Optimizing transition states via kernel-based machine learning

Zachary D. Pozun,[1,2] Katja Hansen,[1,3,a)] Daniel Sheppard,[1,2,b)] Matthias Rupp,[1,3,c)] Klaus-Robert Müller,[1,3,4] and Graeme Henkelman[1,2,d)]

[1] *Institute for Pure and Applied Mathematics, University of California, Los Angeles, Los Angeles, California 90095-7121, USA*

[2] *Department of Chemistry and Biochemistry and the Institute for Computational Engineering and Sciences, The University of Texas at Austin, Austin, Texas 78712-0165, USA*

[3] *Machine Learning Group, Computer Science Department, Technische Universität Berlin, Germany*

[4] *Department of Brain and Cognitive Engineering, Korea University, Anam-dong, Seongbuk-gu, Seoul 136-713, Korea*

(Received 25 January 2012; accepted 11 April 2012; published online 1 May 2012)

We present a method for optimizing transition state theory dividing surfaces with support vector machines. The resulting dividing surfaces require no *a priori* information or intuition about reaction mechanisms. To generate optimal dividing surfaces, we apply a cycle of machine-learning and refinement of the surface by molecular dynamics sampling. We demonstrate that the machine-learned surfaces contain the relevant low-energy saddle points. The mechanisms of reactions may be extracted from the machine-learned surfaces in order to identify unexpected chemically relevant processes. Furthermore, we show that the machine-learned surfaces significantly increase the transmission coefficient for an adatom exchange involving many coupled degrees of freedom on a (100) surface when compared to a distance-based dividing surface. [http://dx.doi.org/10.1063/1.4707167]

I. INTRODUCTION

One of the great challenges of computational chemistry and materials science is the disparity in time scales between atomic vibrations and the material properties that evolve on a human time scale. Molecular dynamics (MD) cannot reach the time scales required to observe the rare events which govern phenomena such as bond-breaking during catalysis and grain boundary migration.

The most successful framework for bridging the time scale gap is transition state theory (TST).[1–3] Within the TST approximation, the description of rare events is transformed from a problem of kinetics to one of equilibrium statistical mechanics by constructing a hypersurface that separates a reactant state from product states. The rate of reaction can be approximated by the equilibrium flux out of this hypersurface as

$$k_{\mathrm{TST}} = \frac{1}{2}\langle \delta(x - x^*)|\bar{v}|\rangle_{\mathrm{R}}, \tag{1}$$

where $\langle \dots \rangle_{\mathrm{R}}$ is a Boltzmann average over the reactant region, $x = x^*$ is the location of the TS surface, and $\bar{v}$ is the average velocity through the surface. For TST to be a meaningful approximation, the TS surface should capture the bottleneck regions through which reactive trajectories pass.

Given that there are at least as many crossing points as reactive trajectories, the TST rate is always greater than or equal to the true rate. From the relationship

$$k_{\mathrm{True}} = \kappa\, k_{\mathrm{TST}}, \tag{2}$$

the transmission coefficient, $\kappa \in [0, 1]$, quantifies the fraction of successful trajectories which cross the TS surface.

In any complex system, however, finding such a TS is a difficult problem. Even for systems in which the reaction mechanisms are known, an analytic expression of the TS surface can be intractable. In a suboptimal surface, not all crossing points will lead to reactive trajectories, and reactive trajectories may also re-cross the surface.

The challenge of choosing a TS dividing surface is shown in Fig. 1. Consider a case where there is *a priori* knowledge that a product state P_1 is separated from the reactant state R along the x direction. Choosing $x = x^*$ is then a logical choice for a transition state (TS_1). Such a surface, however, is demonstrably suboptimal for separating R and P_1; even if x^* is variationally optimized, regions of the surface lie in either the reactant or product states. An even more serious problem is the presence of the second product state, P_2, which could go undetected when calculating the TST rate through TS_1.

A variational TS can only be as good as the parametrization of the surface. If the system reacts via an unexpected mechanism, the TST approximation will fail. The problem of an assumed reaction coordinate is particularly severe in the case of a condensed phase system where there can be many possible reactions involving collective degrees of freedom.[4–6] Obtaining a dividing surface that contains all of the low-energy processes is a formidable task.

Significant effort has been expended in order to determine the nature of the optimal dividing surface. Although the dividing surface is frequently defined along a reaction

[a)] Also at Theory Department, Fritz Haber Institute of the Max Planck Society, Berlin, Germany.

[b)] Present address: Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA.

[c)] Present address: Institute of Pharmaceutical Sciences, ETH Zurich, Zurich, Switzerland.

[d)] Electronic mail: henkelman@mail.utexas.edu.

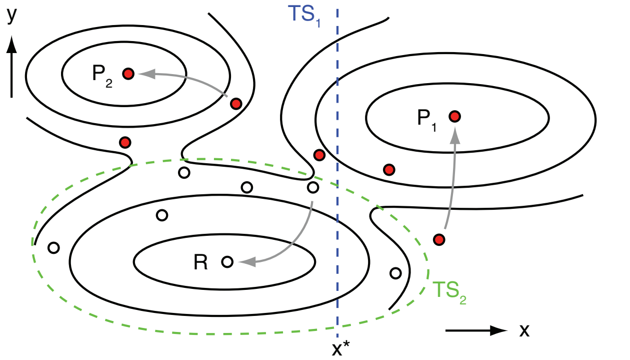

FIG. 1. Transition state TS$_1$, placed along an assumed reaction coordinate *x*, separates reactant R and product P$_1$ but fails to describe the transition to P$_2$. TS$_2$ is a surface which can be determined by training a machine to distinguish a set of points as reactant or product.

