.; Martínez, T. J. Ab Initio Quantum Molecular Dynamics. *Adv. Chem. Phys.* **2002**, *121*, 439−512.
(43) Richings, G. W.; Worth, G. A. A Practical Diabatisation Scheme for Use with the Direct-Dynamics Variational Multi-Configuration Gaussian Method. *J. Phys. Chem. A* **2015**, *119*, 12457−12470.
(44) Yonehara, T.; Hanasaki, K.; Takatsuka, K. Fundamental Approaches to Nonadiabaticity: Toward a Chemical Theory beyond the Born-Oppenheimer Paradigm. *Chem. Rev.* **2012**, *112*, 499−542.
(45) Lu, J. F.; Zhou, Z. N. Improved Sampling and Validation of Frozen Gaussian Approximation with Surface Hopping Algorithm for Nonadiabatic Dynamics. *J. Chem. Phys.* **2016**, *145*, 124109.
(46) Heaps, C. W.; Mazziotti, D. A. Accurate Non-Adiabatic Quantum Dynamics from Pseudospectral Sampling of Time-Dependent Gaussian Basis Sets. *J. Chem. Phys.* **2016**, *145*, 064101.
(47) Herman, M. F. Toward an Accurate and Efficient Semiclassical Surface Hopping Procedure for Nonadiabatic Problems. *J. Phys. Chem. A* **2005**, *109*, 9196−9205.
(48) Curchod, B. F. E.; Tavernelli, I.; Rothlisberger, U. Trajectory-Based Solution of the Nonadiabatic Quantum Dynamics Equations: An on-the-Fly Approach for Molecular Dynamics Simulations. *Phys. Chem. Chem. Phys.* **2011**, *13*, 3231−3236.
(49) Langer, H.; Doltsinis, N. L. Nonradiative Decay of Photoexcited Methylated Guanine. *Phys. Chem. Chem. Phys.* **2004**, *6*, 2742−2748.
(50) Craig, C. F.; Duncan, W. R.; Prezhdo, O. V. Trajectory Surface Hopping in the Time-Dependent Kohn-Sham Approach for Electron-Nuclear Dynamics. *Phys. Rev. Lett.* **2005**, *95*, 163001.
(51) Mitrić, R.; Werner, U.; Wohlgemuth, M.; Seifert, G.; Bonačić-Koutecký, V. Nonadiabatic Dynamics within Time-Dependent Density Functional Tight Binding Method. *J. Phys. Chem. A* **2009**, *113*, 12700−12705.
(52) Klein, S.; Bearpark, M. J.; Smith, B. R.; Robb, M. A.; Olivucci, M.; Bernardi, F. Mixed State 'on the Fly' Non-adiabatic Dynamics: The Role of the Conical Intersection Topology. *Chem. Phys. Lett.* **1998**, *292*, 259−266.
(53) Yu, L.; Xu, C.; Lei, Y. B.; Zhu, C. Y.; Wen, Z. Y. Trajectory-Based Nonadiabatic Molecular Dynamics without Calculating Nonadiabatic Coupling in the Avoided Crossing Case: Trans < -> Cis Photoisomerization in Azobenzene. *Phys. Chem. Chem. Phys.* **2014**, *16*, 25883−25895.
(54) Tully, J. C. Perspective: Nonadiabatic Dynamics Theory. *J. Chem. Phys.* **2012**, *137*, 22A301.
(55) Rohrdanz, M. A.; Zheng, W. W.; Clementi, C. Discovering Mountain Passes via Torchlight: Methods for the Definition of Reaction Coordinates and Pathways in Complex Macromolecular Reactions. *Annu. Rev. Phys. Chem.* **2013**, *64*, 295−316.
(56) Bishop, C. M. *Pattern Recognition and Machine Learning*; Springer Science+Business Media: New York, 2006.
(57) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. Essential Dynamics of Proteins. *Proteins: Struct., Funct., Genet.* **1993**, *17*, 412−425.
(58) Jolliffe, I. *Principal Component Analysis*; Wiley Online Library, 2002.
(59) Hardle, W.; Simar, L. *Applied Multivariate Statistical Analysis*; Springer: Berlin, 2007.
(60) Borg, I.; Groenen, P. J. F. *Modern Multidimensional Scaling: Theory and Applications*; Springer Science & Business Media: USA, 2005.
(61) De Silva, V.; Tenenbaum, J. B. *Sparse Multidimensional Scaling Using Landmark Points*; Technical report: Stanford University, 2004.
(62) Balasubramanian, M.; Schwartz, E. L. The Isomap Algorithm and Topological Stability. *Science* **2002**, *295*, 295.
(63) Tenenbaum, J. B.; de Silva, V.; Langford, J. C. A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science* **2000**, *290*, 2319−2323.
(64) Coifman, R. R.; Lafon, S.; Lee, A. B.; Maggioni, M.; Nadler, B.; Warner, F.; Zucker, S. W. Geometric Diffusions as a Tool for Harmonic Analysis and Structure Definition of Data: Diffusion Maps. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 7426−7431.
