Interpolating Moving Least-Squares Methods for Fitting Potential Energy Surfaces: Applications to Classical Dynamics Calculations

Yin Guo, Akio Kawano, and Donald L. Thompson

Department of Chemistry, Oklahoma State University, Stillwater, OK 74078

Albert F. Wagner

Chemistry Division, Argonne National Laboratory, Argonne, IL 60439

Michael Minkoff

Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL 60439

Abstract

As a continuation of our efforts to develop efficient and accurate interpolating moving least-squares (IMLS) methods for generating potential energy surfaces, for the first time we carry out classical trajectories and compute kinetics properties on higher-degree IMLS surfaces. In this study, we have investigated the choice of coordinate system, the range of points (i.e., the cut-off radius) used in fitting, strategies for selections of data points, and basis elements. We illustrate and test the method by applying it to hydrogen peroxide (HOOH). In particular, reaction rates for the O-O bond breaking in HOOH are calculated on fitted surfaces by /Ousing the classical trajectory approach to test the accuracy of the IMLS method for providing potentials for dynamics calculations.

I. INTRODUCTION

Constructing a potential energy surface (PES) is the first step in many theoretical studies of chemical applications. While many PES are obtained by empirical fitting, significant advances in quantum chemistry and computer power in the past two decades have made it increasingly possible to construct a PES from accurate *ab initio* energy values and sometimes derivatives of the energy. Over the past several decades, a variety of data fitting methods have been explored[1] for constructing a reliable PES from a minimum of expensive *ab initio* electronic structure calculations. These methods include the use of cubic splines,[2] least-squares fitting,[3] reproducing kernel Hilbert space interpolation methods,[4] and hybrid methods.[5,6] These methods often lack robustness (e.g., they may require special techniques for each PES) and scalability (e.g., they involve computations that increase rapidly with the reaction complexity). One of the most successful methods developed in the past decade that is both robust and scalable is the method of Ischtwan and Collins. This method is based on a modified Shepard[8] interpolation—the modification is necessary to overcome poor derivative properties near the *ab initio* points. The modification requires a Taylor expansion through second order at each *ab initio* point, implying that *ab initio* gradients and Hessians must be calculated. This approach is mathematically coherent and has many advantages, but the gradients and Hessians often cannot be provided inexpensively (particularly for high-level *ab initio* calculations).

Recently we and others have sought to develop an alternative approach based on the use of interpolating moving least-squares (IMLS) method.[9-14] The approach is straightforward to implement and has the potential to provide a highly automated procedure for generating an accurate global PES for practical applications. IMLS methods can be of any polynomial degree and thus

have generality. In fact, the Shepard method is a zero-degree IMLS technique. These methods do not require any derivative information and are thus efficient for high-level *ab initio* techniques.

In prior studies[9,10] we have examined 1-D and 3-D cases. However, further development and improvement are needed to take the method to the application stage. The main focus of the present study is to further investigate various aspects of the approach and to develop and test procedures for generating IMLS surfaces for classical dynamics calculations. The most important issue here is the cost of computational time. The IMLS values for energy and derivatives at a given point in configuration space are obtained by constructing and solving a matrix equation. The sizes of the matrices involved in the equation depend on the numbers of data points and basis elements, both of which increase as the size of the system increases. Since derivatives are evaluated at every time step, meaning the matrix equation needs to be solved millions of times in a trajectory study, the IMLS method can become prohibitively expensive as the size of the system increases. Thus, some measures have to be taken to reduce the computational cost. We present in this paper the procedures we have used and developed in this respect, using hydrogen peroxide (HOOH) as a test case.

In Sec. II, we describe procedures for fitting surfaces using the IMLS approach. We first give a brief description of the basic formulism of the IMLS method. We then discuss the coordinate systems, the cut-off radius, selection of data points, and selection of basis elements, with calculations performed on HOOH. Since our purpose is to develop and test the method, instead of performing *ab initio* calculations, we have used the six-dimensional (6-D) analytical PCPSDE potential developed by Kuhn et al.[15] to generate data points and assess accuracy. In Sec. III, we present rate constants for the O-O bond fission in HOOH obtained by running classical trajectories on fitted surfaces. A summary and the conclusions are given in Sec. IV.

II. AN IMLS APPROACH FOR GENERATING POTENTIAL ENERGY SURFACES

A. IMLS Method

A brief description of the IMLS method is given here for the convenience of discussing the issues related to the method and the procedures employed to further refine the method.

In the IMLS method, the fitted potential is expressed in terms of a linear combination of the basis functions $\{ b_i(\mathbf{Z}) \}$:

$$V_{\text{fitted}}(\mathbf{Z}) = \sum_{i=1}^{M} a_i(\mathbf{Z}) b_i(\mathbf{Z}), \tag{1}$$

where M is the total number of basis functions, $\{b_i(\boldsymbol{Z})\}$ are polynomials, and $\{a_i(\boldsymbol{Z})\}$ are determined by minimizing the sum of weighted squared-deviations,

$$D[V_{\text{fitted}}(\mathbf{Z})] = \sum_{i=1}^{N} w_i(\mathbf{Z}) [V_{\text{fitted}}(\mathbf{Z}^{(i)}) - V(\mathbf{Z}^{(i)})]^2 \ , \tag{2}$$

that is, by the conditions $\partial D/\partial a_i = 0$. Here N is the total number of *ab initio* data points located at $\{\mathbf{Z}^{(1)},\mathbf{Z}^{(2)},\ldots,\mathbf{Z}^{(\mathrm{N})}\}$ with energies $\{V(\mathbf{Z}^{(1)}),V(\mathbf{Z}^{(2)}),\ldots,V(\mathbf{Z}^{(\mathrm{N})})\}$, respectively, and $w(\mathbf{Z})$ is a weight function with the property that the data points closer to $\mathbf{Z}$ have larger weights than the more distant ones. This minimization procedure produces the following equation:

$$\mathbf{B}^{\mathrm{T}}\mathbf{W}\mathbf{B}\mathbf{a} = \mathbf{B}^{\mathrm{T}}\mathbf{W}\mathbf{V}, \tag{3}$$

where $\mathbf{B}$ is a $N \times M$ matrix,

$$\mathbf{B} = \begin{pmatrix} b_1(\mathbf{Z}^{(1)}) & b_2(\mathbf{Z}^{(1)}) & \cdots & b_M(\mathbf{Z}^{(1)}) \\ b_1(\mathbf{Z}^{(2)}) & b_2(\mathbf{Z}^{(2)}) & \cdots & b_M(\mathbf{Z}^{(2)}) \\ \vdots & \vdots & \ddots & \vdots \\ b_1(\mathbf{Z}^{(N)}) & b_2(\mathbf{Z}^{(N)}) & \cdots & b_M(\mathbf{Z}^{(N)}) \end{pmatrix}, \tag{4}$$