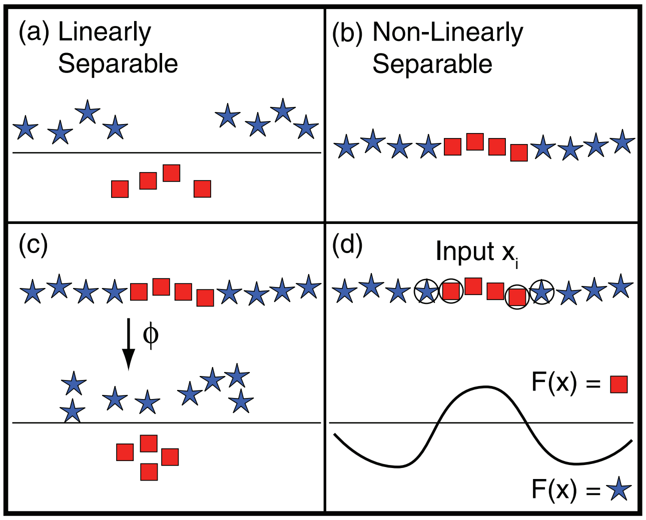

FIG. 2. For data that is linearly separable (a), the SVM method seeks to identify the plane which separates the data with a maximum margin. When the data are not linearly separable (b), a transform to a higher dimensional space where the data are linearly separable (c) is required through the use of a kernel. A decision function which properly classifies all data points (d) is produced through a linear combination of support vectors, which are circled in black. Note that the explicit form of ϕ is not required to produce $F(\mathbf{x})$.

coordinate in configuration space, it is possible to define a dividing surface in the higher dimensional phase space such that trajectories which cross the surface never recross.[7–11] This definition seeks to identify the minimal set of states that all reactive trajectories must encounter and all non-reactive trajectories never encounter.[12,13] This method is formally exact but, in practice, requires perturbative expansions and coordinate transforms in order to apply to higher-dimensional systems.[14]

Ideally, one wishes to avoid having to perform a computationally expensive coordinate transformation and perturbative expansion around some critical point. Support vector machines (SVMs) allow for one to quickly and efficiently map observed data from configuration space into a higher dimensional space in which the data are linearly separable without having to explicitly determine the mapping function. SVMs are a powerful machine learning technique for estimating nonlinear relationships in data. Most often used for classification tasks, SVMs have been applied in engineering for search, image processing, and intrusion detection, as well as in genetics, bioinformatics, neuroscience, physics, and chemistry.[15–17]

Here, we demonstrate the application of an SVM to create a TS surface which divides reactants from products without parametrization or physical intuition. Figure 1 illustrates a set of points that are identified as reactant or product by following the steepest descent paths to the local minimum. The TS surface TS$_2$ represents the hypersurface which divides the points by this classification.

II. COMPUTATIONAL METHODS

A. Support vector machines

The SVM method, as proposed by Cortes and Vapnik,[18] was originally an algorithm for binary classification. It constructs a plane that separates the data classes with a maximum margin between the two classes. As long as the data are linearly separable, as in Fig. 2(a), no coordinate transformation is required in order to generate this plane. If, as in Fig. 2(b), the data are not linearly separable, then an implicit projection via a kernel function into a high-dimensional feature space is required.

For data $\mathbf{x}_i$ and $\mathbf{x}_j$ that are defined on input space I, the kernel function K must satisfy $K(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle_S$, where ϕ is a transform from I to the feature space S. This feature space is a high-dimensional reproducing kernel Hilbert space in which the data are separable.[15,16] In this case, $\langle \cdot , \cdot \rangle_S$ refers to the inner product in S. The so-called "kernel trick" is that one does not need to define an explicit form for ϕ as long as $K(\mathbf{x}_i, \mathbf{x}_j)$ is an inner product in S.

When applied to a chemical system, the input vectors for the SVM are the position coordinates of the atoms. For a given set of n input vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^{3N}$, each $\mathbf{x}_i$ denotes a vector of the position coordinates of the N atoms in the system. With corresponding class labels $y_1, \ldots, y_n \in \{-1, 1\}$ for products or reactants, the SVM classification function is given by

$$F(\mathbf{x}) = \sum_{i=1}^{n} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}), \tag{3}$$

where K denotes the nonlinear kernel function.[19]

The expansion in Eq. (3) extends over all data points. We note, however, that only few α_i are non-zero; these are the coefficients of the support vectors that define the surface. The parameters α_i are computed by solving the underlying dual optimization problem for soft-margin SVMs, wherein the quantity

$$\sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \tag{4}$$

is maximized subject to the constraints

$$\sum_{i=1}^{n} \alpha_i y_i = 0, \quad C \geq \alpha_i \geq 0. \tag{5}$$

In Eq. (5), the regularization parameter C controls the number of support vectors used to define $F(\mathbf{x})$ and, thus, controls the complexity of the surface in order to avoid over-fitting.

