Nonlinear dimensionality reduction for nonadiabatic dynamics: The influence of conical intersection topography on population transfer rates

Aaron M. Virshup, Jiahao Chen, and Todd J. Martínez

Citation: J. Chem. Phys. **137**, 22A519 (2012); doi: 10.1063/1.4742066
View online: http://dx.doi.org/10.1063/1.4742066
View Table of Contents: http://jcp.aip.org/resource/1/JCPSA6/v137/i22
Published by the American Institute of Physics.

Nonlinear dimensionality reduction for nonadiabatic dynamics: The influence of conical intersection topography on population transfer rates

Aaron M. Virshup,[1] Jiahao Chen,[2] and Todd J. Martínez[3,4]
[1]*Department of Chemistry, Duke University, Durham, North Carolina 27710, USA*
[2]*Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
[3]*Department of Chemistry and The PULSE Institute, Stanford University, Stanford, California 94305, USA*
[4]*SLAC National Accelerator Laboratory, Menlo Park, California 94025, USA*

(Received 20 May 2012; accepted 19 July 2012; published online 9 August 2012)

Conical intersections play a critical role in the nonadiabatic relaxation of excited electronic states. However, there are an infinite number of these intersections and it is difficult to predict which are actually relevant. Furthermore, traditional descriptors such as intrinsic reaction coordinates and steepest descent paths often fail to adequately characterize excited state reactions due to their highly nonequilibrium nature. To address these deficiencies in the characterization of excited state mechanisms, we apply a nonlinear dimensionality reduction scheme (diffusion mapping) to generate reaction coordinates directly from *ab initio* multiple spawning dynamics calculations. As illustrated with various examples of photoisomerization dynamics, excited state reaction pathways can be derived directly from simulation data without any *a priori* specification of relevant coordinates. Furthermore, diffusion maps also reveal the influence of intersection topography on the efficiency of electronic population transfer, providing further evidence that peaked intersections promote nonadiabatic transitions more effectively than sloped intersections. Our results demonstrate the usefulness of nonlinear dimensionality reduction techniques as powerful tools for elucidating reaction mechanisms beyond the statistical description of processes on ground state potential energy surfaces. © *2012 American Institute of Physics*. [http://dx.doi.org/10.1063/1.4742066]

I. INTRODUCTION

The faithful modeling of excited state reaction dynamics is a major challenge since it places great demands on both electronic structure theory (the calculation of electronic excited states for highly nonequilibrium molecular geometries) and dynamics (quantum effects must be included to describe the nonadiabatic transitions which allow for the transfer of energy from the electronic to nuclear degrees of freedom). *Ab initio* molecular dynamics (AIMD) methods, where the dynamics and electronic structure problems are solved simultaneously, have been developed for nonadiabatic dynamics[1–10] in order to avoid the need to fit potential energy surfaces and their couplings to analytic functional forms. An oft-quoted advantage of AIMD methods is that all nuclear degrees of freedom can be included in the dynamics, i.e., there is little or no incentive to create reduced dimensionality models. Ironically, while this is indeed an advantage for the realistic modeling of excited state dynamics, it can be an obstacle to extracting chemical understanding. Simply put, as more degrees of freedom (electronic and/or vibrational) are included, it can become harder to see which are the most important. Thus, a means of automatically identifying the important degrees of freedom would be most welcome.

For chemical processes occurring entirely on the ground electronic state, one can use statistical theories as a guide to identify important degrees of freedom. Thus, one identifies local minima and transition states connecting these minima as the first step in this process. One can further find minimal energy pathways connecting these points and thereby identify possible reactions and their mechanisms. This approach is not so straightforward for many excited state reactions because they are intrinsically far from equilibrium. Excited state reaction dynamics typically take place on femtosecond timescales, as the electronic excitation event is rapid and the molecule often finds itself on a steeply sloped part of the excited state potential energy surface after photon absorption. The nonequilibrium nature of photodynamics thus presents some difficulty in establishing quantitative rate theories. The standard methods for determining reaction paths and rates for ground state reactions rely primarily on the assumptions of transition state theory; in particular, the dynamical evolution of the system is either neglected entirely or at best treated only in a thermodynamically averaged sense. Reaction coordinates arising from such assumptions, such as the one-dimensional intrinsic reaction coordinate (IRC) that parameterizes the minimal energy pathway,[11] are therefore unlikely to provide a realistic description of ultrafast processes which are completed in picoseconds or even femtoseconds. This problem is not unique to excited state reactions: dynamical correlations are already known to disrupt the minimal energy path picture even in some thermally activated ground state reactions.[12–14] While transition path sampling methods based on dynamical trajectories have been developed with great success as a means of correctly sampling a variety of candidate reaction pathways,[15] they have so far only been explored for reactions on the ground electronic state.

Conical intersections (CIs), i.e., molecular geometries where two or more electronic states are exactly degenerate,

are now recognized to play a central role in ultrafast photochemistry.[16,17] The Born-Oppenheimer approximation breaks down at and near these geometries, thus allowing nonadiabatic population transfer between different electronic states. Thus, these geometries may be viewed as analogs to the transition state in a ground state chemical reaction (but here for the "reaction" A* → A). However, CIs are not isolated points, but rather high-dimensional seams. For a CI involving two electronic states, the degeneracy between the potential energy surfaces (PESs) is lifted linearly in exactly two dimensions (the "branching plane") around the CI, forming the shape of a double cone. To first order, the two electronic states remain degenerate with respect to molecular displacements along all other vibrational degrees of freedom. Thus, for a two state intersection in a molecule with N vibrational degrees of freedom, the CI seam is an N-2 dimensional manifold. It can be difficult to predict which parts of this manifold will be most important in a given excited state reaction.

In this work, we aim to construct coarse-grained representations of dynamically accessed reaction paths and portions of the CI seam directly from dynamics simulations. One goal is to identify the dynamically accessible seam of CIs. In general, the shape of the CI manifold is difficult to determine, although in some cases, certain points on the CI manifold can be deduced from simple symmetry considerations. Furthermore, the high dimensionality of the seam usually represents far more degrees of freedom than can be visualized straightforwardly. Traditionally, the CI seam is characterized by its local minima, which are termed minimal energy CIs (MECIs).[18,19] This parallels the IRC approach for ground state reactions in providing a convenient set of molecular geometries that can be used to describe a reaction mechanism. However, the ultrafast nature of photochemical processes makes the relevance of these points uncertain. In fact, it has been shown in some cases that they have little or no role in excited state reaction mechanisms.[16,20,21] Nevertheless, they do retain some utility for descriptive purposes, i.e., as "signposts" on the PES that can be used to describe and distinguish alternative mechanisms, if only because there are few easily computed alternatives available.

