

# Efficient global representations of potential energy functions: Trajectory calculations of bimolecular gas-phase reactions by multiconfiguration molecular mechanics

Oksana Tishchenko<sup>a)</sup> and Donald G. Truhlar<sup>b)</sup>

*Department of Chemistry and Supercomputing Institute, University of Minnesota, Minneapolis, Minnesota 55455-0431, USA*

(Received 22 September 2008; accepted 14 November 2008; published online 13 January 2009)

Multiconfiguration molecular mechanics (MCMM) was previously applied to calculate potential energies, gradients, and Hessians along a reaction path and in the large-curvature tunneling swath, and it was shown that one could calculate variational transition state theory rate constants with optimized multidimensional tunneling without requiring more than a few electronic structure Hessians. It was also used for molecular dynamics simulations of liquid-phase potentials of mean force as functions of a reaction coordinate. In the present article we present some improvements to the formalism and also show that with these improvements we can use the method for the harder problem of trajectory calculations on gas-phase bimolecular reactive collisions. In particular, we apply the MCMM algorithm to the model reaction  $\text{OH} + \text{H}_2 \rightarrow \text{H}_2\text{O} + \text{H}$ , for which we construct the global full-dimensional interpolated potential energy surfaces with various numbers of electronic structure Hessians and various molecular mechanics force fields, and we assess the quality of these fits by quasiclassical trajectory calculations. We demonstrate that chemical accuracy (1–2 kcal/mol) can be reached for a MCMM potential in dynamically important regions with a fairly small number of electronic structure Hessians. We also discuss the origins of the errors in the interpolated energies and a possible way to improve the accuracy. © 2009 American Institute of Physics. [DOI: 10.1063/1.3042145]

## I. INTRODUCTION

Computational chemical reaction dynamics usually involves two steps: (i) calculating the potential energy function<sup>1</sup> (PEF) or its gradient field (which is minus the force field) and (ii) calculating the dynamics *per se*. Step (i) may be accomplished in advance by developing an analytic representation of the PEF for a given reaction or, in principle, by choosing a general set of molecular mechanics parameters. Unfortunately, with a few exceptions,<sup>2</sup> molecular mechanics with general parameters is restricted to nonreactive systems. Furthermore the analytic representation of PEFs for specific reactions requires a certain artfulness and a considerable amount of human time and ingenuity. For these reasons the modern approach is to use direct dynamics calculations,<sup>3</sup> which have been defined<sup>4</sup> as calculations in which “instead of using a predefined PEF, all required energies and forces for each geometry that is important for evaluating dynamical properties are obtained directly from electronic structure calculations.” Considerable attention has been paid to developing efficient algorithms for direct dynamics calculations.<sup>5</sup> An ultimate goal would be to make the algorithm so efficient that the electronic structure calculations no longer dominate the cost or at least are quite affordable.

Some approaches are intermediate between fitting surfaces and direct dynamics. For example, a robust, systemati-

cally improvable, semiautomatic fitting scheme would eliminate the need for time-consuming human trial and error and artfulness, but since dynamics calculations would eventually be carried out by using the fit, it would become possible to carry out extensive dynamical explorations quite efficiently and affordably. The first available algorithm to come close to this ideal was the “Grow” algorithm of Collins and co-workers.<sup>6</sup> Using this algorithm, they were able to generate global potential energy surfaces for quasiclassical trajectory calculations on the reactions  $\text{OH} + \text{H}_2 \rightarrow \text{H}_2\text{O} + \text{H}$ ,<sup>6(a)</sup>  $\text{H} + \text{CH}_4 \rightarrow \text{H}_2 + \text{CH}_3$ ,<sup>6(b)</sup> and  $\text{H} + \text{N}_2\text{O} \rightarrow \text{OH} + \text{N}_2$ ,<sup>6(g)</sup> with 200–400, 1100–1300, and 1400 Hessians, respectively. Later<sup>6(e)</sup> they refined the  $\text{OH} + \text{H}_2$  surface using over 1600 Hessians, and a version based on 1000 Hessians was used for quantum scattering calculations.<sup>6(d)</sup> This method is based on using a modified Shepard interpolation to interpolate between the second-order Taylor expansions defined by the Hessians of the PEF at a set of geometries called Shepard points. A succinct summary with references to related work is provided in the introduction of Ref. 6(a). A more recent method, closely related to the method of Collins and co-workers,<sup>6</sup> is multiconfiguration molecular mechanics (MCMM). MCMM is also based on a modified version of Shepard interpolation, but the quantity interpolated is the off-diagonal element of a  $2 \times 2$  matrix corresponding to a valence bond treatment of two molecular mechanics configurations, one for products and one for reactants. In this respect, the MCMM method is reminiscent of the original  $2 \times 2$  valence-bond type treatment by London,<sup>7</sup> and is similar to the empirical valence bond