This function is demonstrated graphically in Figs. 2(c) and 2(d). The explicit form of ϕ, which is the coordinate transformation to the feature space in which the data are linearly separable, is difficult to obtain. One, however, does not need to know the form of ϕ but instead uses the kernel to determine the inner product in feature space. For the set of data in Fig. 2(d), complete classification of all data may be achieved by generating a surface of the form of Eq. (3) with only four support vectors that define the surface. In this manner, one does not need to explicitly compute the transform into feature space but instead uses the kernel function in order to quickly determine similarity.

We use a Gaussian radial basis function (RBF) kernel given by

$$K(\mathbf{x}_i, \mathbf{x}_j) = e^{-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2}, \tag{6}$$

where the parameter $\gamma > 0$ controls the kernel width and, thus, the smoothness of the underlying nonlinear classifier.[20] The RBF kernel is chosen due to its stationarity and its performance in classification as compared to polynomial, sigmoidal, or linear kernels. The RBF kernel function has the added benefit that the kernel value is guaranteed to fall on [0, 1], which is not always the case for other kernels.

The parameter γ represents the influence a single data point has on its local environment. γ is related to σ, the width of the Gaussian, by $\gamma = \frac{1}{2\sigma^2}$. As γ grows large, the width of the associated Gaussian function shrinks. As $\gamma \rightarrow \infty$, the feature space consists of vectors that are orthogonal to one another, and the kernel matrix approaches the identity matrix. In contrast, smaller γ values indicate that all data points contribute to the classification. In our methodology, γ has units of 1/Å^2 and C is unitless.

Our SVM implementation is based on the scikits.learn python package[21] and libsvm.[22] The parameters C and γ are determined using a grid search and 5-fold cross-validation.[15,23] In this parameter optimization step, the data are first split into five equally sized folds, and each fold serves once as the test fold with the remaining folds as the training data. The underlying optimization problem is solved on the training data, and the resulting SVM is evaluated on the test fold. The combination of C and γ with best performance in classifying the test folds is then chosen as the optimal set of parameters. In this manner, one does not need to know the structure of the potential energy surface, especially in regards to the curvature of the saddle point regions, in order to optimize the parameters of the SVM surface.

B. Generating input vectors

There are many suitable ways to generate $\mathbf{x}_i$ and y_i; our SVM method only requires one to accumulate points in configuration space and assign a label corresponding to either reactant or product. The full class of methods for sampling a potential energy surface in a computationally efficient way with either dynamics or Monte Carlo is too broad to be fully summarized in this work. We note that only a sampling method and a rule for classification is required. We present a simple scheme for generating an optimal dividing surface either with *a priori* information or without. In order to classify points, we have chosen to follow steepest descent paths from that point to the local minimum. This classification is computationally efficient and identifies the basin of attraction for a given reactant state.

The first step is to collect an initial set of points in both reactant and product states in order to generate an initial guess of the surface. If one has *a priori* information about reaction mechanisms, points may be collected from sampling an assumed dividing surface. If no such information is present, high-temperature MD trajectories may be initiated from which points $\mathbf{x}_i$ are collected regularly and minimized in order to determine y_i. We employ the Bussi-Donadio-Parrinello thermostat in order to sample the canonical ensemble.[24]

After generating an initial guess for the dividing surface, we sample the surface using a harmonic potential and MD. The SVM is then retrained on the new data set, and a new surface is generated. The process is iterated with multiple sampling/learning cycles. We note that this process is inherently parallelizable and that multiple independent trajectories result in faster convergence by simultaneously sampling the surface.

By iterating through the re-learning process, the problem of identifying a dividing surface is transformed from one of parametrization to one of sampling, which is a significantly more tractable problem. If the low free energy regions of the surface have been fully sampled, then the surface will contain all of the relevant bottleneck regions.

C. Molecular dynamics sampling

Once a dividing surface of the form from Eq. (3) has been generated, MD sampling is necessary to calculate TST rates through the surface, to refine the surface through increased sampling, or to determine the free energy of the surface. At a given point in configuration space, the value of the decision function and the gradient of the decision function are known analytically. In order to generate a spring force that attracts the MD trajectory to the $F(\mathbf{x}) = 0$ surface, the Cartesian distance to the closest point on this surface is required. This information is not known when only given a local gradient and the value of $F(\mathbf{x})$.

In order to determine a distance in Cartesian space from a hypersurface defined in a high-dimensional Hilbert space, we first expand the decision function $F(\mathbf{x})$ in a Taylor series around the point $\mathbf{x}_0$, which lies on the dividing surface, by

$$F(\mathbf{x}) = \nabla F(\mathbf{x}_0) \cdot \Delta\mathbf{x} + \mathcal{O}(\Delta\mathbf{x}^2). \tag{7}$$

The dividing surface $F(\mathbf{x}_0) = 0$ represents the inflection of $F(\mathbf{x})$ between reactants and products, so we assume that the local curvature vanishes, $\nabla^2 F(\mathbf{x}_0) \approx 0$. The gradients of the SVM are calculated according to Baehrens *et al.*[25] The signed Cartesian distance to the surface is, thus,

$$D(\mathbf{x}) = \frac{F(\mathbf{x})}{\|\nabla F(\mathbf{x})\|}. \tag{8}$$

A spring force to the decision surface from this point in the direction of the unit gradient can then be applied

$$f_{\mathrm{surf}}(\mathbf{x}) = -k\frac{F(\mathbf{x})\nabla F(\mathbf{x})}{\|\nabla F(\mathbf{x})\|^2}, \tag{9}$$

where k is a spring constant for the restraint.