In contrast, powerful statistical techniques exist for analyzing high-dimensional data sets, which could allow us to circumvent this somewhat *ad hoc* approach. These so-called dimensionality reduction techniques characterize general Riemannian manifolds by transformations of point clouds sampled from the manifold. Such methods take as input a set of datapoints and in return give a description of that dataset in terms of a reduced set of variables. Consider, for instance, a one-dimensional manifold (i.e., a curve or trajectory) embedded in a high dimensional space. Ideally, a dimensionality reduction technique applied to points sampled from this manifold should return a single coordinate describing motion along the manifold whose value can be interpreted as the displacement from some reference point on the curve. One of the most common and mature of these techniques is principal component analysis (PCA),[22] which fits, in the least-squares sense, an N-dimensional hyperplane to a point cloud embedded in $\mathbb{R}^N$. The variances associated with each axis of the hyperplane give a measure of the importance of each coordinate. Typically, only a small number, M, of axes are retained, thus creating an M-dimensional representation of the original data based only on the most strongly varying degrees of freedom and projecting out all other, more weakly varying dimensions. However, PCA's inherent linearity strongly limits its applicability to large-scale chemical dynamics. In particular, the assumption of linear variance of the coordinates in the hyperplane fitting process introduces a problematic dependence on the choice of coordinate system when the variations become large. For example, a large amplitude torsional motion may be viewed as a linear displacement in a dihedral angle or alternatively a highly nonlinear collective motion in the Cartesian coordinates of the atoms. In applications to chemical dynamics, this precludes the use of PCA for the characterization of many sequential reactions and the results are strongly dependent on the choice of coordinate system.

To address such issues, there has been increasing interest in nonlinear dimensionality reduction (NLDR) techniques.[23–26] As in PCA, these techniques quantify "important" coordinates, and project out "unimportant" degrees of freedom. However, in contrast to PCA, these techniques are often invariant to coordinate system, and can recover parameters describing nonlinear manifolds. Rather than focusing on data variance in a specified coordinate system, NLDR techniques construct low-dimensional subspaces that preserve short-range spatial relationships from the original high-dimensional space – these are constructed by projections that preserve distances between nearby datapoints in the original space, but not necessarily those between distant points. In larger, equilibrium chemical systems, NLDR techniques have been applied with notable successes; for instance, the ISOMAP NLDR technique has been shown to be effective in identification of reaction coordinates and transition states directly from dynamics data.[24] The diffusion map algorithm has also been applied to molecular systems, allowing a reaction coordinate to be identified[25] and its associated reduced equation of motion to be derived[27] without any *a priori* assumptions about the nature of this coordinate. We also employ diffusion maps in this paper. The diffusion map technique is particularly well suited for ultrafast nonadiabatic chemical dynamics because of its nonlinearity, coordinate independence, computational simplicity (relative to other NLDR techniques), and robustness to noise.[28]

In this paper, we specifically focus on two photoisomerization reactions that have already been well studied and have known excited state mechanisms. Our goal is to show that diffusion mapping recovers the known mechanisms from full dimensional dynamics data without any guidance or foreknowledge.

The first example is the photoisomerization of ethylene after excitation to the lowest $\pi\pi^*$ bright electronic state. Ethylene may be thought of as the paradigmatic molecule for photoisomerization, being perhaps the simplest molecule with a C=C double bond. It has therefore long been of both experimental and theoretical interest.[29–37] There is a barrierless path from the Franck-Condon point in ethylene to a variety of CIs connecting S_1 and S_0. Several geometries that have been highlighted in previous discussions of the photodynamics in ethylene are shown in Fig. 1. (Figure 1 also shows the energet-

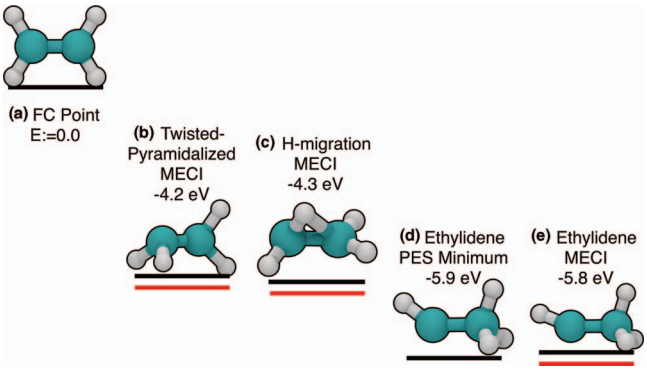

FIG. 1. Important points on the $\pi\pi^*$ excited state PES of ethylene, obtained from SA-3 CAS(2/2) / 6-31G** optimizations. Energy values are reported relative to the Franck-Condon (FC) point. Minimal energy conical intersections (MECIs) are doubly underlined. Note that the level of theory employed overestimates the Frank-Condon point energy, and that higher of levels of theory show that the ethylidene and pyramidalized intersections should, in fact, be more similar in energy.

ics that correspond to the level of theory we use in this paper for reference below. For more accurate values of these energies, we refer the reader to other publications as referenced above.) Because all of these structures are accessible from the Franck-Condon point with no barrier, dynamical effects play a crucial role in determining the excited state reactivity of ethylene. Furthermore, the structure of the reaction path is not easily represented linearly in standard internal coordinates, thus requiring *ad hoc*, or, as here, nonlinear definitions of the reaction coordinates and CI seam.

The second example is a model of the chromophore in photoactive yellow protein (PYP). This chromophore has two isomerizable bonds in the excited state, and thus multiple excited state reaction mechanisms may be expected. Interestingly, there are indications that the surrounding environment can strongly influence the branching ratio between these mechanisms.[38,39] It is also considerably larger than ethylene (21 atoms) and thus serves to show that the success of diffusion maps for the interpretation and analysis of excited state dynamics is not limited to small molecules.

We also use the diffusion map approach to address a long-standing question in excited state dynamics around CIs – namely, the role of intersection topography in the efficiency (or lack thereof) of nonadiabatic transitions.[40–42] As shown schematically in Fig. 2, the local topography around a CI may be "peaked" or "sloped"; this can be quantified using the overall slope introduced by Yarkony:[40]

$$\mathbf{S} = \lim_{\mathbf{R}_0 \to \mathbf{R}_{CI}} \nabla \left(\frac{E_1(\mathbf{R}) + E_2(\mathbf{R})}{2} \right) \Bigg|_{\mathbf{R}=\mathbf{R}_0}. \quad (1)$$

The length of the projection of this vector on the branching plane provides a measure of the sloped character of the intersection – for an ideal peaked intersection, the vector $\mathbf{S}$ has no projection on the branching plane. Dynamical considerations suggest that peaked intersections should more efficiently funnel trajectories towards the point of degeneracy $\mathbf{R}_{CI}$;[40–42] however, a quantitative relationship between these static PES properties and nonadiabatic dynamics has yet to be firmly established. By using diffusion maps, we are able to quantify

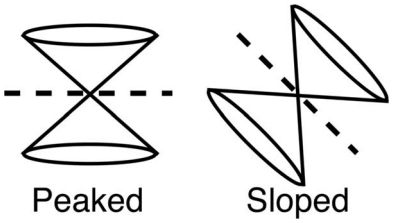

FIG. 2. Schematic depiction of peaked and sloped conical intersections (CIs). The character of the CI is distinguished by the norm of the gradient of the branching plane, represented by a dashed line. It has been suggested that peaked intersections more efficiently transfer population because their pronounced funnel shape tends to direct trajectories toward the point of degeneracy.

the influence of intersection topography on the efficiency of nonadiabatic transitions, demonstrating that peaked intersections are indeed more likely to lead to efficient population transfer.