<sup>a)</sup>Electronic mail: o\_t@t1.chem.umn.edu.

<sup>b)</sup>Electronic mail: truhlar@umn.edu.

approach by Warshel and co-workers<sup>8</sup> except for the choice of  $V_{12}$  (which is crucial for realistic representation of a potential energy surface). The present implementation of the MCMM method allows one to treat multiple reactive intermediates, and a  $2 \times 2$  electronic diabatic Hamiltonian matrix allows one to treat reactions with a single bimolecular reaction channel (single reactant asymptote and single product asymptote, plus any additional asymptotes that are equivalent by symmetry). Extensions of the present model to a  $3 \times 3$  Hamiltonian are conceivable and can be implemented in the future.

The MCMM equations were originally presented without enforcing identical particle symmetry.<sup>9</sup> This is quite reasonable for methods such as variational transition state theory (VTST)<sup>10</sup> where one only needs to represent a potential energy surface in the vicinity of a single reaction path and the large-curvature tunneling swath.<sup>10</sup> The effect of identical particle symmetry can then be included by symmetry factors.<sup>11</sup> The original formalism<sup>9</sup> has been employed in subsequent work,<sup>12–16</sup> and along the way we have corrected some errors in the original equations. For completeness, the Appendix presents, for the first time, a corrected set of equations for the unsymmetrized case. The Appendix also includes refinements that are expected to lead to stable results in full dynamics calculations, where the refinement is required because general dynamics calculations depend on the potential energy surface more globally.

As an example of the need to consider more than just a single reaction path, consider a trajectory calculation<sup>17</sup> on collisions of  $H_aH_b$  with  $OH_c$ ; such a trajectory will likely visit not only geometries along the reaction paths leading to  $H_a+H_bOH_c$  and  $H_b+H_aOH_c$  but also intermediate geometries and perhaps even hydronium radical geometries. To treat such collisions realistically, we should use a potential energy surface with the correct permutational symmetry. Permutational symmetry of a potential energy surface has been enforced in various ways and emphasized by several groups.<sup>6,18,19</sup> A formalism for enforcing identical particle symmetry in the MCMM algorithm was presented in Ref. 19. A procedure with further refinements is presented in Sec. II in a notation consistent with the Appendix.

Section III presents details of the present applications of MCMM to construct potential energy surfaces for the example reaction  $OH+H_2 \rightarrow H_2O+H$ . Section IV discusses trajectory calculations for this reaction. The goal of these calculations is to demonstrate the usefulness of MCMM for full molecular dynamics simulations. Section V presents concluding remarks.

## II. MCMM ALGORITHM (SYMMETRIZED CASE)

The procedure for constructing a potential energy surface that is invariant with respect to the exchange of identical nuclei using MCMM is described in Ref. 19. The algorithm, with further refinements introduced here, is as follows.

(i) Select  $k=1, 2, \dots, N$  molecular geometries  $\mathbf{x}^{(k)}$  to be used as electronic structure Shepard points, and calculate the energies  $V^{(k)}$ , gradients  $\mathbf{G}^{(k)}$ , and Hessians  $\mathbf{F}^{(k)}$  at these geometries.