The system is initially placed on the surface and an MD trajectory is initiated with a total force given by

$$f(\mathbf{x}) = -\nabla U(\mathbf{x}) + f_{\mathrm{surf}}(\mathbf{x}). \tag{10}$$

In this manner, sampling is restricted to the region near the decision surface. Using only local information, the new spring force points along the smallest Cartesian distance to the point at which $F(\mathbf{x}) = 0$. In addition, the spring force may be chosen in the conventional units of energy per Cartesian distance.

III. RESULTS AND DISCUSSION

A. Low-dimensional systems

1. The Voter97 potential

We first apply this method to the two-dimensional Voter97 potential,[26] which has the form

$$U(x, y) = \cos(2\pi x)(1 + 4y) + \frac{1}{2}(2\pi y)^2. \tag{11}$$

This potential is periodic in the x direction and harmonic in y. Periodic boundary conditions are appropriately set such that there are three minima and three saddle points. Trajectories may exit from one edge of the simulation box and come back at the opposite edge without a discontinuity in either the potential or the force.

If the center basin is chosen as the reactant, the optimal dividing surface consists of two vertical lines through the saddle points (dashed lines in Figs. 3(b) and 3(d)). The initial high-temperature MD sampling and the resulting SVM surface is shown in Figs. 3(a) and 3(b). For this surface, 100 points were collected via high-temperature MD. Sampling the surface with a force given by Eq. (10) and re-learning a new SVM surface in an iterative fashion results in the converged surface shown in Figs. 3(c) and 3(d). The final surface has 600 total points, the later 500 of which were collected in sets of 15 at a time before re-learning the surface.

From Fig. 3(a), the initial fraction of points which are support vectors is 15%; however, with increased sampling, this fraction of points with a non-zero α in Eq. (3) steadily trends downward. By the time the initial data set has grown to 600 points, the fraction of points which are support vectors is only 2.6%. The absolute number of support vectors between the two surfaces is essentially the same. The difference is that with more sampling, the vectors align more closely along the ridge between basins. Since they are closer together, the value of γ increases to reduce their range. In the limit of complete sampling, the decision function approaches a step function at the ridge surrounding the saddle region. Additional sampling then does not change the structure of the surface.

The structure of the resulting surface can, however, be dependent on the spring constant which is used to sample the potential energy surface. As shown in Fig. 4(a), a weak spring constant will not hold the trajectory to the SVM surface. Although the resulting dividing surface is still reasonably accurate, Fig. 4(b) indicates that a serious problem may occur in this sampling scheme. The norm of the gradient in the basins becomes quite small. If one has chosen a small spring constant in order to use a larger MD time step, the trajectory could relax into this basin and encounter a very large force due to the gradient norm in the denominator of Eq. (9). In contrast, a large spring constant, although requiring a smaller time step, does not allow for the trajectory to reach a region where the gradient is small. The resulting surface, shown in Fig. 4(c), has the support vectors aligned narrowly along the ridge around the saddle points.

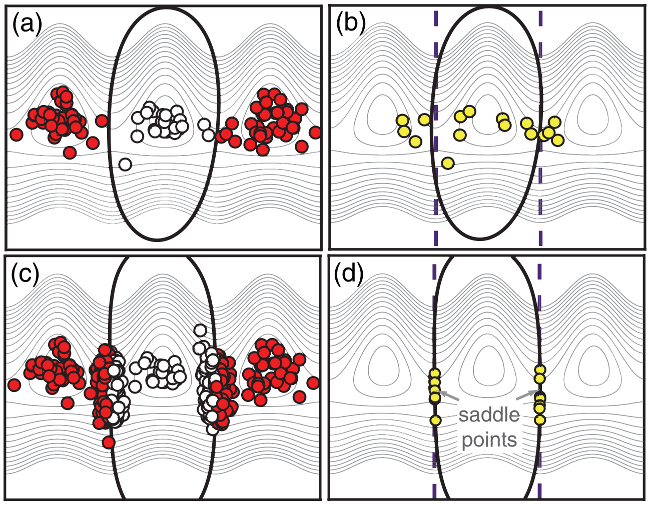

FIG. 3. The process of sampling/re-learning a hypersurface is demonstrated graphically for the Voter97 potential. An initial surface (a), defined by a set of support vectors (b), is generated from high temperature dynamics. The final surface (c) is defined by a small set of support vectors (d) along the reaction bottlenecks.

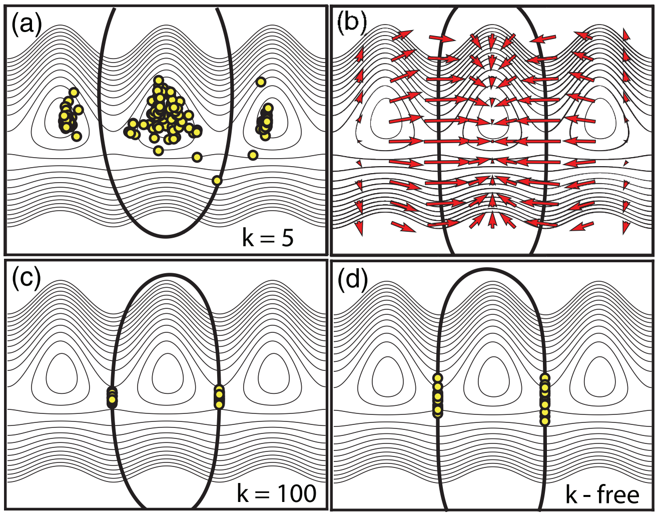

FIG. 4. On the Voter97 potential, a weak spring constant (a) results in a wider spacing of support vectors and can cause sampling issues due to the small norm of the gradient (b) near the basin. In contrast, a large spring constant (c) results in a narrower spacing of the support vectors. A spring-free sampling method (d) produces a similar surface without explicitly defining a spring constant.