II. METHODS

A. Diffusion maps

The process of finding low-dimensional manifolds embedded in high-dimensional spaces (NLDR) is an area of active algorithmic research. In this article, we employ diffusion maps, an example of one such recently developed algorithm, to quantitatively characterize the configuration space manifold underlying excited state reaction pathways sampled from *ab initio* multiple spawning (AIMS) simulations of the nonadiabatic dynamics. The use and construction of diffusion maps from high-dimensional data has been described extensively elsewhere[27,43–45] and we restrict ourselves to a brief overview here. The basic idea is to model a diffusion process over a set of points sampled from the manifold of interest. The diffusion process elucidates the underlying degrees of freedom by identifying local directions along which neighboring data points are connected, i.e., accessible via diffusion. In doing so, a manifold containing the entire data set can be traced out. Keeping track of the local directions employed over the course of the diffusion results in a parameterization of the original dataset with a dimensionality that tends toward that of the embedded manifold and not the larger dimension of the embedding space. It is important to note that "diffusion mapping" is a computational, not physical, process. Although diffusion maps have special connections to manifolds that arise from diffusive processes, they can be used to map geometrical manifolds of arbitrary origin, such as the excited state dynamics described here.[27]

The mapping takes as input a finite collection of points $\{\mathbf{x}_i\}_{i=1}^{M}$ in the original space. In this article, these points are molecular geometries ($\mathbf{x}_i \in \mathbb{R}^{3N}$) sampled from AIMS trajectories at regular time intervals, as described below. A diffusion kernel is constructed on the dataset, representing the unnormalized amplitude of a random walker diffusing from datapoint i to point j,

$$k_{ij} = e^{-(d_{ij}/D)^2}, \quad (2)$$

where d_{ij} is the distance between datapoints i and j, and D is an empirically chosen length scale as described below. A diffusion operator is then constructed from the kernel, representing the probability of transition from datapoint i to j:

$$a(\mathbf{x}_i, \mathbf{x}_j) = \frac{1}{Z_i} \frac{k_{ij}}{(v_i v_j)^\alpha}$$

$$\text{where } v_i = \sum_j k_{ij} \text{ and } Z_i = \sum_j \frac{k_{ij}}{(v_i v_j)^\alpha}, \tag{3}$$

where the parameter α determines the specific type of diffusion modeled (discussed below), and the quantities v_i and Z_i are normalization and renormalization constants, respectively. We refer the reader to earlier in-depth discussions of this renormalization procedure.[27,43]

An m-dimensional representation is found by solving for the m largest eigenvalues of the diffusion operator a. The m corresponding eigenvectors $\{\phi_j\}_{j=1,\ldots,m}$ provide the low-dimensional parameterization of the original dataset. In this respect, the diffusion operator plays a role analogous to that of the covariance matrix in PCA. However, while the original high-dimensional coordinates are the basis of the covariance matrix, the data points themselves are the basis of the diffusion kernel. Thus, the dimensionality of each eigenvector is not the dimensionality of the original coordinate system, as in PCA, but instead M, the number of data points used to construct the mapping. The low dimensional representation of the i^{th} datapoint, x'_i, is given by the i^{th} components of the dominant m eigenvectors,

$$x'_i := (\phi_1(i), \phi_2(i), \ldots, \phi_m(i)). \tag{4}$$

Thus, we have constructed a mapping of the original dataset from $\mathbb{R}^{3N}$ to $\mathbb{R}^m$. We refer to this low-dimensional space as the "reduced space," and the m new coordinates as the "diffusion coordinates." The dimensionality of the parameterization, m, is determined by examination of the eigenvalue spectrum; typically, all eigenvectors corresponding to eigenvalues beneath the first significant spectral gap are discarded.

The parameter D sets the distance scale of the diffusion kernel (cf. Eq. (2)). This should be much larger than the nearest neighbor spacings between data points to smooth out discretization effects, but small enough so that large-scale features of the manifold are not obscured. The choice of α depends on the significance of data point density; $\alpha = 0$ corresponds to standard Laplacian diffusion, which is heavily influenced by the local density of datapoints; $\alpha = 1$ gives an approximation to Laplace-Beltrami heat diffusion, resulting in a diffusion map that is invariant the density of data points; and $\alpha = 1/2$ approximates the backward Fokker-Planck operator on the dataset, resulting in a map that emphasizes, but is not dominated by, the density of sampled points along the manifold.[44] For this work, we employ $\alpha = 1$, as it provides an approximately consistent reproduction of distances across the manifold (that is, the magnitude of the Jacobian between diffusion space and coordinate space depends only weakly on location).

Finally, a distance between data points needs to be defined (d_{ij} in Eq. (2)). Here we use the root-mean-square displacement between each pair of geometries, after moving the two geometries into maximum coincidence through rotation[46] and symmetry operations (including permutation of identical nuclei).

B. *Ab initio* multiple spawning dynamics

The dynamical results presented here were generated with the *ab initio* multiple spawning (AIMS) technique for molecular quantum dynamics, which has been described extensively elsewhere.[4,5,47] The AIMS wavefunction is a linear combination of Born-Oppenheimer basis functions

$$\Psi_{\text{AIMS}}(\mathbf{r}, \mathbf{R}, t) = \sum_{i=1}^{N(t)} c_i(t)\, \psi_{M_i}(\mathbf{r}; \mathbf{R})\, \chi_i(\mathbf{R}; \tilde{\mathbf{R}}_i(t), \tilde{\mathbf{P}}_i(t)), \tag{5}$$

where ψ_J is the electronic wavefunction on the Jth adiabatic state, $\chi(\mathbf{R}; \tilde{\mathbf{R}}(t), \tilde{\mathbf{P}}(t))$ is a Gaussian wavepacket centered at $\tilde{\mathbf{R}}(t)$ with momentum $\tilde{\mathbf{P}}(t)$, and c_i(t) are the complex amplitudes associated with each vibronic basis function. The parameters $\tilde{\mathbf{R}}(t)$ and $\tilde{\mathbf{P}}(t)$ have an obvious classical interpretation as nuclear positions and momenta, and are propagated classically on the PES for the electronic state with which they are associated. Hence, the vibronic basis functions are denoted trajectory basis functions (TBFs). At points where the Born-Oppenheimer approximation breaks down, i.e., in the vicinity of conical intersections, population can be transferred between electronic states (through the evolution of the complex amplitudes as dictated by the solution of the Schrödinger equation in the time-evolving basis set). In order to ensure that the appropriate basis functions are available to represent this population transfer, new TBFs are "spawned" when the coupling between electronic states exceeds a predetermined threshold – it is for this reason that the upper limit N(t) in the sum of Eq. (5) is explicitly time-dependent.

The PESs for dynamics were generated on the fly using *ab initio* quantum chemical techniques. The electronic structure calculations of potential energy surfaces, gradients, and nonadiabatic couplings were carried out using MOLPRO.[9,48] Specifically, the calculations for ethylene used the state-averaged complete active space self-consistent field (SA-CASSCF) method,[49,50] with state averaging over the lowest three singlet states, an active space of two electrons in two orbitals, and the 6-31G** basis set,[51] i.e., SA-3-CAS(2/2)/6-31G**. For the photoactive yellow protein (PYP) chromophore, the electronic wavefunction is described with the SA-2-CAS(8/6)/6-31G* method. Initial conditions in both cases are sampled from the harmonic approximation to the Wigner distribution of the ground vibrational state.