(65) Coifman, R. R.; Lafon, S. Diffusion Maps. *Appl. Comput. Harmon.* **2006**, *21*, 5−30.
(66) Das, P.; Moll, M.; Stamati, H.; Kavraki, L. E.; Clementi, C. Low-Dimensional, Free-Energy Landscapes of Protein-Folding Reactions by Nonlinear Dimensionality Reduction. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 9885−9890.
(67) Rohrdanz, M. A.; Zheng, W. W.; Maggioni, M.; Clementi, C. Determination of Reaction Coordinates via Locally Scaled Diffusion Map. *J. Chem. Phys.* **2011**, *134*, 124116.
(68) Zheng, W. W.; Rohrdanz, M. A.; Maggioni, M.; Clementi, C. Polymer Reversal Rate Calculated via Locally Scaled Diffusion Map. *J. Chem. Phys.* **2011**, *134*, 144109.

DOI: 10.1021/acs.jctc.7b00394
J. Chem. Theory Comput. 2017, 13, 4611−4623

(69) 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.
(70) Tribello, G. A.; Ceriotti, M.; Parrinello, M. Using Sketch-Map Coordinates to Analyze and Bias Molecular Dynamics Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 5196−5201.
(71) Virshup, A. M.; Chen, J. H.; Martínez, T. J. Nonlinear Dimensionality Reduction for Nonadiabatic Dynamics: The Influence of Conical Intersection Topography on Population Transfer Rates. *J. Chem. Phys.* **2012**, *137*, 22A519.
(72) Belyaev, A. K.; Domcke, W.; Lasser, C.; Trigila, G. Nonadiabatic Nuclear Dynamics of the Ammonia Cation Studied by Surface Hopping Classical Trajectory Calculations. *J. Chem. Phys.* **2015**, *142*, 104307.
(73) Zauleck, J. P. P.; Thallmair, S.; Loipersberger, M.; de Vivie-Riedle, R. Two New Methods to Generate Internal Coordinates for Molecular Wave Packet Dynamics in Reduced Dimensions. *J. Chem. Theory Comput.* **2016**, *12*, 5698−5708.
(74) Capano, G.; Penfold, T. J.; Tavernelli, I.; Chergui, M. Photophysics of a Copper Phenanthroline Elucidated by Trajectory and Wavepacket-based Quantum Dynamics: A Synergetic Approach. *Phys. Chem. Chem. Phys.* **2017**, *19*, 19590.
(75) Granucci, G.; Persico, M. Critical Appraisal of the Fewest Switches Algorithm for Surface Hopping. *J. Chem. Phys.* **2007**, *126*, 134114.
(76) Kabsch, W. Solution for Best Rotation to Relate 2 Sets of Vectors. *Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr.* **1976**, *32*, 922−923.
(77) Kabsch, W. Discussion of Solution for Best Rotation to Relate 2 Sets of Vectors. *Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr.* **1978**, *34*, 827−828.
(78) Dijkstra, E. W. A Note on Two Problems in Connexion with Graphs. *Numerische Mathematik* **1959**, *1*, 269−271.
(79) Cormen, T. H.; Leiserson, C. E.; Rivest, R. L.; Stein, C. *Introduction to Algorithms*, 2nd ed.; MIT Press and McGraw−Hill, 2001; pp 595−601. ISBN 0-262-03293-7.
(80) Floyd, R. W. Algorithm-97 - Shortest Path. *Commun. ACM* **1962**, *5*, 345−345.
(81) Warshall, S. A Theorem on Boolean Matrices. *J. Assoc. Comput. Mach.* **1962**, *9*, 11−12.
(82) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, 2825−2830.
(83) Buitinck, L.; Louppe, G.; Blondel, M. et al. *API Design for Machine Learning Software: Experiences from the Scikit-learn Project*; European Conference on Machine Learning and Principles and Practices of Knowledge Discovery in Databases; 2013; https://arxiv.org/abs/1309.0238.
(84) Werner, H. J.; Knowles, P. J.; Knizia, G. et al. *MOLPRO*, version 2012.1, a package of ab initio programs; TTI, 2012.
(85) Barbatti, M.; Aquino, A. J. A.; Lischka, H. Ultrafast Two-Step Process in the Non- Adiabatic Relaxation of the $CH_2NH_2^+$ Molecule. *Mol. Phys.* **2006**, *104*, 1053−1060.
(86) Weingart, O.; Altoe, P.; Stenta, M.; Bottoni, A.; Orlandi, G.; Garavelli, M. Product Formation in Rhodopsin by Fast Hydrogen Motions. *Phys. Chem. Chem. Phys.* **2011**, *13*, 3645−3648.
(87) Vreven, T.; Bernardi, F.; Garavelli, M.; Olivucci, M.; Robb, M. A.; Schlegel, H. B. Ab Initio Photoisomerization Dynamics of a Simple Retinal Chromophore Model. *J. Am. Chem. Soc.* **1997**, *119*, 12687−12688.