TABLE I. Parameters for SVM dividing surfaces of the Voter97 potential.

	C	γ/Å^2	Support vectors
$k = 5$	0.067	0.5	34.6%
$k = 100$	25 000	1.0	5.3%
k-free	67 000	1.0	4.6%

In Fig. 4(d), we demonstrate a spring-free data collection method. Instead of refining the surface with MD sampling, the same high temperature MD trajectories are initiated. Each time the sign on the decision function $F(\mathbf{x})$ changes, a new data point is collected. After the same number of re-learning cycles as in the two spring-sampled surfaces, the SVM surface is nearly indistinguishable from a surface sampled with a spring. Due to the elevated temperature, the distribution of support vectors along the ridge is wider for the spring-free method as compared to the lower-temperature surface with a strong spring. This result demonstrates that the spring-based sampling method detailed in Sec. II C is not strictly necessary to generate a refined dividing surface. Our method simply requires a logical and systematic method for collecting and classifying configuration space points.

The parameters that generate the surfaces in Fig. 4 are summarized in Table I. When the spring constant is too weak to hold the trajectory to the dividing surface, no support vectors can reach the true ridge between states. The surface, then, has a smaller γ, which corresponds to a larger distance between points and a wider Gaussian kernel. When the data points are collected at the ridges, the width of the Gaussian kernel drops and significantly fewer points are required to accurately identify the ridges between states.

2. A mobile adatom on a frozen surface

A more challenging test is to find the mechanism of diffusion for an adsorbed Al atom on an Al(100) surface with an embedded-atom potential.[27] Starting with a frozen (100) surface on which only the adatom can move, as shown in Fig. 5, there are four saddle points between the four surface atoms which define the minimum. Unlike the Voter97 potential, the ridges which contain the saddle points intersect at maxima on top of each of the four frozen atoms. In this case, a single high-temperature trajectory can sample all regions of the dividing surface.

Specifically, a full surface is generated by running high-temperature MD in order to determine an initial surface and then iterated through 50 learning cycles with 10 points collected per cycle. This process is shown in Fig. 5. The single high-temperature trajectory, restricted to the surface by Eq. (10), is able to fully sample all regions of the surface. With sufficient sampling, the support vectors completely enclose the reactant's basin of attraction, as shown in Fig. 5(d). This result demonstrates that, given sufficient sampling, the SVM method can provide complete information about the structure of the potential energy surface.

Although suitable for this low-dimensional case, sampling the potential energy surface with a single high-temperature trajectory scales very poorly with dimensionality. A single high temperature trajectory may take a long time to sample all regions of a high-dimensional surface. More importantly, a high-dimensional system may have many degrees of freedom that are almost entirely uncoupled from the modes that contribute to reactive trajectories. Single-trajectory, high temperature sampling will fail for these cases.

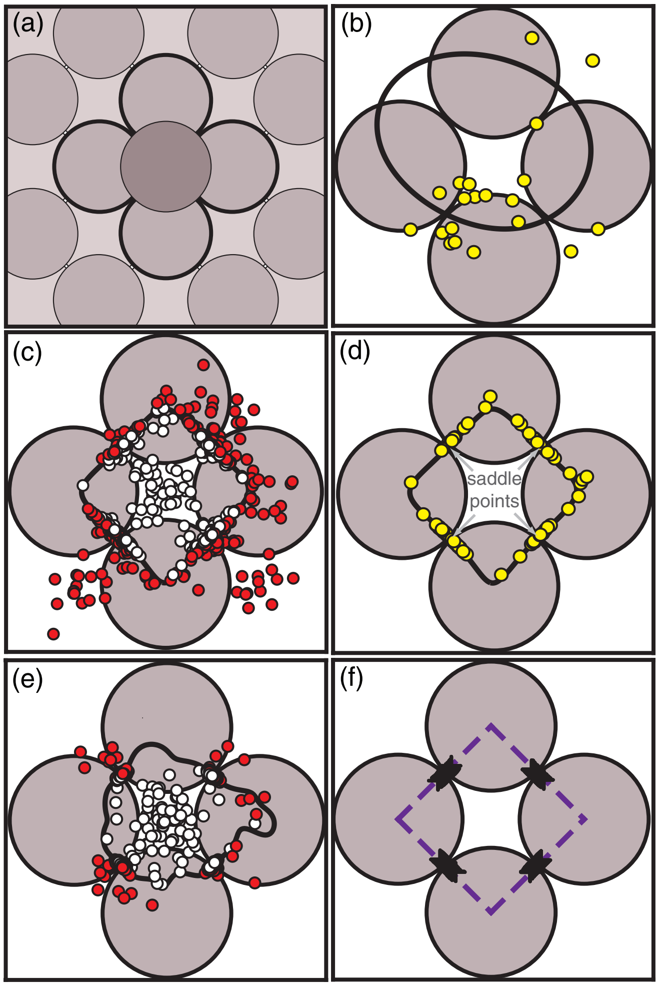

FIG. 5. For an Al adatom on a frozen Al(100) surface (a), an initial high-temperature MD surface (b) may be refined through high temperature sampling (c) to produce a set of support vectors (d) that align surrounding the basin of attraction for the reactant state. Parallel tempering sampling for the initial surface produces a dividing surface (e) that is refined at the saddle points such that the crossing points (f) align along the true dividing ridges (purple lines). All small points indicate locations of the adatom, with respect to the four nearest surface atoms shown.