(ii) For each of these data points  $\mathbf{x}^{(k)}$  generate  $m!$  symmetrically equivalent data sets  $\{\mathbf{x}^{(k,i)}, \mathbf{G}^{(k,i)}, \mathbf{F}^{(k,i)}\}$ , where

$$\mathbf{x}^{(k,i)} = \mathbf{P}^{(i)} \mathbf{x}^{(k)}, \quad (1)$$

$$\mathbf{G}^{(k,i)} \equiv \frac{\partial}{\partial \mathbf{x}} V = \mathbf{P}^{(i)} \mathbf{G}^{(k)}, \quad (2)$$

$$\mathbf{F}^{(k,i)} \equiv \frac{\partial^2}{\partial \mathbf{x}^2} V = \mathbf{P}^{(i)} \mathbf{F}^{(k)} \mathbf{P}^{(i)}, \quad (3)$$

where  $\mathbf{P}$  is the nuclear permutation operator that interchanges Cartesian coordinates of  $m$  identical nuclei. We note that the strategy for enforcing symmetry by replicating electronic structure data was also employed by Collins and co-workers.<sup>6</sup>

(iii) Define a set of  $m!$  MM energies, gradients, and Hessians at point  $(k, i)$  by

$$V_{MM,n}^{(k,i)} \equiv V_{nn}(\mathbf{x}^{(k,i)}), \quad (4)$$

$$\mathbf{G}_{MM,n}^{(k,i)} \equiv \left. \frac{\partial}{\partial \mathbf{x}} V_{nn} \right|_{\mathbf{x}=\mathbf{x}^{(k,i)}}, \quad (5)$$

and

$$\mathbf{F}_{MM,n}^{(k,i)} \equiv \left. \frac{\partial^2}{\partial \mathbf{x}^2} V_{nn} \right|_{\mathbf{x}=\mathbf{x}^{(k,i)}}, \quad (6)$$

for  $n=1, 2; k=1, 2, \dots, N; \text{ and } i=1, 2, \dots, m!$ .

(iv) Define a symmetrized MM potential and its gradient and Hessian at point  $(k)$  (where a tilde denotes a symmetrization) by

$$\tilde{V}_n^{(k)} = -\frac{1}{\alpha} \ln \left( \frac{1}{\sigma_{mm}} \sum_i e^{-\alpha V_{MM,n}^{(k,i)}} \right), \quad (7)$$

where  $\alpha$  is a parameter, and  $\sigma_{mm}$ , which is called the symmetry factor, is the number of times the lowest-energy MM configuration occurs among the  $m!$  symmetrically equivalent MM configurations at a general geometry;

$$\tilde{\mathbf{G}}_n^{(k)} \equiv \frac{\partial}{\partial \mathbf{x}} \tilde{V}_n^{(k)} = \frac{\sum_i \mathbf{G}_{MM,n}^{(k,i)} e^{-\alpha V_{MM,n}^{(k,i)}}}{\sum_i e^{-\alpha V_{MM,n}^{(k,i)}}} \quad (8)$$

and

$$\begin{aligned} \tilde{\mathbf{F}}_n^{(k)} \equiv \frac{\partial^2}{\partial \mathbf{x}^2} \tilde{V}_n^{(k)} = & \frac{\sum_i (\mathbf{F}_{MM,n}^{(k,i)} - \alpha \mathbf{G}_{MM,n}^{(k,i)} \mathbf{G}_{MM,n}^{(k,i)\mathbf{T}}) e^{-\alpha V_{MM,n}^{(k,i)}}}{\sum_i e^{-\alpha V_{MM,n}^{(k,i)}}} \\ & + \alpha \tilde{\mathbf{G}}_n^{(k)} \tilde{\mathbf{G}}_n^{(k)\mathbf{T}}. \end{aligned} \quad (9)$$

Notice that the symmetrized MM potential is dominated by the  $\sigma_{mm}$  lowest-energy MM configurations among the  $m!$  permutations of the Cartesian coordinates.