(88) Schapiro, I.; Ryazantsev, M. N.; Frutos, L. M.; Ferré, N.; Lindh, R.; Olivucci, M. The Ultrafast Photoisomerizations of Rhodopsin and Bathorhodopsin Are Modulated by Bond Length Alternation and HOOP Driven Electronic Effects. *J. Am. Chem. Soc.* **2011**, *133*, 3354−3364.
(89) Rockwell, N. C.; Su, Y. S.; Lagarias, J. C. Phytochrome Structure and Signaling Mechanisms. *Annu. Rev. Plant Biol.* **2006**, *57*, 837−858.
(90) Alvey, R. M.; Biswas, A.; Schluchter, W. M.; Bryant, D. A. Attachment of Noncognate Chromophores to CpcA of Synechocystis sp PCC 6803 and Synechococcus sp PCC 7002 by Heterologous Expression in Escherichia coli. *Biochemistry* **2011**, *50*, 4890−4902.
(91) Strambi, A.; Durbeej, B. Initial Excited-State Relaxation of the Bilin Chromophores of Phytochromes: A Computational Study. *Photoch. Photobio. Sci.* **2011**, *10*, 569−579.
(92) Zhuang, X. H.; Wang, J.; Lan, Z. G. Tracking of the Molecular Motion in the Primary Event of Photoinduced Reactions of a Phytochromobilin Model. *J. Phys. Chem. B* **2013**, *117*, 15976−15986.
(93) Mroginski, M. A.; Murgida, D. H.; von Stetten, D.; Kneip, C.; Mark, F.; Hildebrandt, P. Determination of the Chromophore Structures in the Photoinduced Reaction Cycle of Phytochrome. *J. Am. Chem. Soc.* **2004**, *126*, 16734−16735.
(94) Mroginski, M. A.; Murgida, D. H.; Hildebrandt, P. The Chromophore Structural Changes during the Photocycle of Phytochrome: A Combined Resonance Raman and Quantum Chemical Approach. *Acc. Chem. Res.* **2007**, *40*, 258−266.
(95) Keal, T. W.; Koslowski, A.; Thiel, W. Comparison of Algorithms for Conical Intersection Optimisation Using Semiempirical Methods. *Theor. Chem. Acc.* **2007**, *118*, 837−844.
(96) Koslowski, A.; Beck, M. E.; Thiel, W. Implementation of a General Multireference Configuration Interaction Procedure with Analytic Gradients In a Semiempirical Context Using the Graphical Unitary Group Approach. *J. Comput. Chem.* **2003**, *24*, 714−726.
(97) Weber, W.; Thiel, W. Orthogonalization Corrections for Semiempirical Methods. *Theor. Chem. Acc.* **2000**, *103*, 495−506.
(98) Thiel, W. *MNDO Program*, version 7.0; Max-Planck-Institut für Kohlenforschung: Mülheim an der Ruhr, Germany, 2005.
(99) Rupp, M.; Tkatchenko, A.; Müller, K. R.; von Lilienfeld, O. A. Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning. *Phys. Rev. Lett.* **2012**, *108*, 058301.
(100) Behler, J.; Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Phys. Rev. Lett.* **2007**, *98*, 146401.
(101) Bartók, A. P.; Payne, M. C.; Kondor, R.; Csányi, G. Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons. *Phys. Rev. Lett.* **2010**, *104*, 136403.
(102) Bartók, A. P.; Kondor, R.; Csányi, G. On Representing Chemical Environments. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2013**, *87*, 184115.
(103) Majumder, M.; Hegger, S. E.; Dawes, R.; Manzhos, S.; Wang, X. G.; Tucker, C.; Li, J.; Guo, H. Explicitly Correlated MRCI-F12 Potential Energy Surfaces for Methane Fit with Several Permutation Invariant Schemes and Full-Dimensional Vibrational Calculations. *Mol. Phys.* **2015**, *113*, 1823−1833.
(104) Hansen, K.; Biegler, F.; Ramakrishnan, R.; Pronobis, W.; von Lilienfeld, O. A.; Müller, K. R.; Tkatchenko, A. Machine Learning Predictions of Molecular Properties: Accurate Many-Body Potentials and Nonlocality in Chemical Space. *J. Phys. Chem. Lett.* **2015**, *6*, 2326−2331.
(105) Hansen, K.; Montavon, G.; Biegler, F.; Fazli, S.; Rupp, M.; Scheffler, M.; von Lilienfeld, O. A.; Tkatchenko, A.; Müller, K. R. Assessment and Validation of Machine Learning Methods for Predicting Molecular Atomization Energies. *J. Chem. Theory Comput.* **2013**, *9*, 3404−3419.
(106) Behler, J. Perspective: Machine Learning Potentials for Atomistic Simulations. *J. Chem. Phys.* **2016**, *145*, 170901.

DOI: 10.1021/acs.jctc.7b00394
J. Chem. Theory Comput. 2017, 13, 4611−4623