III. RESULTS

A. Ethylene

The first 200 fs of ethylene photodynamics following vertical excitation from the ground state minimum to the $\pi\pi^*$ state (S_1 for the electronic wavefunction ansatz used here) were simulated with 600 uncoupled AIMS simulations (each starting with a single TBF and spawning new TBFs as dictated by the nonadiabatic coupling matrix elements). Due to computational requirements (specifically, the cost of diago-

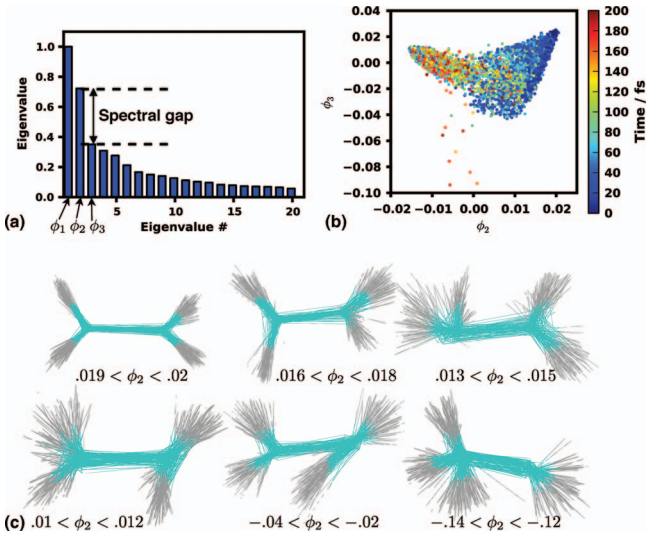

FIG. 3. (a) Eigenvalue spectrum for diffusion map from ethylene S_1 dynamics. There is a large spectral gap after the first nontrivial eigenvalue (ϕ_2), indicating that the dynamics are predominantly one-dimensional. (b) Locations of the sampled geometries as a function of ϕ_2 and ϕ_3, colored to indicate the time following photoexcitation. (c) Overlays of all geometries at several different values of ϕ_2, illustrating the "width" of the reaction path at several different points. Geometries are clustered fairly tightly at each value, illustrating that ϕ_2 is the dominant controlling variable for the dynamics.

nalizing the $M \times M$ matrix representing the diffusion kernel, where M is the number of geometries), the diffusion map was constructed using geometries sampled every 5 fs from the initial excited state trajectory in the first 200 simulations, yielding a total of 6738 molecular geometries.

The distance between each pair of molecular geometries was computed as described above. This pairwise distance matrix was then transformed into a diffusion kernel using Eqs. (2) and (3) with $\alpha = 1$ and $D = 0.5$ Å. This diffusion kernel was then diagonalized. Note that the first eigenvector (ϕ_1) of this and any other diffusion kernel is trivial, with a uniform value for the entire data set – this indicates that all data points occupy a single, compact cluster.[27] As shown in Fig. 3(a), the eigenvalue spectrum of the diffusion kernel from the ethylene excited state dynamics simulations contains a large spectral gap after the second eigenvector. Therefore, this diffusion map describes a dominantly one-dimensional reaction path parameterized by the second eigenvector ϕ_2. In Fig. 3(b), we show the geometries (data points) in the two-dimensional space characterized by the second and third eigenvectors (ϕ_2 and ϕ_3), colored according to the time when the geometry was accessed in the dynamics ($t = 0$ corresponds to the photoexcitation event). The coordinates of the ith geometry (data point) along the ϕ_2 and ϕ_3 axes are given by the ith element of the second and third diffusion kernel eigenvectors, as given in Eq. (4). Of course, restricting the diffusion map to a single diffusion coordinate necessarily implies that there are many molecular geometries which are described by the same value of the diffusion coordinate. One can get some insight into this by superimposing molecular geometries corresponding to restricted ranges of the diffusion coordinate. This is done in Fig. 3(c), where a set of geometries from the data set generating the diffusion kernel are superimposed according to their diffusion coordinate value. Geometries with similar values of the diffusion coordinate can be seen to be quite similar, but with a variance that reflects the "width" of the data set, i.e., the degree to which further diffusion coordinates would need to be included to completely specify the geometries in the data set.

The values of the diffusion coordinates for the geometries in the data set as given in Eq. (4) are expressed in terms of unitless "diffusion modes." In order to convert these to more meaningful real space distances, we compute the magnitude of the Jacobian between the diffusion coordinate ϕ_2 and the real space coordinates $\mathbf{R}$. The value of the Jacobian at ϕ can be calculated by averaged finite difference,

$$J(\phi) = \left.\frac{\partial \mathbf{R}}{\partial \phi_2}\right|_{\phi} = \lim_{\delta \to 0} \left\langle \frac{\mathbf{R} - \bar{\mathbf{R}}}{\phi_2 - \phi} \right\rangle_{\{\phi-\delta, \phi+\delta\}}, \tag{6}$$

where $\langle \rangle_{\{a,b\}}$ indicates an average over all points where $a < \phi_2 < b$, and $\bar{\mathbf{R}}$ is the average value of $\mathbf{R}$ in the same interval. The real-space distance along this coordinate can then be calculated as

$$\mathrm{DRC}(\phi_2) = \int_O^{\phi_2} |J(\phi)|\, d\phi, \tag{7}$$

where O is an arbitrarily chosen reference point (here, the Franck-Condon geometry as depicted in Fig. 2).

Selected molecular geometries corresponding to motion along this coordinate are shown in Fig. 4. The coordinate is here referred to as the "dynamical" (as opposed to intrinsic) reaction coordinate (DRC). This one-dimensional excited state reaction path is highly nonlinear in both Cartesian and internal coordinates. Motion along the DRC schematically follows the excited state reaction mechanism described above. At low values of the dynamical reaction coordinate (DRC), the molecule is planar and similar to the Franck-Condon geometry. More positive values of the DRC correspond to twisting and pyramidalization about the C=C bond, followed by hydrogen atom migration across the C=C bond to form ethylidene; finally, the newly formed methyl group twists and bends. The time evolution of several representative TBFs along the DRC is also shown. Initially, the TBFs

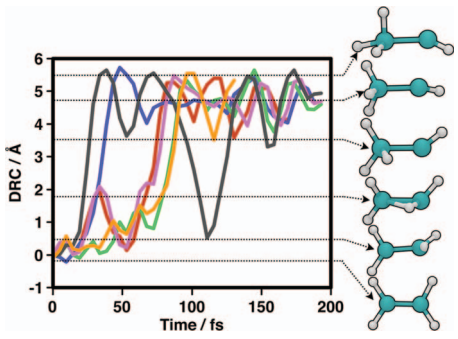

FIG. 4. Several representative trajectory basis functions (TBFs) along the dynamical reaction coordinate (DRC) of the S_1 state of ethylene and their corresponding geometries. The unitless diffusion coordinate ϕ_2 (see Fig. 3), was converted to the DRC distance shown here using Eqs. (6) and (7), with the Franck-Condon point set to a DRC value of 0 Å. TBFs may undergo torsional motion for tens of fs (DRC 0–0.5 Å) before proceeding through pyramidalization (DRC 0.5–1.5 Å) and hydrogen migration (DRC 1.5–3 Å) to form ethylidene (DRC 3–6 Å).

oscillate for a short time between the planar and twisted geometries, before falling through the pyramidalized structures to the ethylidene minimum.