(v) Generate  $m!$  values of  $\tilde{\mathbf{G}}_n^{(k,i)}$  and  $\tilde{\mathbf{F}}_n^{(k,i)}$  from each  $\tilde{\mathbf{G}}_n^{(k)}$  and  $\tilde{\mathbf{F}}_n^{(k)}$  by applying  $\mathbf{P}^{(i)}$ , as in step (ii).

(vi) Transform  $\mathbf{G}^{(k,i)}$ ,  $\mathbf{F}^{(k,i)}$ ,  $\tilde{\mathbf{G}}_n^{(k,i)}$ , and  $\tilde{\mathbf{F}}_n^{(k,i)}$  to the set of internal coordinates  $\mathbf{r}$  by the Wilson  $\mathbf{B}$  matrix and  $\mathbf{C}$  tensor. This yields

$$\mathbf{g}^{(k,i)} \equiv \left. \frac{\partial}{\partial \mathbf{r}} V \right|_{\mathbf{r}=\mathbf{r}(\mathbf{x}^{(k,i)})}, \quad (10)$$

$$\mathbf{f}^{(k,i)} \equiv \left. \frac{\partial^2}{\partial \mathbf{r}^2} V \right|_{\mathbf{r}=\mathbf{r}(\mathbf{x}^{(k,i)})}, \quad (11)$$

$$\tilde{\mathbf{g}}_n^{(k,i)} \equiv \left. \frac{\partial}{\partial \mathbf{r}} \tilde{V}_n \right|_{\mathbf{r}=\mathbf{r}(\mathbf{x}^{(k,i)})}, \quad (12)$$

$$\tilde{\mathbf{f}}_n^{(k,i)} \equiv \left. \frac{\partial^2}{\partial \mathbf{r}^2} \tilde{V}_n \right|_{\mathbf{r}=\mathbf{r}(\mathbf{x}^{(k,i)})}. \quad (13)$$

(vii) Define a matrix  $\mathbf{V}^{(k,i)}$  at each geometry  $(k,i)$  by

$$\mathbf{V}^{(k,i)}(\mathbf{r}) = \begin{pmatrix} \tilde{V}_n^{(k)}(\mathbf{r}) & V_{12}(\mathbf{r}, k, i) \\ V_{12}(\mathbf{r}, k, i) & \tilde{V}_n^{(k)}(\mathbf{r}) \end{pmatrix} \quad (14)$$

and construct Taylor series expansions of  $V_{12}$  around each data point  $(k,i)$  using the Taylor series reversion<sup>20</sup> in the same way as in Eq. (A6) for nonsymmetrized calculations, with the only difference being that  $(k)$  is replaced with  $(k,i)$ . Notice that  $\tilde{V}_n$  is a symmetrized molecular mechanics potential, and it does not depend on  $i$ .

(viii) Calculate the Taylor series coefficients  $D$ ,  $\mathbf{b}$ , and  $\mathbf{C}$  [as in Eqs. (A8)–(A10)] for each symmetrically equivalent data point  $(k,i)$ ,

$$D^{(k)} = (\tilde{V}_1^{(k)} - V^{(k)})(\tilde{V}_2^{(k)} - V^{(k)}), \quad (15)$$

$$\mathbf{b}^{(k,i)} = \frac{\tilde{\mathbf{g}}_1^{(k,i)} - \mathbf{g}^{(k,i)}}{\tilde{V}_1^{(k)} - V^{(k)}} + \frac{\tilde{\mathbf{g}}_2^{(k,i)} - \mathbf{g}^{(k,i)}}{\tilde{V}_2^{(k)} - V^{(k)}}, \quad (16)$$