$\mathbf{W}$ is a $N \times N$ diagonal matrix,

$$\mathrm{W} = \mathrm{diag}(w_1(\mathrm{Z}), w_2(\mathrm{Z}), \ldots, w_N(\mathrm{Z})), \tag{5}$$

and $\boldsymbol{V}$ is a column vector,

$$\boldsymbol{V} = (V(\boldsymbol{Z}^{(1)}), V(\boldsymbol{Z}^{(2)}), \ldots, V(\boldsymbol{Z}^{(\mathrm{N})}))^{\mathrm{T}}. \tag{6}$$

Once the coefficient vector **a** is obtained by solving Eq. (3), the fitted energy at a given point $\boldsymbol{Z}$ is computed by using Eq. (1). Equation (3) was solved by using the singular value decomposition (SVD) technique.

In classical trajectory calculations, the first derivatives of the potential are evaluated. The efficient procedure[10] that reuses the matrix operations for solving **a** can be employed for computing the derivatives **a'**. Briefly, taking the derivative of Eq. (3) leads to a similar equation:

$$\mathbf{B}^{\mathrm{T}}\mathbf{W}\mathbf{B}\mathbf{a}' = (\mathbf{B}^{\mathrm{T}}\mathbf{W}\mathbf{V})' - (\mathbf{B}^{\mathrm{T}}\mathbf{W}\mathbf{B})'\mathbf{a}. \tag{7}$$

Thus, the matrix decomposition of $\mathbf{B}^{\mathrm{T}}\mathbf{W}\mathbf{B}$ by the SVD technique for solving Eq. (3) can be reused to solve Eq. (7).

Since **a** is a function of $\boldsymbol{Z}$, it needs to be evaluated at every time step in a trajectory propagation. This process amounts to solving Eq. (3) at each step. The time to solution goes as NM^2. Clearly, the IMLS approach can become prohibitively expensive as the size of the molecule increases, and thus measures must be taken to minimize N and M. The procedures employed in this

regard are discussed below in this section after a description of the coordinate systems used in this study.

B. Coordinate System

We have investigated two coordinate systems: one is interatomic distance coordinates that are commonly used in the modified Shepard interpolation method[16] and that we used in our previous studies;[11] the other is valence internal coordinates, which are physically more intuitive for many systems.

In the system of interatomic distances $\{R_i\}$, as in our previous work on HOOH,[11] the reciprocal interatomic distances $\{Z_i=1/R_i\}$ are used as the fitting coordinates and the un-normalized weight function is

$$v_i(\mathrm{Z}) = v(\Delta\rho) = \frac{1}{[(\Delta\rho)^2 + \varepsilon^2]^p}, \tag{8}$$

where $(\Delta\rho)^2 = \sum_{k=1}^{6}(\rho_k - \rho_k^{(i)})^2$ and $\{\rho_k = R_k\}$.

The valence internal coordinates are bond lengths, bond angles, and dihedral angles. For the 6-D case (HOOH) considered here, the coordinates $\boldsymbol{Z} = (r_{\mathrm{OH}}, r_{\mathrm{OO}}, r_{\mathrm{OH}}, \theta_1, \theta_2, \cos\tau)$ (with θ_1 and θ_2 being the two bending angles) are used as the fitting coordinates. Here, $\cos\tau$ instead of τ is used as the torsional coordinate because τ is defined in [-π,π] and is thus discontinuous at π. The un-normalized weight function is of the same form as that of Eq. (8) except that $\rho=(r_{\mathrm{OH}}, r_{\mathrm{OO}}, r_{\mathrm{OH}}, a_1\theta_1, a_2\theta_2, a_3\cos\tau)$, where $\{a_i\}$ are constants of unity value with units such that $\{\rho_i\}$ all have the units of length.

To determine the values of the weight-function parameters, we computed rms errors for energy and gradient as functions of ε and p. The results obtained with the valence internal coordinates are shown in Fig. 1, where the total number of symmetry distinct data points used for fitting is 502, and 5000 points generated by using the efficient microcanonical sampling (EMS) method[17] for E = 77.625 kcal mol^{-1} on the fitted surface were used for computing the errors. The selections of the data points and basis elements are described in the following subsections. The cut-off radius (described in the next subsection) was 1.5 Å, which is the value used in computing the rate constants presented in Sec. III. The value p=4 was used in computing the results shown in Figs. 1(a) and 1(b).

As shown in Figs. 1(a) and 1(b), the rms errors are nearly independent of ε once it is below a certain value. This is understandable considering that a small enough ε should have no effect on the weight function except near the origin when a data point is very close to the fitted point; such a situation would happen infrequently, especially in high-dimensional space. Similar results were obtained by using the interatomic distance coordinates. This result suggests that one need not test the value for ε in future studies as long as a reasonably small value is used. Similar conclusions were reached in our earlier studies.[10,11]

Figures 1(c) and 1(d) illustrate the rms errors for energy and gradient as a function of p. Here ε=0.01, which is the value used in all the subsequent calculations with either coordinate system. One can see that the rms error for energy remains near its minimum value for p equals 4 to 6. However, the rms error for the gradient behaves somewhat differently from that for energy; it remains near its minimum value at p equals 1 to 4. We note that these results are only qualitative: they vary somewhat with the number of data points and their locations. Similar behaviors were also

observed for interatomic distance coordinates. Based on these results, we have used p=4 in all the calculations presented here. Similar conclusions were reached in our earlier studies.[10,11]

C. Cut-Off Radius

As discussed in Sec. II A, the time for solution goes as NM^2. Obviously, one way to reduce the computational cost is to reduce the number of data points N included in an evaluation. This is possible because of the local nature of the IMLS fits—the weight function acts to make the contributions from distant data points negligible. Therefore, a cut-off radius r_{cut} can be introduced to exclude those distant points. That is, the weight function is set to zero for $\Delta\rho \geq r_{\text{cut}}$.