Even though we constructed the diffusion map from a limited subset of simulations, it is possible in practice to map the data from all 600 original simulations into the reduced dimensionality space. The diffusion coordinates for geometries in the 400 trajectories that were not included in the construction of the diffusion map can be estimated with the Nyström extension.[52] For each new geometry $\mathbf{y}$, a new diffusion kernel b is constructed which includes the new point as well as all the original geometries in the set S (those used to construct the diffusion map). The reduced coordinates for the new point are given by

$$y_i' = \frac{1}{\lambda_i} \sum_{x_j \in S} b(\mathbf{x}_j, \mathbf{y}) \phi_i(\mathbf{x}_j) \tag{8}$$

where λ_i and $\phi_i(\mathbf{x}_j)$ are the eigenvalues and eigenvectors from the original diagonalization of a. This so-called "restriction" provides a transformation from configuration space $\mathbb{R}^{3N}$ to the reduced dimensional space $\mathbb{R}^m$. Using this procedure, we can carry out a reduced dimensionality analysis of the data from all 600 dynamical AIMS simulations, even though only 200 were utilized in the construction of the diffusion map. We note in passing that the inverse "lifting" transformation (determining the molecular geometries in the full 3N-dimensional space corresponding to a given point in the reduced dimensional space) is necessarily more complicated since information has been discarded in the reduced space. Attempts to define such a transformation have so far been restricted to an inverse transformation in a stochastic sense, for example by running stochastic dynamics with a constraint on the reduced space coordinates.[52] In the present work, we do not require this "lifting" transformation and we do not consider it further.

The value of the excited state potential energy for each sampled geometry along the DRC is shown in Fig. 5, where the important geometries from Fig. 2 are also shown. A reaction path similar to the one in Fig. 2 is recovered, but with most points lying well above the excited state minima. This is to be expected given that the Franck-Condon point is several eV above the twisted/pyramidalized MECI as well as the other depicted MECIs and minima on S_1.

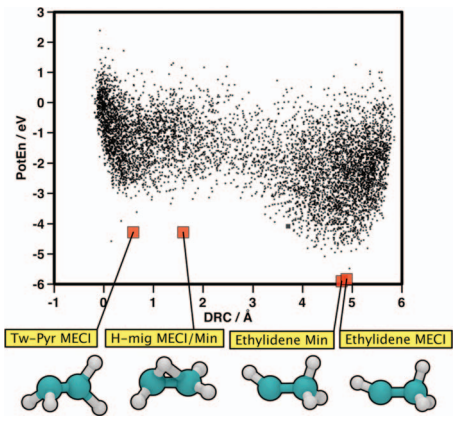

FIG. 5. Energies of sampled points from dynamics along the reaction path of the S_1 state of ethylene.

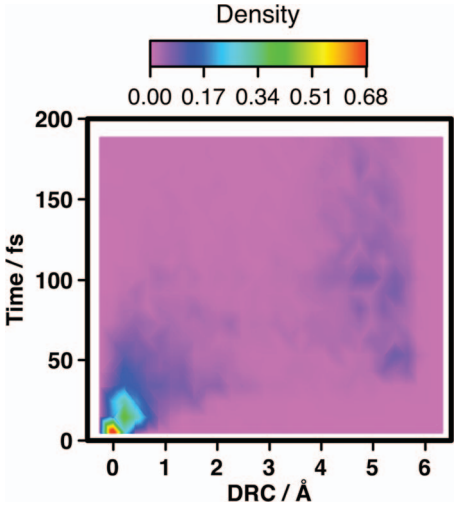

FIG. 6. Time evolution of the reduced probability density on the excited electronic state as a function of the DRC for the photodynamics of ethylene.

We can also represent the nuclear wavepacket probability density as a function of the diffusion coordinate, using the same Monte Carlo integration that allows us to represent the time-evolving wavepacket in an arbitrary set of internal coordinates.[53] In Fig. 6, we show the time evolution of the probability density on the excited electronic state in terms of the DRC. Note that as the reaction progresses, the total wavefunction density drops off quickly as population quenches to the ground state.

In order to characterize the dynamically relevant portion of the CI seam, we select the molecular geometries at which "spawning" takes place, i.e., the initial locations of the new TBFs that are added to the simulation to allow the description of nonadiabatic events. We choose these geometries as those with the largest magnitude of the nonadiabatic coupling vector connecting S_0 and S_1 during the placement of the newly spawned TBF. Population transfer to these newly spawned TBFs is evaluated by monitoring the complex amplitudes c(t) as given in Eq. (5) until the associated population (evaluated as a Löwdin-like population[54] to account for nonorthogonality) of the TBF becomes approximately constant, typically within twenty femtoseconds of the introduction of the new TBF. In Fig. 7, the population transferred to ground state TBFs is shown as a function of both time and the location of the associated spawning geometries in the reduced space. The spawning geometries were projected in the reduced space using the Nyström extension as discussed above. Population transfer is localized to geometries in two regions of the DRC. As can be determined from Fig. 4, the first of these (0.5–1.5 Å) corresponds to twisted/pyramidalized geometries and the second (3–6 Å) region corresponds to ethylidene-like geometries where a hydrogen atom has migrated to form CH_3CH. Furthermore, there is a strong temporal/spatial correlation with early population transfer occurring in the twisted/pyramidalized region and later population transfer events being mediated by ethylidene-like geometries.

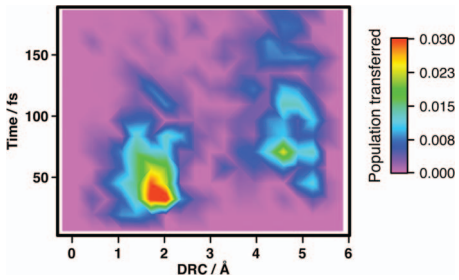

FIG. 7. Population transferred to ground state as a function of DRC and time. Population is initially quenched in the pyramidalized region, while the longer timescale decay takes place in the ethylidene region.

The efficiency of population decay through nonadiabatic transitions is not constant along the reaction path and it would be interesting to correlate the transition rates with features of the potential energy surface such as the energy or topography of the closest conical intersection. We evaluate a time-averaged rate constant for ground state quenching at each value along the excited state DRC directly from the simulation data. Specifically, we compare the amount of population transfer through each point to the total wavepacket probability density that passed through that region of the path

$$\alpha(x) = \frac{T(x)}{\int_{t_i}^{t_f} dt \rho(x,t)}, \tag{9}$$

where x is the DRC, $T(x)$ is the amount of population transferred through that point (integral over the time coordinate in Fig. 7), and $\rho(x, t)$ is the probability density shown in Fig. 6. The DRC-dependent nonadiabatic transition rate $\alpha(x)$ was evaluated by a histogramming procedure, integrating over the time coordinate in Figs. 6 and 7. The upper panel of Fig. 8 shows the results, where the red bars denote the amount of population transferred to the ground state as a function of the DRC and the blue bars denote the time-integrated probability