$$\begin{aligned} \mathbf{C}^{(k,i)} = & (1/D^{(k,i)})(\tilde{\mathbf{g}}_1^{(k,i)} - \mathbf{g}^{(k,i)})(\tilde{\mathbf{g}}_2^{(k,i)} - \mathbf{g}^{(k,i)})^T \\ & + (\tilde{\mathbf{g}}_2^{(k,i)} - \mathbf{g}^{(k,i)})(\tilde{\mathbf{g}}_1^{(k,i)} - \mathbf{g}^{(k,i)})^T \\ & + \frac{\tilde{\mathbf{f}}_1^{(k,i)} - \mathbf{f}^{(k,i)}}{\tilde{V}_1^{(k)} - V^{(k)}} + \frac{\tilde{\mathbf{f}}_2^{(k,i)} - \mathbf{f}^{(k,i)}}{\tilde{V}_2^{(k)} - V^{(k)}}. \end{aligned} \quad (17)$$

Notice that  $D$  and  $V$  do not depend on  $i$ . The Taylor series of  $(V_{12})^2$  for each  $(k,i)$  at an arbitrary geometry  $\mathbf{r}=\mathbf{r}(\mathbf{x})$  can now be written as

$$\begin{aligned} [V_{12}(\mathbf{r}, k, i)]^2 = & D^{(k)} \left( 1 + \mathbf{b}^{(k,i)T} \Delta \mathbf{r}^{(k,i)} \right. \\ & \left. + \frac{1}{2} \Delta \mathbf{r}^{(k,i)T} \mathbf{C}^{(k,i)} \Delta \mathbf{r}^{(k,i)} \right). \end{aligned} \quad (18)$$

(ix) Define the matrix  $\mathbf{V}$  at the input geometry  $\mathbf{x}$  by

$$\mathbf{V}(\mathbf{x}) = \begin{pmatrix} \tilde{V}_n(\mathbf{x}) & V_{12}^S(\mathbf{x}) \\ V_{12}^S(\mathbf{x}) & \tilde{V}_n(\mathbf{x}) \end{pmatrix}. \quad (19)$$

The lowest-energy eigenvalue of this matrix is the MCMM PEF. The diagonal matrix elements  $\tilde{V}_n(\mathbf{x})$  and their derivatives,  $\tilde{\mathbf{G}}_n(\mathbf{x})$  and  $\tilde{\mathbf{F}}_n(\mathbf{x})$ , are obtained as follows: First we define

$$V_{MM,n}^{(j)} \equiv V_{nm}^{(j)}(\mathbf{x}), \quad j = 1, \dots, m!, \quad (20)$$

$$\mathbf{G}_{MM,n}^{(j)} \equiv \frac{\partial}{\partial \mathbf{x}} V_{nm}^{(j)}, \quad j = 1, \dots, m!, \quad (21)$$

and

$$\mathbf{F}_{MM,n}^{(j)} \equiv \frac{\partial^2}{\partial \mathbf{x}^2} V_{nm}^{(j)}, \quad j = 1, \dots, m!, \quad (22)$$

where each value of  $j$  corresponds to one of the  $m!$  connectivity patterns. Then,  $\tilde{V}_n(\mathbf{x})$ ,  $\tilde{\mathbf{G}}_n(\mathbf{x})$ , and  $\tilde{\mathbf{F}}_n(\mathbf{x})$  are calculated as

$$\tilde{V}_n(\mathbf{x}) = -\frac{1}{\alpha} \ln \left( \frac{1}{\sigma_{mm}} \sum_j e^{-\alpha V_{MM,n}^{(j)}(\mathbf{x})} \right), \quad (23)$$

$$\tilde{\mathbf{G}}_n(\mathbf{x}) = \frac{\sum_j \mathbf{G}_{MM,n}^{(j)}(\mathbf{x}) e^{-\alpha V_{MM,n}^{(j)}(\mathbf{x})}}{\sum_j e^{-\alpha V_{MM,n}^{(j)}(\mathbf{x})}}, \quad (24)$$