Truncating the weight function at $\Delta\rho = r_{\text{cut}}$ introduces a discontinuity in the function that causes jumps in the fitted energy and derivatives whenever a data point crosses r_{cut}. This problem can be circumvented by introducing a damping function $S(\Delta\rho)$ to make the weight function go smoothly to zero at r_{cut}. That is, we write the new weight function as

$$w^{new}(\Delta\rho) = w(\Delta\rho)\, S(\Delta\rho), \tag{9}$$

where $(\Delta\rho)^2 = \sum_{k=1}^{6} (\rho_k - \rho_k^{(i)})^2$ as defined earlier and $S(\Delta\rho)$ is chosen as

$$S(\Delta\rho) = \begin{cases} \tanh^3\left[\alpha\left(1 - \dfrac{(\Delta\rho)^2}{r_{\text{cut}}^2}\right)\right] & \text{if } \Delta\rho < r_{\text{cut}}, \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

Here α is used as a parameter to adjust the fall-off region in $S(\Delta\rho)$; the value α=7 was used. The power 3 is used so that the function and its first and second derivatives are all continuous at $\Delta\rho = r_{\text{cut}}$.

Figure 2 illustrates the effect of using a cut-off radius with the valence internal coordinates. Same as in Fig. 1, the number of symmetry distinct data points N used for fitting is 502, and 5000

points generated via EMS method on the fitted surface were used for calculating the rms errors. Panels (a) and (b) show the dependence of the rms errors for energy and gradient on r_{cut}, respectively. It is seen that as r_{cut} decreases, the rms errors also decrease until reaching a minimum before turning back up again. The rms errors decrease with decreasing r_{cut} at large r_{cut} because the distant data points actually have negative effects on the fits. They increase with decreasing r_{cut} at small r_{cut} because the fits become less accurate as the number of data points included gets too small. Since the derivatives are used in a trajectory calculation and since the rms error for gradient is minimal at approximately $r_{\text{cut}} = 1.5$ Å as shown in Fig. 2(b), $r_{\text{cut}} = 1.5$ Å was used in all the subsequent calculations in the valence internal coordinates for $N_{\text{SYM}} = 502$ (among which there are four points with C_2 symmetry and thus $N = 1000$). Figure 2(c) illustrates the averaged number of data points enclosed inside the cut-off radius as a function of r_{cut}. For the value of r_{cut} used here, r_{cut}=1.5 Å, the averaged number of data points inside r_{cut} is approximately 300 out of 1000 total points. Thus, using a cutoff provides nearly 70% of savings in computer time in this case. For the calculations performed in this study, it is generally true that the rms errors for gradients are near their minima if a value of r_{cut} is used such that about 250—300 points on average are inside r_{cut}. Thus, the approximate value of r_{cut} can be readily estimated by computing the averaged number of data points inside r_{cut}.

A neighbor list can be implemented to further reduce the computational time on evaluating distances that are greater than r_{cut}. In this bookkeeping scheme a list of neighboring data points that lie within a distance slightly larger than r_{cut} is maintained. The same list is used over several consecutive time steps and updated periodically. However, the procedure does not provide significant savings through our tests in studying the O-O bond fission in HOOH, likely because the cost of evaluating distances is minor compared to that of solving the matrix equation.

D. Selection of Data Points

How to locate the data points is another important aspect of the IMLS approach. In principle, the data points should be placed in regions of interest. For studying chemical reactions using the classical trajectory approach, the data points should be located in regions where trajectories travel, and they should be somewhat uniformly distributed to avoid sparsity in any regions that trajectories frequently visit. The data points used in this study consist of two sets. The first set includes 25 symmetry distinct predetermined points: 5 points along the O-O minimum energy path (among which 4 points with C_2 symmetry, including the equilibrium configuration), 3 points at large r_{OH}, and 17 points in the low-energy regions (E<40 kcal mol^{-1}). These 25 predetermined points were used in all the calculations with either of the two coordinate systems. The second set consists of randomly selected points under two criteria: (1) the distance from the existing points is no less than a specified value, and (2) the energy is less than a maximum value E_{max}. In this study, E_{max} was chosen as 85 kcal mol^{-1}, the largest energy for which the rate constant was calculated. The total number of points $N = 2N_{SYM} - 4$. We note that although this random selection scheme is applicable here, a more efficient random walk procedure needs to be developed for general applications.

The reason that a random selection procedure, instead of a trajectory-based approach, is employed here lies mainly in the computational efficiency. We have calculated rate constants for the O-O bond fission in HOOH at various energies. Running trajectories repeatedly using the IMLS surface to pick the data points is computationally demanding and likely less efficient. With the random selection procedure employed here, the data points have the probability to be in any regions with E<E_{max}, which works well when the intramolecular dynamics is chaotic and trajectories cover

all regions of the energetically allowed phase space. However, if trajectories cover only limited regions, such as when the reaction is highly mode-selective, a trajectory-based approach is probably more efficient. We will investigate further the selection of data points in future work.

E. Selection of Basis Elements

As discussed in Sec. II A, the number of basis elements used in fitting is another important factor in the computational cost. To achieve a given level of accuracy, one needs to carefully select the basis elements to make the set as small as possible. As the dimension of the system n increases, the number of diagonal terms scales as n, and the number of off-diagonal (cross) terms scales combinatorially. Thus, reducing the number of cross terms becomes increasingly important as the size of the system increases.

Since a fitted energy is an expansion over basis elements (see Eq. (1)), the contribution of each basic element b_i to the fit is determined by the corresponding coefficient a_i. Therefore, one can examine $\{a_i\}$ to determine the basis elements. In this study, the averaged values of $\{a_i\}$ were calculated for a set of fitted energies at randomly selected configurations, and the basis elements with large values of $< a_i >$ were selected. For the valence internal coordinate system, there are 25 terms in the basis set; they are 19 diagonal terms with the bond terms up to the fourth order and the angle terms up to the second order, and 6 cross terms concerning couplings between the O-O bond and the two bending angles: $r_{\mathrm{OO}}\theta_1$, $r_{\mathrm{OO}}\theta_2$, $r_{\mathrm{OO}}^2\theta_1$, $r_{\mathrm{OO}}^2\theta_2$, $r_{\mathrm{OO}}{\theta_1}^2$, and $r_{\mathrm{OO}}{\theta_2}^2$. The selection was also confirmed by computing the rms errors for different selections. For the interatomic distance coordinate system, 25 terms were also used; they are 3 third-order terms, $1/r_{\mathrm{O_1H_1}}^3$, $1/r_{\mathrm{O_2H_2}}^3$, and $1/r_{\mathrm{OO}}^3$, and all the terms up to the second order except 6 second-order cross terms: $1/(r_{\mathrm{HH}}r_{\mathrm{O_1H_1}})$, $1/(r_{\mathrm{HH}}r_{\mathrm{O_2H_2}})$, $1/(r_{\mathrm{O_1H_1}}r_{\mathrm{O_1H_2}})$, $1/(r_{\mathrm{O_2H_2}}r_{\mathrm{O_2H_1}})$, $1/(r_{\mathrm{O_1H_2}}r_{\mathrm{O_2H_1}})$, and $1/(r_{\mathrm{O_1H_1}}r_{\mathrm{O_2H_2}})$. With the valence internal