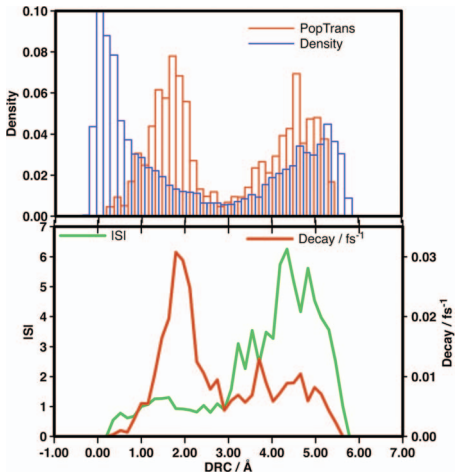

FIG. 8. Upper panel: Population transferred (red) and time-integrated wavepacket probability density (blue) along the diffusion reaction coordinate of the S_1 state of ethylene. Lower panel: Average slope of conical intersections closest to the spawning points (green, left axis) and effective population decay rate (red, right axis). Although the pyramidalized/hydrogen geometries are higher in energy than ethylidene (with the electronic structure ansatz used in this work), they nevertheless serve to promote nonadiabatic transitions more effectively.

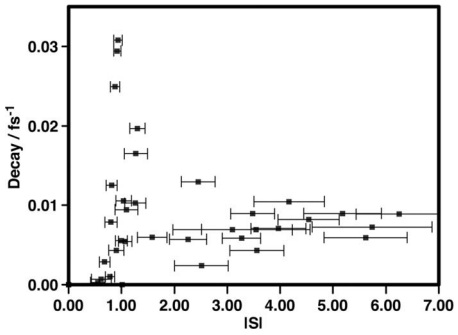

FIG. 9. Relationship between the rate of population transfer induced by nonadiabatic transitions and the slope of the nearest conical intersection (obtained from the data in Fig. 8) in ethylene. Error bars indicate the width of the distribution of slope parameters for intersections corresponding to the data point.

of finding the nuclei in a geometry corresponding to a particular value of the DRC.

We then calculated the average slope of conical intersections as a function of the DRC by calculating the magnitude of the overall slope vector S (Eq. (1)) for each spawning geometry. The average of this distribution for each value of the DRC is shown as the green line in the lower panel of Fig. 8, showing that intersections in the twisted/pyramidalized region of the DRC tend to be quite peaked (small overall slope) and those in the ethylidene region of the DRC tend to be strongly sloped. This panel also shows the nonadiabatic transition rate (cf. Eq. (9)) as a red line, demonstrating that nonadiabatic transitions are far more efficient in the region of the DRC corresponding to twisted/pyramidalized geometries, as compared to the ethylidene geometries. From this data, it is easy to extract the rate of nonadiabatic transitions as a function of conical intersection slope. This is shown by plotting the decay rate as a function of the intersection slope in Fig. 9. The error bars along the x axis correspond to the width of the distribution of intersection slopes for each value of the DRC. Of course, there is only limited data and this is for a specific example, but Fig. 9 does show that highly sloped intersections tend to lead to less efficient nonadiabatic transitions, while the most efficient nonadiabatic transitions occur around peaked intersections. Although there is no clear one-to-one relationship between intersection slope and nonadiabatic transition rate (and neither is this to be expected since the momentum of the nuclei should also play a role), the maximal values along the plot suggest a limiting relationship, supporting previous suggestions[40–42] about the importance of intersection topography.

B. Photoactive yellow protein chromophore

As a second example, showing that diffusion mapping also successfully recovers low-dimensional manifolds of the dynamics of larger systems, we turn to the case of photoisomerization in the chromophore of photoactive yellow protein (PYP). We choose a model chromophore here, anionic *p*-coumaric acid (**pCK-**), shown in Fig. 10. We have previously discussed the photodynamics of this chromophore in isolation and in condensed phase environments.[10,38,55]

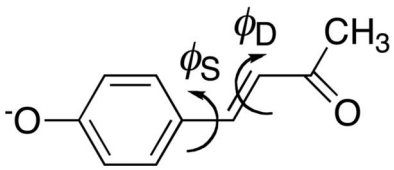

FIG. 10. The *p*-coumaric acid anion, a model for the chromophore of photoactive yellow protein (PYP), denoted as **pCK-** in the text. After photoexcitation, there are two possible photoisomerization pathways – torsion about the phenyl-adjacent single bond (ϕ_S) or the side chain double bond (ϕ_D), which are indicated in the figure.

Upon photoabsorption in the protein environment, the PYP chromophore undergoes *trans-cis* isomerization, initiating the negative photoaxis pathway of *Halorhodospira halophile*.[56] Various experimental[57,58] and theoretical[55,59–64] studies have shown rotation of the ethylenic double bond (D) and the phenyl-adjacent single bond (S) are the principal reaction coordinates and these can compete with each other.

Here, we simulate the excited state dynamics of isolated **pCK-** with 16 initial AIMS trajectories. Electronic structure calculations were at the SA2-CAS(8/6) level of theory with the 6-31G* basis set. In contrast to ethylene, the larger size and longer excited state lifetime of **pCK-** limits the amount of dynamical data available. Even with a relatively small dataset, however, the diffusion mapping recovers a one-dimensional representation of the reaction path that agrees with the qualitative picture of the reaction proceeding through torsion about either the S or D bond (these torsions are indicated in Fig. 10). In contrast to ethylene, where the excited state reaction dynamics largely follows a sequential path from the Franck-Condon region to twisted/pyramidalized and finally ethylidene geometries, two potential outcomes compete in the photodynamics of **pCK-**. Specifically, the molecule starts in the Franck-Condon region, and then may isomerize about either the S or the D bond. As shown in Fig. 11, the dynamical reaction coordinate revealed by diffusion mapping captures this behavior nicely. Negative values of the DRC correspond to torsion about the D bond, and positive values correspond to torsion about the S bond. Geometries at the extremes (both positive and negative) of the DRC involve some degree of torsion about both bonds. Of course, a more complete picture would be obtained by including the second eigenvector of the diffusion operator to build a two-dimensional reduced space picture of the excited state dynamics.

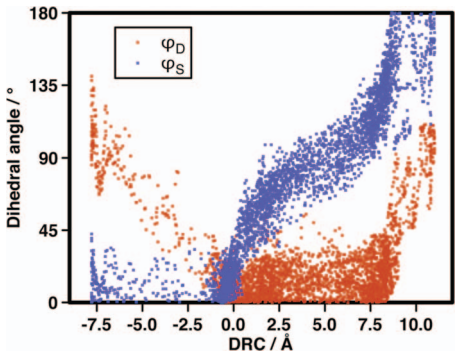

FIG. 11. Plot of the two relevant torsion angles along the DRC for the model PYP chromophore. The diffusion map correctly identifies these as the two most important coordinates in the photodynamics of the chromophore, as shown by the fact that the DRC is dominated by the ϕ_S torsion angle for the right half of the plot (from 0 to 9 Å) and by the ϕ_D torsion angle for the left half of the plot (from 0 to −7.5 Å).