and

$$\begin{aligned} \tilde{\mathbf{F}}_n(\mathbf{x}) = & \frac{\sum_j (\mathbf{F}_{MM,n}^{(j)}(\mathbf{x}) - \alpha \mathbf{G}_{MM,n}^{(j)}(\mathbf{x}) \mathbf{G}_{MM,n}^{(j)T}(\mathbf{x})) e^{-\alpha V_{MM,n}^{(j)}(\mathbf{x})}}{\sum_j e^{-\alpha V_{MM,n}^{(j)}(\mathbf{x})}} \\ & + \alpha \tilde{\mathbf{G}}_n \tilde{\mathbf{G}}_n^T, \end{aligned} \quad (25)$$

where  $V_{MM,n}^{(j)}(\mathbf{x})$ ,  $\mathbf{G}_{MM,n}^{(j)}(\mathbf{x})$ , and  $\mathbf{F}_{MM,n}^{(j)}(\mathbf{x})$  are sets of  $m!$  MM energies, gradients, and Hessians at the geometry  $\mathbf{x}$ .

The off-diagonal elements are obtained via Shepard interpolation for  $V_{12}$  as in Eq. (A13) but with  $m!N$  terms in the sum

$$V_{12}^S(\mathbf{r}) = \sum_{k=1}^N \sum_{i=1}^{m!} w_{ki}(\mathbf{r}) V'_{12}(\mathbf{r}, k, i), \quad (26)$$

where  $w_{ki}$  are normalized weights, and  $V'_{12}$  is defined by

$$V'_{12}(\mathbf{r}, k, i) = \sqrt{V_{12}(\mathbf{r}, k, i)^2 u(\mathbf{r}, k, i)}, \quad (27)$$

where  $V_{12}(\mathbf{r}, k, i)^2$  is given in Eq. (18), and

$$u(\mathbf{r}, k, i) = \begin{cases} \frac{1}{1 + (\Delta/V_{12}(\mathbf{r}, k, i))^{2n}}, & V_{12}(\mathbf{r}, k, i)^2 > 0 \\ 0 & \text{otherwise.} \end{cases} \quad (28)$$

The normalized weight function is

$$w_{ki}(\mathbf{s}) = \frac{Y_{ki}(\mathbf{r})}{d_{ki}(\mathbf{s})^4} \frac{1}{\sum_{k=1}^{(N+2)} \sum_{i=1}^{m!} \frac{Y_{ki}(\mathbf{r})}{d_{ki}(\mathbf{s})^4}}, \quad (29)$$

where  $d_{ki}$  is the generalized distance between  $\mathbf{s}$  and  $\mathbf{s}^{(k,i)}$  defined as

$$d_{ki}(\mathbf{s}) = \sqrt{\sum_{\gamma=1}^{\Gamma m!} (s_{\gamma} - s_{\gamma}^{(k,i)})^2}, \quad (30)$$

where  $\mathbf{s} = \{s_1, s_2, \dots, s_{\gamma}, \dots, s_{\Gamma}\}$ . The current implementation only supports the following cases: (i) set  $\mathbf{s}$  is the same as set  $\mathbf{r}$ . and (ii) set  $\mathbf{s}$  is a subset of  $\mathbf{r}$ . The scaling coefficients  $Y_{ki}(\mathbf{r})$  are chosen such that the reactive system is described

by pure molecular mechanics in the asymptotic regions. In particular, we define  $Y_{ki}$  at a geometry  $\mathbf{r}$  as

$$Y_{ki}(\mathbf{r}) = \frac{1}{1 + \left( \frac{V'_{12}(\mathbf{r}, k, i)^2 - D^{(k)}}{A^2} \right)^\mu}, \quad (31)$$

where  $A$  and  $\mu$  are parameters.