coordinates, only 2 out of 15 second-order cross terms are important, whereas 9 out of 15 are important with the interatomic distance coordinates. Since the valence internal coordinate system is physically more intuitive and better for describing potential interactions in general, generally it may be more suitable for describing potential couplings; that is, fewer cross terms are needed.

III. RATE CONSTANTS FOR O-O BOND FISSION IN HOOH

To test the accuracy of the IMLS method described in Sec. II for computing rate constants, we have studied O-O bond fission in HOOH using the classical trajectory approach.

A. Details of Trajectory Calculations

The overtone-induced dissociation of HOOH into OH fragments has been extensively studied both experimentally and theoretically.[18] In a recent study using the same analytical surface, we have shown that the O-O bond dissociation rates are similar for the local mode sampling and EMS.[19] Thus, to compute the reaction rate for OH-overtone excitations, one can use the EMS to select the initial conditions at energies corresponding to the experimental values.

The rate constant for the O-O bond fission in HOOH was calculated by using the fitted surface for several energies. The initial conditions were selected by using the EMS method on the fitted potential. Ensembles of 500 trajectories were used for each set of initial conditions. The trajectories were propagated by using Hamilton's equations of motion in lab-fixed Cartesian coordinates using a fourth-order Runge-Kutta-Gill integrator with a fixed step size of 0.07 fs, and were followed for 1 to 3 ps depending on the energy. The total angular momentum is zero in all cases.

Some calculations were also performed on the analytical surface to provide the "exact" rates for comparisons. Since the calculations are much faster in this case, the rate was computed for three independent ensembles of 2,000 trajectories, and the average value is reported.

B. Rate Constants

The reaction was assumed to be first order, so the rate constant k was obtained by fitting the decay curve to

$$\ln(N_t / N_0) = -kt,$$

where N_t is the number of unreacted trajectories at time t and N_0 is the total number of trajectories in the ensemble. A trajectory is assumed to have reacted once the O-O bond length exceeds 3.2 Å.

We show in Table I the dependence of the rate constant on the number of data points. The rate calculated on the analytical surface was also listed for comparison. Here $E = E(v_{\mathrm{OH}} = 7) = 77.625$ kcal mol^{-1}, N is the total number of data points, and N_{SYM} is the number of symmetry distinct points. The results for using either the valence internal or interatomic distance coordinates are listed. Note that statistical errors due to the sampling size in the rates computed on the fitted surfaces are probably around 10% with 500 trajectories in an ensemble. In the valence internal coordinates, the calculated rate agrees well with the exact rate even when N_{SYM} is only 152. On the other hand, the rate is much faster in the interatomic distance coordinates and does not converge to the exact rate even when N_{SYM} is as high as 2002. Nevertheless, the result is much improved with the use of the zeroth-order potential, as shown in parentheses in the last column. The same zeroth-order potential given in Ref. 11 was used.

We have also tested using a smaller number of points ($N_{\mathrm{SYM}} = 102$). With this small number of points, however, a trajectory often goes into regions with so sparse points that the fitted

energy is an unphysically large negative number. Obviously, the number of points cannot be too small for generating a PES for dynamics calculations.

Figure 3 shows two decay plots obtained on the analytical surface (solid curve) and on a fitted surface (dotted curve). The result on the fitted surface was obtained for $N_{\mathrm{SYM}} = 502$ with the valence internal coordinates. The two curves agree well with each other, both showing the initial statistical regime with a faster rate and then the dynamical regime with a slower rate. The rate constants were computed from the slopes of the dynamical regime. The curve for the analytical surface is smoother because of the larger ensemble size—it contains 2000 trajectories as opposed to 500 for the fitted surface.

An interesting question is how accurate a PES must be in order to obtain accurate rates. The answer for HOOH is that even a rather crude PES can yield accurate rates. To illustrate this point, we list in Table II the dependence of rms errors for energy and gradient on N_{SYM} calculated by using the 5,000 points generated as described in Sec. II B. Clearly, the average qualities of the fits are not good, but this result is expected because the number of data points used is not enough to accurately describe all the regions considered here of this 6-D potential. In spite of this, the computed rates are quite accurate, at least when using the valence internal coordinates. For a statistical system this can be understood from statistical theories that rate constants depend only on a few quantities characterizing the surface (the barrier height and the frequencies) and not on the accuracy of the entire surface. However, HOOH is not a completely statistical system. As discussed previously,[19] the intramolecular couplings in HOOH are strong except that the two OH bonds are so weakly coupled that when one OH bond is excited the energy does not flow to the other one and thus the unexcited OH mode does not contribute to reaction. This nonstatistical behavior needs to be correctly incorporated in the potential for an accurate description of the

reaction rate. The difference in the results for the two coordinate systems is probably because this weak coupling is more accurately described in the valence internal coordinates with the basis elements used, even though the overall qualities of the fitted surfaces are similar for the two coordinate systems as shown in Table II. In the case of using the interatomic distance coordinates, the potential couplings may be too strong, and the calculated rate is closer to the statistical value, which is faster than the true dynamical one.

To illustrate this nonstatistical behavior, we show in Fig. 4 a RRK plot of the rate constant as a function of energy, where diamonds and circles represent the rates calculated on the analytical and fitted surfaces, respectively. The results were obtained by using the valence internal coordinates with $N_{\mathrm{SYM}} = 502$. The middle point in the plot is the one listed in Table I. A least-squares fit of the data to the RRK expression $k(E) = v(1 - E_b / E)^{s-1}$ (E_b is the dissociation energy of O-O bond) yields $s \approx 3.4$ and 3.2 for the rates obtained on the analytical and fitted surfaces, respectively, which are less than 5, the theoretical number for a statistical system, indicating that the intramolecular energy transfer is not complete during reaction.