Ideally, sampling would be restricted to the low-energy bottleneck regions through which all reactive trajectories must pass at the temperature of interest. This goal can be accomplished with parallel tempering, in which many trajectories are initiated at different temperatures. After a given number of MD steps, swaps of configurations are attempted with a probability of acceptance given by

$$p = \min[1, \exp((U(\mathbf{x}_i) - U(\mathbf{x}_j))(\beta_i - \beta_j))], \qquad (12)$$

where $\beta = 1/k_B T$. When parallel tempering is implemented, the resulting dividing surface, shown in Fig. 5(e), is not well defined along the ridges. Importantly, however, parallel tempering sampling of the SVM surface concentrates support vectors in the low free energy bottleneck regions around the saddle points. The optimal choice between local and global accuracy of the SVM surface will depend upon the application. Even in this low-dimensional case, however, parallel tempering generates a dividing surface that is optimized in the important regions and requires fewer support vectors than would be required to enclose the entire basin of attraction.

B. High-dimensional systems

Higher-dimensional systems present a greater challenge for the SVM method. The degrees of freedom grow as $3N$ with N free particles in the system; however, not all of these extra modes contribute information about successful reaction pathways. Due to equipartition, each mode has an average of $\frac{1}{2}k_BT$ in thermal energy. The displacements along these modes steadily grow with temperature but only contributes noise to the SVM decision surface rather than mechanistic information. This extra information steadily increases the $\|\mathbf{x}_i - \mathbf{x}_j\|$ term in Eq. (6). We demonstrate that the SVM classifier can still generate a dividing surface that outperforms several purely geometric dividing surfaces.

1. A mobile adatom on a frozen (100) surface coupled to harmonic oscillators

In order to demonstrate the dimensionality problem, we return to the case of the adatom on a frozen (100) surface. The only information necessary to represent this system as an input vector, $\mathbf{x}$, is the Cartesian positions of the adatom. If this system were coupled to a series of fictitious harmonic oscillators, the reactive pathways would still be defined only by the adatom's position but the full Gaussian kernel would have noise corresponding to the uncoupled modes of the harmonic oscillators.

In Table II, we present this case of a mobile adatom on a frozen (100) surface. The system is coupled to a set of independent harmonic oscillators with a spring constant of 0.25 eV/Å. The set of data points, shown in Fig. 5(e) at 100 K is augmented by a random number drawn from the position distribution of an oscillator at the same temperature. These extra points do not affect the classification of each point as reactant or product, y_i; however, the structure of the optimal dividing surface is affected.

In Table II, we note several trends. The adatom on the frozen surface without any added oscillators has a γ of 5 Å^{-2}, which is a narrow Gaussian kernel surrounding each support vector. The sharp Gaussian width implies that points are close to one another in input space and that the kernel matrix is sparse. With only three oscillators coupled to this system–analogous to a free atom in the bulk away from the surface–the γ parameter shrinks due to the greater distance between points in input space. The extra noise added by the oscillators also creates a steady downward trend in the fraction of points in the data set which are accurately classified during 5-fold cross-validation.

TABLE II. Optimal SVM surfaces for a mobile adatom on a frozen (100) surface coupled to harmonic oscillators.

Oscillators	γ/Å^2	Support vectors	Classification success
0	5.0	32%	93%
3	0.75	32%	90%
15	0.01	37%	85%
50	0.002	48%	74%

As the number of oscillators coupled to the system increases, the parameter γ steadily trends downward and the number of support vectors increases. The SVM machinery requires a wider Gaussian kernel in order to average out the noise dimensions and to classify the points properly. With an increasing number of oscillators, the Gaussian width grows and the kernel matrix becomes dense. The scaling with dimensionality presents a problem for optimizing a dividing surface for a high-dimensional system; however, as we demonstrate in Sec. III B 2, the SVM surface still outperforms purely geometric dividing surfaces.

2. A mobile adatom on a relaxed (100) surface

When the previously frozen surface is allowed to relax, the dimensionality of the dividing surface increases from 3 to 603 for a cell containing four layers that are free to move atop two frozen layers in a 20 × 20 Å cell. Not all of these degrees of freedom are strongly coupled to the diffusion of the adatom, however. Similar to the case of coupled oscillators, the vibrations of a bulk atom are essentially uncoupled from the vibrational modes that point toward a successful reaction. Despite this dimensionality issue, we show that the SVM method can outperform a purely geometric dividing surface without any *a priori* information about the reaction pathways.

An optimal SVM dividing surface maximizes the transmission coefficient κ, which is calculated as in Lu *et al.*[28] κ values for the SVM method are compared in Table III to those for dividing surfaces that are defined by (i) the adatom displacement from its equilibrium position in the reactants and (ii) the maximum fractional displacement of the adatom to neighboring atoms, as in the bond-boost method.[29] In the case of the frozen Al(100) surface, the escape pathways are well

TABLE III. κ values for different dividing surfaces.