The gradient and Hessian of Eq. (26) with respect to internal coordinates are given by

$$\mathbf{g}^S(\mathbf{r}) \equiv \frac{\partial V_{12}^S(\mathbf{r})}{\partial \mathbf{r}} = \sum_{k=1}^N \sum_{i=1}^{m!} \left[ \frac{\partial w_{ki}}{\partial \mathbf{r}} V'_{12}(\mathbf{r}, k, i) + w_{ki} \mathbf{g}_{12}(\mathbf{r}, k, i) \right], \quad (32)$$

$$\mathbf{f}^S(\mathbf{r}) \equiv \frac{\partial^2 V_{12}^S(\mathbf{r})}{\partial \mathbf{r}^2} = \sum_{k=1}^N \sum_{i=1}^{m!} \left( \frac{\partial^2 w_{ki}}{\partial \mathbf{r}^2} V'_{12}(\mathbf{r}, k, i) + \frac{\partial w_{ki}}{\partial \mathbf{r}} \mathbf{g}_{12}(\mathbf{r}, k, i)^\top + \mathbf{g}_{12}(\mathbf{r}, k, i) \times \left( \frac{\partial w_{ki}}{\partial \mathbf{r}} \right)^\top + w_{ki} \mathbf{f}_{12}(\mathbf{r}, k, i) \right), \quad (33)$$

where

$$\mathbf{g}_{12}(\mathbf{r}, k, i) \equiv \frac{\partial V'_{12}(\mathbf{r}, k, i)}{\partial \mathbf{r}} = \frac{1}{2V'_{12}(\mathbf{r}, k, i)} D^{(k)} (\mathbf{b}^{(k,i)} + \mathbf{C}^{(k,i)} \Delta \mathbf{r}^{(k,i)}) u(\mathbf{r}, k, i) \times \left( 1 + \left( \frac{\Delta}{V_{12}(\mathbf{r}, k, i)} \right)^n nu(\mathbf{r}, k, i) \right) \quad (34)$$

and

$$\mathbf{f}_{12}(\mathbf{r}, k, i) \equiv \frac{\partial^2 V'_{12}(\mathbf{r}, k, i)}{\partial \mathbf{r}^2} = \frac{1}{V'_{12}(\mathbf{r}, k, i)} \left( -\mathbf{g}_{12}(\mathbf{r}, k, i) \mathbf{g}_{12}(\mathbf{r}, k, i)^\top + \frac{D^{(k)2} n^2 (\mathbf{b}^{(k,i)} + \mathbf{C}^{(k,i)} \Delta \mathbf{r}^{(k,i)})^2 u(\mathbf{r}, k, i)^2}{(V_{12}(\mathbf{r}, k, i))^2} \times \left( \frac{\Delta}{(V_{12}(\mathbf{r}, k, i))^2} \right)^n \left\{ nu(\mathbf{r}, k, i) \left( \frac{\Delta}{(V_{12}(\mathbf{r}, k, i))^2} \right)^n - (n-1) \right\} + \frac{1}{2} D^{(k)} \mathbf{C}^{(k,i)} u(\mathbf{r}, k, i) \times \left( nu(\mathbf{r}, k, i) \left( \frac{\Delta}{(V_{12}(\mathbf{r}, k, i))^2} \right)^n + 1 \right) \right). \quad (35)$$

As in the nonsymmetrized MCMM calculations, the first and second derivatives  $\partial w_{ki} / \partial \mathbf{r}$  and  $\partial^2 w_{ki} / \partial \mathbf{r}^2$  in Eqs. (32) and (33) are obtained numerically. Since all operations except for this numerical intermediate step are analytic, the final MCMM derivatives may be called semi-analytical.

These derivatives of  $V_{12}^S$  are then transformed from the internal coordinates  $\mathbf{r}$  to Cartesian coordinates by using the transformation matrices saved in an earlier step, in the same fashion as in the formalism for nonsymmetrized potential energy surfaces.