IV. SUMMARY AND CONCLUSIONS

We have further refined the IMLS method for fitting potential surfaces. We have investigated the effect of using a cut-off radius and selection of the data points and basis elements. And, for the first time, we have applied the method to dynamics calculations and computed rate constants for the O-O bond fission in HOOH using the classical trajectory approach.

Our results indicate that preliminary random samplings of IMLS evaluations with large basis sets can be monitored for basis function coefficient size. Smaller coefficients on the average can then be eliminated in the production runs. Our investigations of a cut-off radius suggests that

for this 6-D system, a cut-off radius should contain on the average a few hundred *ab initio* points for reasonable accuracy.

Our results also indicate that, for a nonstatistical system, using a proper coordinate system can be important in accurately describing intramolecular couplings and hence the reaction rate. For HOOH, the valence internal coordinates are better than the interatomic distances. The calculated rate is accurate for N_{SYM} as low as 152. If one uses a dual-level approach where IMLS is applied to the difference of the true PES with a reference PES, any apparent dependence on accuracy with the choice of coordinate system essentially disappears, demonstrating the effectiveness of employing a zeroth-order PES.

The IMLS method requires only energy values from *ab initio* calculations, whereas the modified Shepard method requires also first and second derivatives, which are computationally very demanding for high-dimensional systems. Thus, the IMLS method has the advantage of taking less computer time on electronic structure calculations. Another advantage is that higher than second-order terms can be easily included in doing the IMLS fits to improve accuracy, while higher-order *ab initio* derivatives need to be evaluated for the higher-order terms to be included in the modified Shepard method. However, the IMLS fitting can be more time consuming. A useful comparison of efficiency must take into account the combined cost of the *ab initio* calculations and the total number of evaluations needed for the application.

The results presented here suggest that the IMLS approach has the potential to be a useful means for generating surfaces for studying chemical reactions in high-dimensional systems. More improvements are to be made, such as a better procedure for selecting data points and using variable cut-offs to improve efficiency. Any improvements in IMLS have to be scalable to higher dimensions. That will be the direction of our future work.

ACKNOWLEDGMENTS This work was supported by the U. S. Department of Energy, Office of Basic Energy Sciences, Division of Chemical Sciences, and the Mathematical, Information, and Computational Sciences Division subprogram of the Office of Advanced Scientific Computing Research, Office of Science, U. S. Department of Energy under Contract No. W-31-109-Eng-38 (Argonne) and Contract No. DE-FG02-01ER15231 (OSU).

REFERENCES

[1] G. C. Schatz, *Proceedings of the European School on Computational Chemistry*, **75**, 15, (2000); G. O. de Aspuru; M. L. Hernandez, *Proceedings of the European School on Computational Chemistry*, **75**, 193 (2000).

[2] G. K. Chawla, G. C. McBane, P. L. Houston, and G. C. Schatz, J. Chem. Phys. **88**, 5481 (1988).

[3] J. N. Murrell, S. Carter, S. C. Farantos, P. Huxlay, and A. Varandas, *Molecular Potential Energy Functions* (Wiley, Chichester, 1984).

[4] G. C. Schatz, A. Papaioannou, L. A. Pederson, L. B. Harding, T. Hollebeek, T. S. Ho, and H. Rabitz, J. Chem. Phys. **107**, 2340 (1997).

[5] H. Koizumi, G. C. Schatz, and S. P. Walch, J. Chem. Phys. **95**, 4130 (1991).

[6] A. J. Dobbyn, J. N. L. Connor, N. A. Besley, P. J. Knowles, and G. C. Schatz, Phys. Chem. Chem. Phys. **1**, 957 (1999).

[7] J. Ischtwan and M. A. Collins, *J. Chem. Phys.* **100**, 8080 (1994).

[8] R. Farwig, in *Algorithms for Approximation*, edited by J. C. Mason and M. G. Cox (Clarendon, Oxford, 1987).

[9] G. G. Maisuradze and D. L. Thompson, J. Phys. Chem. **A. 107**, 7118 (2003).

[10] G. G. Maisuradze, D. L. Thompson, A. F. Wagner, and M. Minkoff, J. Chem. Phys. **119**, 1002 (2003).

[11] A. Kawano, Y. Guo, D. L. Thompson, A. F. Wagner, and M. Minkoff, J. Chem. Phys. **120**, 6414 (2004).

[12] T. Ishida and G. C. Schatz, *Chem. Phys. Lett.* **314**, 369 (1999).

[13] G. C. Schatz, in *Reaction and Molecular Dynamics, Lecture Notes in Chemistry, Vol. 14,* edited by A. Lagana and A. Riganelli (Springer, Berlin, 2000), p. 15.

[14] T. Ishida and G. C. Schatz, *J. Comput. Chem.* **24**, 1077 (2003).

[15] B. Kuhn, T. R. Rizzo, D. Luckhaus, M. Quack, and M. A. Suhm, J. Chem. Phys. **111**, 2565 (1999).

[16] See, for example, K. C. Thompson, and M. A. Collins, J. Chem. Soc., Faraday Trans. **93**, 871 (1997).

[17] (a) H. W. Schrantz, S. Nodholm, and G. Nyman, J. Chem. Phys. **94**, 1487 (1991).

[18] See, for example, B. Kuhn and T. R. Rizzo, J. Chem. Phys. **112**, 7461 (2000) and references therein.

[19] Y. Guo and D. L. Thompson, Chem. Phys. Lett. **382**, 654 (2003).

Table I. Computed rate constants on fitted surfaces in either valence internal or interatomic distance coordinates for E = 77.625 kcal mol^{-1}.

	Valence Internal Coordinates	Interatomic Distance Coordinates
$N(N_{SYM})$	k	$k(k')$[a]
300(152)	0.56	0.61(0.50)
500(252)	0.55	0.74(0.60)
1000(502)	0.58	0.78(0.53)
2000(1002)	—	0.73(0.53)
4000(2002)	—	0.70(0.50)
8000(4002)	—	0.56(0.52)
	0.54±0.02[b]	

[a] The values in parentheses in the last column are obtained with the use of the zeroth-order function.