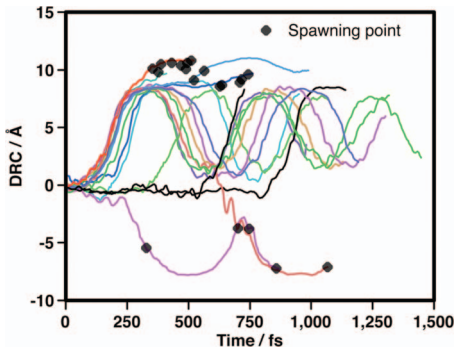

FIG. 12. Time evolution of the TBFs along the DRC of the model PYP chromophore. The majority undergo torsion about the phenyl-adjacent single bond (denoted by positive values of the DRC). Points where new TBFs are spawned are indicated by diamonds, and comparison with Fig. 11 shows that these correspond to a ϕ_D torsion angle around 90°.

As shown in Fig. 12, most of the TBFs undergo torsion around the S bond after photoexcitation and continue to oscillate about $\phi_S = 90°$ (DRC = 5 Å) for at least 1 ps. However, no spawning events occur in this region of the reaction path. Instead, spawning events occur only in regions where the molecule is twisted about the D bond. This can occur in two ways; either by immediate twisting about the D bond or by torsion about the S bond followed by torsion about the D bond (so that both bonds are twisted in a "hula-twist" like motion).[65]

IV. CONCLUSIONS

Our results strongly suggest that diffusion maps provide a useful low-dimensional description of ultrafast nonadiabatic dynamics on excited state manifolds, as demonstrated in the two reactions studied. For the $\pi\pi^*$ excited state dynamics of ethylene, the reduced dimensional model produced a one-coordinate representation that agreed with previous dynamical studies. The dynamical reaction coordinate derived from the diffusion map has a natural interpretation of a highly nonlinear interpolation between torsional rotation, pyramidalization, and hydrogen migration coordinates, without any *a priori* choice of coordinate parameterization. This nonlinearity along an unusual set of coordinates underlines the need for nonlinear analyses of such processes, as linear methods such as PCA will only yield useful results if a "good" set of reaction coordinates is known in advance.

We find an inverse correlation between overall CI slope and the rate of population transfer, as has been suggested in previous work. The peaked (unsloped) CI transfers population more efficiently, despite being higher in energy than the sloped CI. While suggestive of an overall trend, our results also show a need to disentangle aspects of dynamics on the overall PES from those specifically related to the CI topography, so as to further elucidate the detailed relationship between CI topography and nonadiabatic transition rates.

Diffusion map analysis also identified a one-dimensional manifold in the excited state dynamics of anionic *p*-coumaric acid, a model system for the chromophore of photoactive yellow protein. The dynamical reaction coordinate picks out both of the two competing torsional relaxations via the ethylenic double bond (D) and the phenyl-adjacent single bond (S), and thus allowed the classification of trajectories into two categories based on which torsional relaxation pathway was chosen. The data furthermore showed a clear trend of triggering spawning events only when there is significant torsion about the D bond.

The approach in this work can be generalized in a number of different ways. The distance metrics presented here were based only on the 3*N*-dimensional molecular conformation and were restricted to a single electronic surface. A distance metric based on 6*N*-dimensional phase space would be useful in further disentangling the dynamical and static aspects of population transfer near conical intersections. Additionally, an extended distance metric that included both nuclear configurations and electronic information – for instance, allowing nonadiabatic mixing to bring trajectories on different electronic states close to one another – could provide a picture of dynamical time-evolution across multiple electronic states. The NLDR techniques presented here can also be straightforwardly applied to other semi-classical models of chemical dynamics, such as surface hopping[66] and wavepacket dynamics,[67] and a diffusion map of a dynamical quantum system could easily be constructed on a collection of points sampled from the probability density corresponding to the nuclear wavefunction.

We have demonstrated that nonlinear dimensionality reduction techniques such as diffusion maps are powerful new statistical tools for analyzing dynamics simulation data. Dynamically relevant reaction coordinates can be deduced automatically without the need for *a priori* specification of interesting coordinates. Furthermore, these reaction coordinates have illuminated qualitative features of the dynamics themselves, such as the nature and location of spawning points in *ab initio* multiple spawning simulations, the branching ratios of different reaction channels, and the dependence of nonadiabatic transitions on the slope of CIs. The dynamical reaction coordinates remain well defined even far from equilibrium, and thus afford a new and valuable tool for understanding ultrafast processes on electronically excited states.

Although we have restricted our use of diffusion maps to *analysis* of data from dynamics carried out in full dimensionality, one can also imagine using the resulting diffusion map to construct a reduced dimensionality model that might be used in its own right. For example, an explicit reduced dimensionality model could provide a means of reaching longer time scales (as has been explored for ground state reactions[27]) or calculating the wavepacket evolution with more nearly numerically exact quantum dynamics methods.[68] This is an avenue which might be profitable in future work, but the nonequilibrium character of excited state reactions may introduce difficulties in determining how to treat the degrees of freedom outside the set of coordinates selected by the diffusion map.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation through Grant No. CHE-11-24515 with computational support through DOE (Contract No. DE-AC02-7600515). We thank Professor Dr. Dirk Hundertmark (Karlsruher Institut für Technologie) for helpful discussions.