	Frozen surface	Free surface
Al(100) surface at 100 K		
Spherical dividing surface	0.95	0.42
Bond-boost dividing surface	>0.95	0.35
SVM dividing surface	0.97	0.62
Al(100) surface at 400 K		
Spherical dividing surface	0.92	0.22
Bond-boost dividing surface	0.95	0.11
SVM dividing surface	1.00	0.31

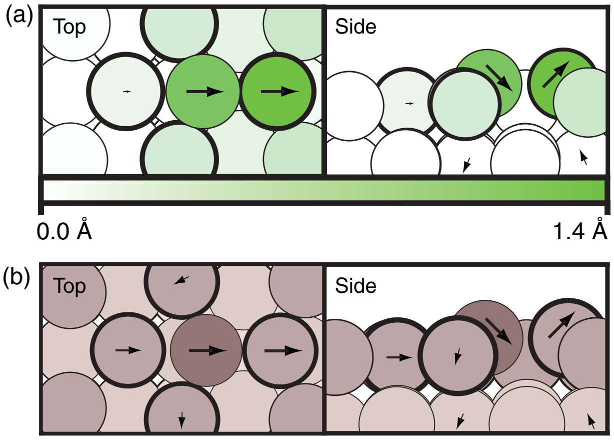

FIG. 6. For an adatom on a relaxed (100) surface, the saddle point for the exchange mechanism (a) displaces nearby surface atoms. The components of the negative mode are shown with arrows. The SVM crossing points cluster at each escape bottleneck, and the centroids of these clusters (b) correctly identify the exchange saddle with a local gradient that points along the negative mode.

defined by only the adatom displacement, so all three surfaces easily divide reactants and products.

In contrast, when the atoms of the Al(100) surface are allowed to relax in addition to the adatom, the dimensionality of the system increases tremendously. As shown in Fig. 6(a), the negative mode at the exchange saddle point, which is the dominant escape pathway at low temperature, involves many coupled degrees of freedom. An accurate TS surface must identify this coupled motion.

The SVM correctly identifies both the configuration of the saddle points as well as the direction of the negative mode at each saddle. The increase in κ noted in Table III is the result of the mechanistic information contained in the SVM surface. An average crossing point through the SVM surface, shown in Fig. 6(b), has a structure which closely approximates the true saddle point for the exchange mechanism. At this point, the gradient of the SVM surface points substantially along the unstable mode of the true saddle point; the inner product between the two vectors is 0.6.

At low temperatures, a plane defined by the saddle point and the eigenvector of the unstable mode produces a near unity κ. Similarly, the mechanistic information–the geometry and the gradient–contained in the SVM surface produces the large value of κ seen in Table III. For comparison, the spherical dividing surface contains the saddle points but provides no information about the unstable mode eigenvector. At high temperatures, however, the plane is a poor approximation for the TS due to anharmonicity away from the saddle points. The SVM and spherical surfaces are better able to follow the TS as it curves between escape mechanisms.

The effect of sampling and the number of vectors input into the SVM machinery are summarized in Table IV. With increasing vectors in the data set, the value of κ steadily grows as the saddle point regions are increasingly well characterized. Unfortunately, the value for γ does not fall correspondingly with increasing vectors. This behavior indicates that each new point is essentially uncorrelated from each other point and that the machinery is attempting to tile a high-dimensional space. The steady increase in κ flattens, however, as the data set grows very large. At this point, the surface is essentially converged and the nonunity of κ indicates that an effect other than undersampling is the cause.

TABLE IV. The effect of sampling size on κ at 100 K.

Input points	C	γ/Å^2	Support vectors	κ
250	2.0	0.05	88%	0.01
500	2.0	0.1	93%	0.05
1000	5.0	0.1	78%	0.07
1500	5.0	0.1	74%	0.17
2000	10.0	0.1	72%	0.41
3000	10.0	0.1	67%	0.59
5600	5.0	0.1	58%	0.62

The trend in κ with increased sampling can be understood in terms of the adatom coupled to fictitious harmonic oscillators. The SVM surfaces whose properties are described in Tables III and IV represent the most general case of our methodology. Every degree of freedom in the system is included, but, as shown in Fig. 6, the only important displacements are local to the adatom. With some intuition about the system, one realizes that only the motion of the atoms near to the adatom need to be included in the SVM machinery.

By only considering the positions of the atoms within a 5.0 Å radius of the adatom, the dimensionality of the free surface is reduced to 54 degrees of freedom from the full 603. As summarized in Table V, when the input into the SVM is localized to the positions of the atoms around the diffusing adatom, the structure of the surface is altered. Fewer points (from the same overall collection) are required to define the SVM classifier. At 100 K, the value of κ significantly increases due to the lack of noise from the vibrating bulk atoms. At 400 K, which is close to the melting point of the surface, the vibrational modes from the bulk are more strongly coupled to the reactive modes and the value of κ is negatively affected due to the reduction in information.

While our aim is to find TS surfaces which maximize κ, it is important to note that the TST rate can be corrected to provide the true rate by evaluating κ. The cost of this evaluation is that of $N \approx 1/\kappa$ short trajectories initiated from the TS surface. If, for example, several hundreds of trajectories can be calculated, a value of $\kappa = 0.01$ can be evaluated and this "poor" TS surface is sufficient for evaluating an accurate rate.