[b] The value is obtained by using the analytical surface.

Table II. Dependence of rms errors for energy and gradient on the number of data points.[a]

	Valence Internal Coordinate		Interatomic Distance Coordinates	
$N(N_{SYM})$	ΔE	ΔG	ΔE	ΔG
300(152)	3.94	64.2	4.30	61.8
500(252)	3.35	54.1	3.23	44.6
1000(502)	1.97	36.6	2.54	40.3

[a] The units of ΔE and ΔG are kcal mol^{-1} and kcal mol^{-1}Å^{-1}, respectively.

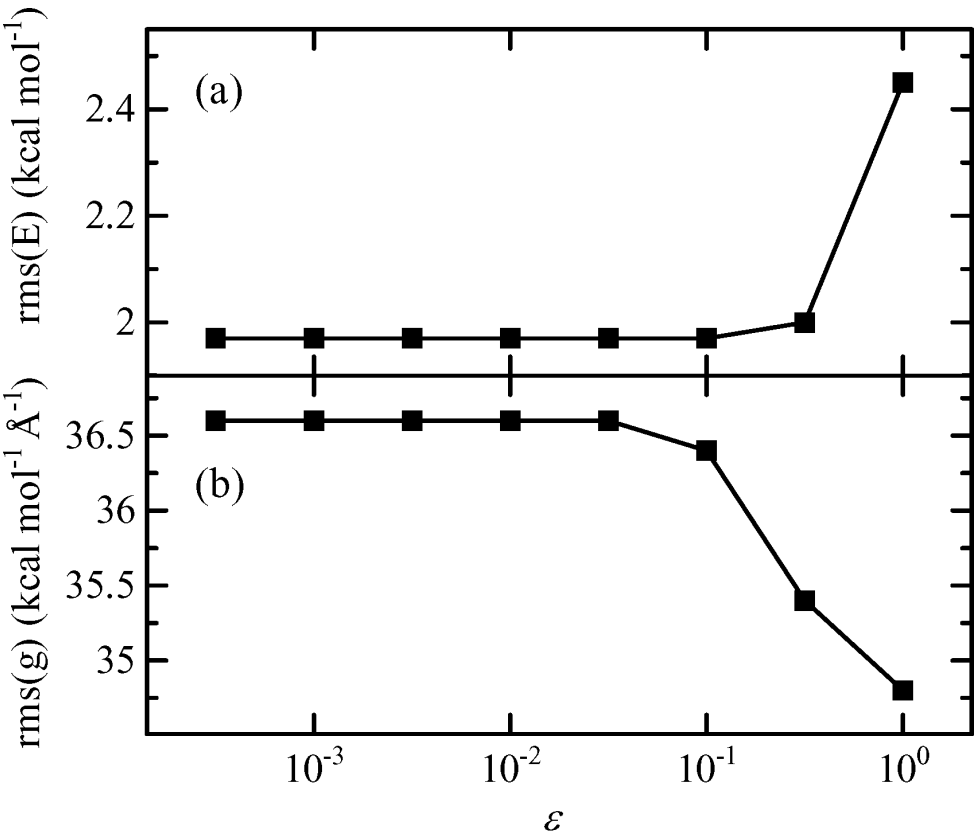
FIGURE CAPTIONS

Fig. 1: The rms errors for energy and gradient versus the parameters in the weight function ε, panels (a) and (b), and *p,* panels (c) and (d).

Fig. 2: The rms errors for energy (panel a) and gradient (panel b) and the average number of data points inside r_{cut} (panel c) versus the cut-off radius r_{cut}.

Fig. 3: Typical decay plots for the O-O bond fission in HOOH. The results were obtained on the analytical surface (solid curve) and a fitted surface (dotted curve) using the valence internal coordinates with $N_{\text{SYM}} = 502$. Here $P = \ln(N_t / N_0)$ and $E = 77.625$ kcal mol^{-1}.

Fig. 4: Rate constants obtained on the analytical surface (diamonds) and the fitted surface with $N_{\text{SYM}} = 502$ (circles). Here E is the total energy, and E_b is the O-O bond dissociation energy.


(a)
rms(E) (kcal mol-1)
2.4
2.2
2
(b)
rms(g) (kcal mol-1 Å-1)
36.5
36
35.5
35
10-3
10-2
10-1
100
ε

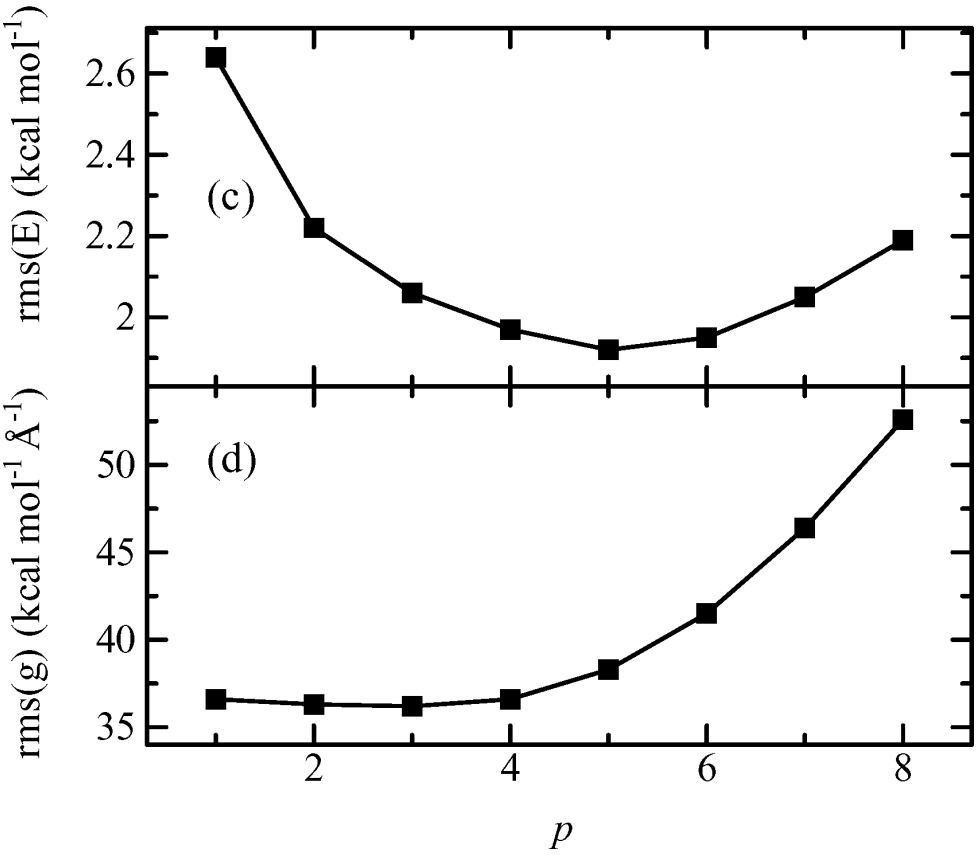

(c)
rms(E) (kcal mol-1)
2.6
2.4
2.2
2
(d)
rms(g) (kcal mol-1 Å-1)
50
45
40
35
2
4
6
8
p