[1] T. J. Martinez and R. D. Levine, J. Chem. Phys. **105**, 6334 (1996).
[2] T. J. Martinez, Chem. Phys. Lett. **272**, 139 (1997).
[3] T. Vreven, F. Bernardi, M. Garavelli, M. Olivucci, M. A. Robb, and H. B. Schlegel, J. Amer. Chem. Soc. **119**, 12687 (1997).
[4] M. Ben-Nun, J. Quenneville, and T. J. Martínez, J. Phys. Chem. A **104**, 5161 (2000).
[5] M. Ben-Nun and T. J. Martínez, Adv. Chem. Phys. **121**, 439 (2002).
[6] N. L. Doltsinis and D. Marx, Phys. Rev. Lett. **88**, 166402 (2002).
[7] B. Lasorne, M. A. Robb, and G. A. Worth, Phys. Chem. Chem. Phys. **9**, 3210 (2007).
[8] M. Barbatti, M. Ruckenbauer, J. J. Symczak, A. J. A. Aquino, and H. Lischka, Phys. Chem. Chem. Phys. **10**, 482 (2008).
[9] B. G. Levine, J. D. Coe, A. M. Virshup, and T. J. Martínez, Chem. Phys. **347**, 3 (2008).
[10] A. M. Virshup, C. Punwong, T. V. Pogorelov, B. Lindquist, C. Ko, and T. J. Martínez, J. Phys. Chem. B **113B**, 3280 (2009).
[11] K. Fukui, J. Phys. Chem. **74**, 4161 (1970).
[12] S. C. Ammal, H. Yamataka, M. Aida, and M. Dupuis, Science **299**, 1555 (2003).
[13] U. Lourderaj, K. Park, and W. L. Hase, Int. Rev. Phys. Chem. **27**, 361 (2008).
[14] L. Sun, K. Song, and W. L. Hase, Science **296**, 875 (2002).
[15] P. G. Bolhuis, C. Dellago, and D. Chandler, Faraday Discuss. **110**, 421 (1998).
[16] B. G. Levine and T. J. Martinez, Ann. Rev. Phys. Chem **58**, 613 (2007).
[17] D. R. Yarkony, Rev. Mod. Phys. **68**, 985 (1996).
[18] M. J. Bearpark, M. Robb, and H. B. Schlegel, Chem. Phys. Lett. **223**, 269 (1994).
[19] M. R. Manaa and D. R. Yarkony, J. Chem. Phys. **99**, 5251 (1993).
[20] J. D. Coe, M. T. Ong, B. G. Levine, and T. J. Martinez, J. Phys. Chem. A **112**, 12559 (2008).
[21] P. Krause and S. Matsika, J. Chem. Phys. **136**, 034110 (2012).
[22] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning* (Springer Science, New York, 2009).
[23] J. B. Tenebaum, V. de Silva, and J. C. Langford, Science **290**, 2319 (2000).
[24] P. Das, M. Moll, H. Stamati, L. E. Kavraki, and C. Clementi, Proc. Natl. Acad. Sci. U.S.A. **103**, 9885 (2006).
[25] M. A. Rohrdanz, W. Zheng, M. Maggioni, and C. Clementi, J. Chem. Phys. **134**, 124116 (2011).
[26] M. Ceriotti, G. A. Tribello, and M. Parrinello, Proc. Natl. Acad. Sci. U.S.A. **108**, 13023 (2011).
[27] B. Nadler, S. Lafon, R. R. Coifman, and I. G. Kevrekidis, Appl. Comput. Harmon. Anal. **21**, 113 (2006).
[28] A. Singer and R. R. Coifman, Appl. Comp. Harm. Anal. **25**, 26 (2008).
[29] M. Ben-Nun and T. J. Martínez, Chem. Phys. **259**, 237 (2000).
[30] J. Quenneville and T. J. Martínez, J. Phys. Chem. A **107**, 829 (2003).
[31] M. Barbatti, J. Paier, and H. Lischka, J. Chem. Phys. **121**, 11614 (2004).
[32] M. Barbatti, M. Ruckenbauer, and H. Lischka, J. Chem. Phys. **122**, 174307 (2005).
[33] K. Kosma, S. A. Trushin, W. Fuss, and W. E. Schmid, J. Phys. Chem. A **112**, 7514 (2008).
[34] H. Tao, T. K. Allison, T. W. Wright, A. M. Stooke, C. Khurmi, J. van Tilborg, Y. Liu, R. W. Falcone, A. Belkacem, and T. J. Martinez, J. Chem. Phys. **134**, 244306 (2011).
[35] T. K. Allison, H. Tao, W. J. Glover, T. W. Wright, A. M. Stooke, C. Khurmi, J. Van Tilborg, Y. Liu, R. W. Falcone, T. J. Martínez, and A. Belkacem, J. Chem. Phys. **136**, 124317 (2012).
[36] T. Mori, W. J. Glover, M. S. Schuurman, and T. J. Martínez, J. Phys. Chem. A **116**, 2808 (2012).
[37] E. F. Cromwell, A. Stolow, M. J. J. Vrakking, and Y. T. Lee, J. Chem. Phys. **97**, 4029 (1992).
[38] C. Ko, A. M. Virshup, and T. J. Martinez, Chem. Phys. Lett. **460**, 272 (2008).

[39]G. Groenhof, M. Bouxin-Cadematory, B. Hess, S. P. de Visser, H. J. C. Berendsen, M. Olivucci, A. E. Mark, and M. A. Robb, J. Am. Chem. Soc. **126**, 4228 (2004).
[40]D. R. Yarkony, J. Chem. Phys. **114**, 2601 (2001).
[41]M. Ben-Nun, F. Molnar, K. Schulten, and T. J. Martinez, Proc. Natl. Acad. Sci. **99**, 1769 (2002).
[42]G. J. Atchity, S. S. Xantheas, and K. Ruedenberg, J. Chem. Phys. **95**, 1862 (1991).
[43]R. R. Coifman, S. Lafon, M. Maggioni, B. Nadler, F. Warner, and S. W. Zucker, Proc. Natl. Acad. Sci. U.S.A. **102**, 7426 (2005).
[44]R. R. Coifman and S. Lafon, Appl. Comput. Harmon. Anal. **21**, 5 (2006).
[45]S. Lafon and A. B. Lee, IEEE Trans. Pattern Anal. Mach. Intell. **28**, 1393 (2006).
[46]D. J. Heisterberg, see http://www.ccl.net/cca/software/SOURCES/FORTRAN/fitest/ (1991).
[47]M. Ben-Nun and T. J. Martínez, Chem. Phys. Lett. **290**, 289 (1998).
[48]H.-J. Werner, P. J. Knowles, R. Lindh, F. R. Manby, M. Schuetz *et al.*, MOLPRO, version 2006.2, a package of *ab initio* programs, 2006, see http://www.molpro.net.
[49]H.-J. Werner and P. J. Knowles, J. Chem. Phys. **82**, 5053 (1985).
[50]B. O. Roos, Adv. Chem. Phys. **69**, 399 (1987).
[51]P. Hariharan and J. Pople, Theor. Chim. Acta **28**, 213 (1973).
[52]R. R. Coifman, I. G. Kevrekidis, S. Lafon, M. Maggioni, and B. Nadler, Multi. Model. Simul. **7**, 842 (2008).
[53]J. D. Coe, B. G. Levine, and T. J. Martínez, J. Phys. Chem. A **111**, 11302 (2007).
[54]P. O. Lowdin, Adv. Quantum Chem. **5**, 185 (1970).
[55]C. Ko, B. Levine, A. Toniolo, L. Manohar, S. Olsen, H.-J. Werner, and T. J. Martinez, J. Am. Chem. Soc. **125**, 12710 (2003).
[56]K. J. Hellingwerf, A. van Leeuwenhoek **81**, 51 (2002).
[57]H. El-Gezawy, W. Rettig, A. Danel, and G. Jonusauskas, J. Phys. Chem. B **109**, 18699 (2005).
[58]A. Espagne, D. H. Paik, C.-B. Pascale, Monique M. Martin, and A. H. Zewail, ChemPhysChem **7**, 1717 (2006).
[59]V. Molina and M. Merchan, Proc. Natl. Acad. Sci. U.S.A. **98**, 4299 (2001).
[60]E. V. Gromov, I. Burghardt, H. Koppel, and L. S. Cederbaum, J. Phys. Chem. A **109**, 4623 (2005).
[61]A. Sergi, M. Gruning, M. Ferrario, and F. Buda, J. Phys. Chem. B **105**, 4386 (2001).
[62]Q.-S. Li and W.-H. Fang, Chem. Phys. **313**, 71 (2005).
[63]E. V. Gromov, I. Burghardt, J. T. Hynes, H. Koppel, and L. S. Cederbaum, J. Photochem. Photobiol., A **190**, 241 (2007).
[64]A. Yamada, S. Yamamoto, T. Yamato, and T. Kakitani, J. Mol. Struct.: (THEOCHEM) **536**, 195 (2001).
[65]R. S. H. Liu, Acc. Chem. Res. **34**, 555 (2001).
[66]J. C. Tully, J. Chem. Phys. **93**, 1061 (1990).
[67]H.-D. Meyer and G. A. Worth, Theor. Chim. Acta **109**, 251 (2003).
[68]K. H. Hughes, C. D. Christ, and I. Burghardt, J. Chem. Phys. **131**, 124108 (2009).