TABLE V. The effect of reducing dimensionality.

	C	γ/Å^2	Support vectors	κ
Free Al(100) surface at 100 K				
Reduced SVM	50.0	0.5	39%	0.82
Full SVM	10.0	0.1	67%	0.59
Free Al(100) surface at 400 K				
	C	γ/Å^2	Support vectors	κ
Reduced SVM	2.0	0.2	39%	0.29
Full SVM	10.0	0.1	89%	0.31

IV. CONCLUSIONS

We have demonstrated a novel method for optimizing TS dividing surfaces using SVMs. Our method is capable of achieving high transmission coefficients for systems containing many degrees of freedom without parametrization. The success of the SVM method is a result of the information density of the surface. In contrast to a dividing surface that requires intuition to be constructed, the SVM surface provides intuition about the structure of the reaction mechanisms. For condensed phase systems in which reactive pathways may be hard to predict, the SVM surface identifies not only the low-energy saddle points but also the mechanisms that result in a successful reaction.

ACKNOWLEDGMENTS

All authors acknowledge support through the Institute of Pure and Applied Mathematics at UCLA. Z.D.P., D.S., and G.H. acknowledge support from the National Science Foundation (Grant No. CHE-0645497). Z.D.P. also acknowledges support from the Dorothy B. Banks Fellowship. K.H., M.R., and K.R.M. acknowledge support by the FP7-ICT programme of the European Community (PASCAL2) and Deutsche Forschungsgemeinschaft (DFG) (Grant No. MU 987/4-2). Computing resources were provided by the Texas Advanced Computing Center. This work was also supported by the World Class University Program through the National Research Foundation of Korea funded by the Ministry of Education, Science, and Technology, under Grant R31-10008.

[1] H. Eyring, J. Chem. Phys. **3**, 107 (1935).
[2] E. Wigner, Trans. Faraday Soc. **34**, 29 (1938).
[3] J. C. Keck, Adv. Chem. Phys. **13**, 85 (1967).
[4] G. Henkelman and H. Jónsson, Phys. Rev. Lett. **90**, 116101 (2003).
[5] L. Xu, G. Henkelman, C. T. Campbell, and H. Jónsson, Phys. Rev. Lett. **95**, 146103 (2005).
[6] G. Henkelman and H. Jónsson, J. Chem. Phys. **115**, 9657 (2001).
[7] E. P. Wigner, J. Chem. Phys. **5**, 720 (1937).
[8] T. Komatsuzaki and R. S. Berry, J. Mol. Struct.: THEOCHEM **506**, 55 (2000).
[9] S. Wiggins, L. Wiesenfeld, C. Jaffé, and T. Uzer, Phys. Rev. Lett. **86**, 5478 (2001).
[10] T. Uzer, C. Jaffé, J. Palacian, P. Yanguas, and S. Wiggins, Nonlinearity **15**, 957 (2002).
[11] H. Waalkens and S. Wiggins, J. Phys. A **37**, L435 (2004).
[12] C. Jaffé, D. Farrelly, and T. Uzer, Phys. Rev. A **60**, 3833 (1999).
[13] R. Hernandez, T. Uzer, and T. Bartsch, Chem. Phys. **370**, 270 (2010).
[14] H. Waalkens, A. Burbanks, and S. Wiggins, Phys. Rev. Lett. **95**, 084301 (2005).
[15] K.-R. Müller, S. Mika, G. Rätsch, K. Tsuda, and B. Schölkopf, IEEE Trans. Neural Netw. **12**, 181 (2001).
[16] B. Schölkopf and A. J. Smola, *Learning with Kernels* (MIT, Cambridge, 2002).
[17] O. Ivanciuc, “Applications of support vector machines in chemistry,” in *Reviews in Computational Chemistry* (Wiley-VCH, 2007), Chap. 6, pp. 291–400.
[18] C. Cortes and V. Vapnik, Mach. Learn. **20**, 273 (1995).
[19] B. Schölkopf, A. Smola, and K.-R. Müller, Neural Comput. **10**, 1299 (1998).
[20] A. Smola, B. Schölkopf, and K.-R. Müller, Neural Networks **11**, 637 (1998).
[21] See http://scikit-learn.sourceforge.net for scikits.learn python module, 2007.
[22] C.-C. Chang and C.-J. Lin, LIBSVM: A library for support vector machines (2001), see http://www.csie.ntu.edu.tw/~cjlin/libsvm.
[23] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, Springer Series in Statistics (Springer, New York, 2001).
[24] G. Bussi, D. Donadio, and M. Parrinello, J. Chem. Phys. **126**, 014101 (2007).
[25] D. Baehrens, T. Schroeter, S. Harmeling, M. Kawanabe, K. Hansen, and K.-R. Müller, J. Mach. Learn. Res. **11**, 1803 (2010).
[26] A. F. Voter, J. Chem. Phys. **106**, 4665 (1997).
[27] A. F. Voter and S. P. Chen, Mater. Res. Soc. Symp. Proc. **82**, 175 (1987).
[28] C.-Y. Lu, D. E. Makarov, and G. Henkelman, J. Chem. Phys. **133**, 201101 (2010).
[29] R. A. Miron and K. A. Fichthorn, J. Chem. Phys. **119**, 6210 (2003).