Fig. 1

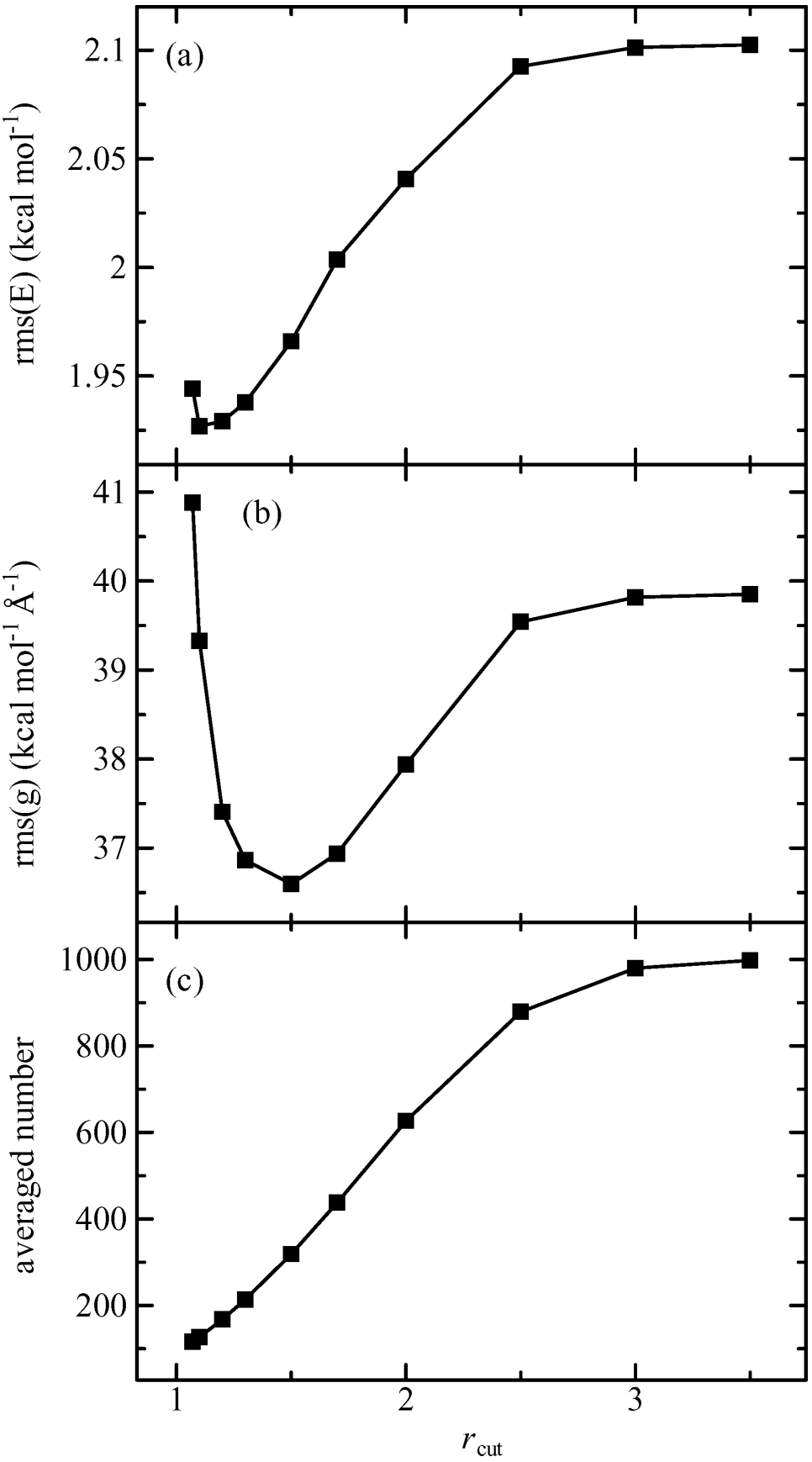

(a)
(b)
(c)
rms(E) (kcal mol^{-1})
rms(g) (kcal mol^{-1} Å^{-1})
averaged number
2.1
2.05
2
1.95
41
40
39
38
37
1000
800
600
400
200
1
2
3
r_{cut}

Fig. 2

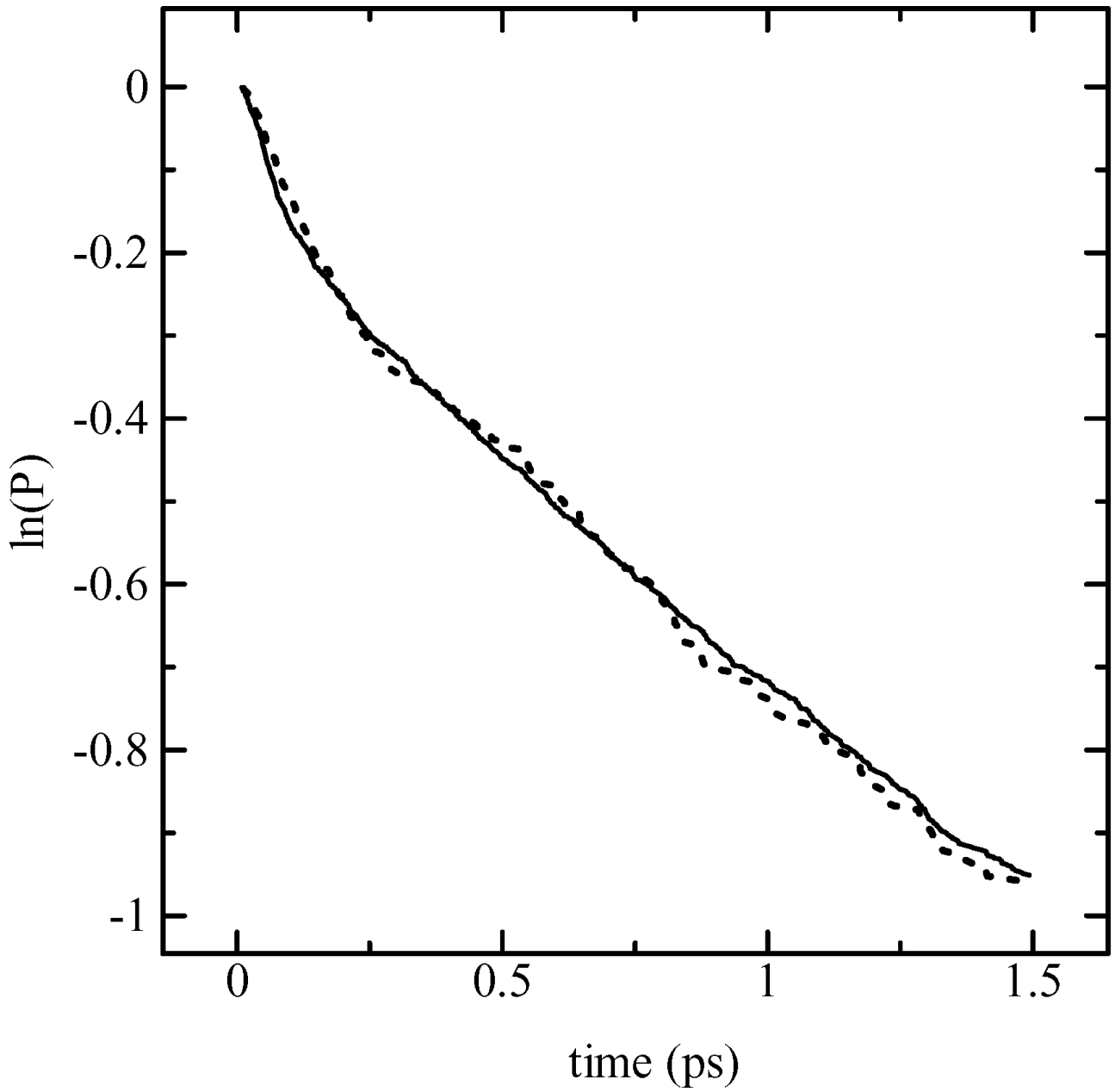


Fig.3

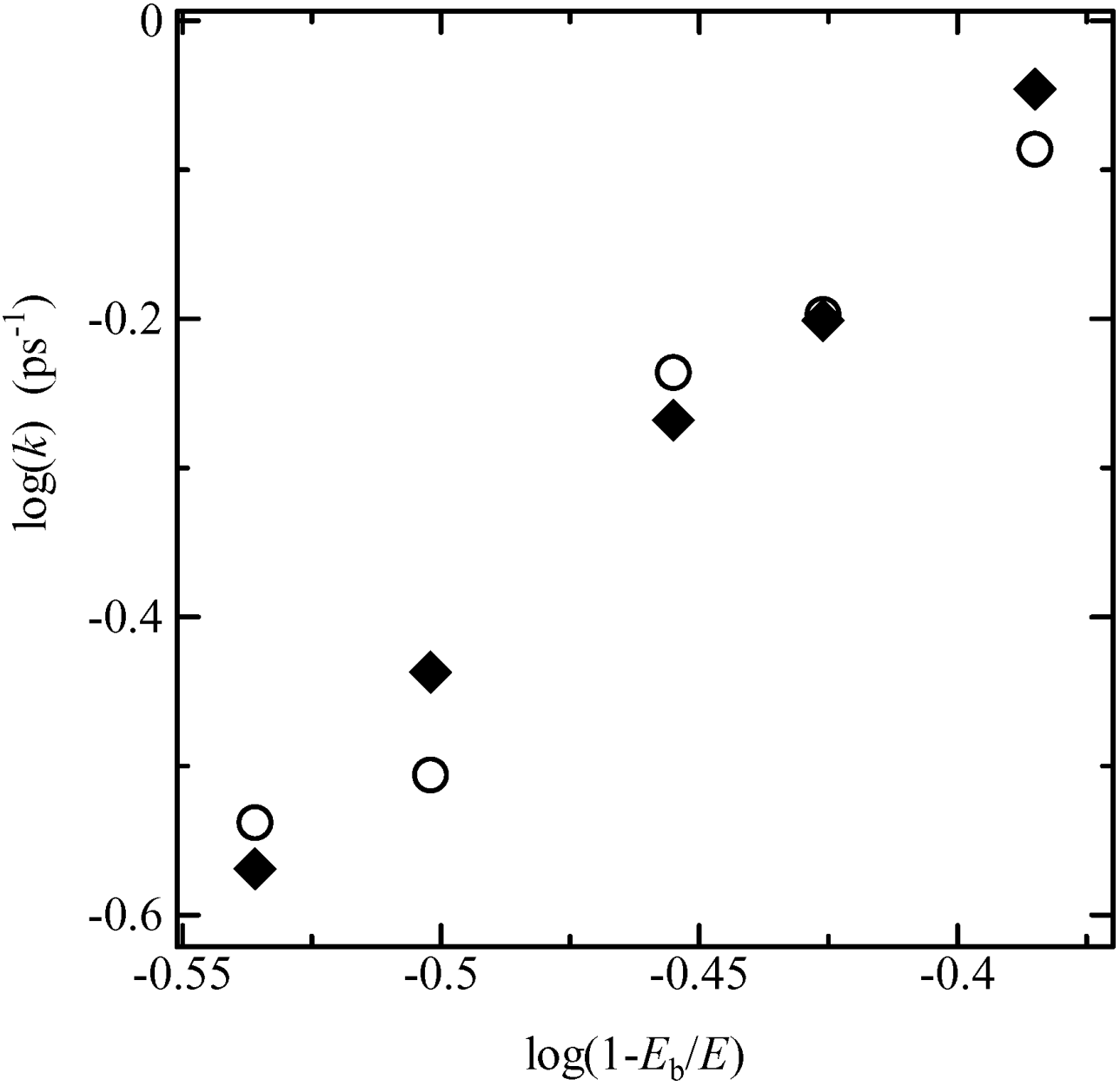


Fig. 4