(x) Find the eigenvalue  $V$  of Eq. (19) and its derivatives in Cartesian coordinates. The lowest eigenvalue of Eq. (19) is given by

$$V(\mathbf{x}) = \frac{1}{2} (\tilde{V}_1(\mathbf{x}) + \tilde{V}_2(\mathbf{x})) - [(\tilde{V}_1(\mathbf{x}) - \tilde{V}_2(\mathbf{x}))^2 + 4(V_{12}^S(\mathbf{x}))^2]^{1/2}, \quad (36)$$

where  $\tilde{V}_n$  are the symmetrized uninterpolated MM potentials given by Eq. (23), and  $V_{12}^S$  is the resonance integral obtained via the  $m!N$ -term Shepard interpolation, Eq. (26). The gradient and Hessian components of  $V$  with respect to Cartesian coordinates are then given by

$$G_i = \frac{\partial V}{\partial x_i} = \frac{1}{2} \left( \tilde{G}_{1i} + \tilde{G}_{2i} - \left( \frac{4V_{12}^S \left( \frac{\partial V_{12}^S}{\partial x_i} \right) + (\tilde{V}_1 - \tilde{V}_2)(\tilde{G}_{1i} - \tilde{G}_{2i})}{((\tilde{V}_1 - \tilde{V}_2)^2 + 4(V_{12}^S)^2)^{1/2}} \right) \right) \quad (37)$$

and

$$F_{ij} = \frac{\partial^2 V}{\partial x_i \partial x_j} = \frac{1}{2} \left( \tilde{F}_{1ij} + \tilde{F}_{2ij} + \frac{\left( 4V_{12}^S \left( \frac{\partial V_{12}^S}{\partial x_i} \right) + (\tilde{V}_1 - \tilde{V}_2)(\tilde{G}_{1i} - \tilde{G}_{2i}) \right)}{((\tilde{V}_1 - \tilde{V}_2)^2 + 4(V_{12}^S)^2)^{3/2}} \left( 4V_{12}^S \left( \frac{\partial V_{12}^S}{\partial x_j} \right) + (\tilde{V}_1 - \tilde{V}_2)(\tilde{G}_{1j} - \tilde{G}_{2j}) \right) \right. \\ \left. - \frac{4 \left( \frac{\partial V_{12}^S}{\partial x_i} \right) \left( \frac{\partial V_{12}^S}{\partial x_j} \right) + (\tilde{G}_{1i} + \tilde{G}_{2i})(\tilde{G}_{1j} + \tilde{G}_{2j})}{((\tilde{V}_1 - \tilde{V}_2)^2 + 4(V_{12}^S)^2)^{1/2}} - \frac{4 \left( \frac{\partial^2 V_{12}^S}{\partial x_i \partial x_j} \right) + (\tilde{V}_1 - \tilde{V}_2)(\tilde{F}_{1ij} - \tilde{F}_{2ij})}{((\tilde{V}_1 - \tilde{V}_2)^2 + 4(V_{12}^S)^2)^{1/2}} \right). \quad (38)$$

### III. REPRESENTATION OF GLOBAL SYMMETRIZED POTENTIAL ENERGY SURFACE FOR THE REACTION $\text{OH} + \text{H}_2 \rightarrow \text{HOH} + \text{H}$ WITH MCMM

The Born–Oppenheimer potential energy at an arbitrary geometry  $\mathbf{x}$  is represented by the lowest eigenvalue of the matrix  $\mathbf{V}$  given in Eq. (19), where  $\tilde{V}_{11}(\mathbf{x})$  and  $\tilde{V}_{22}(\mathbf{x})$  are the symmetrized analytical PEFs that describe molecular mechanics configurations corresponding, respectively, to a reactant and a product [ $\text{OH} + \text{H}_2$  ( $n=1$ ) and  $\text{H}_2\text{O} + \text{H}$  ( $n=2$ ) in the case of reaction  $\text{OH} + \text{H}_2 \rightarrow \text{H}_2\text{O} + \text{H}$ ], and  $V_{12}^S(\mathbf{x})$  is the coupling term obtained via Eq. (26). The set of internal coordinates  $\mathbf{r}$  used in the present work consists of the six internuclear distances. This set is the same as set  $\mathbf{s}$  used to calculate the generalized distance of Eq. (30):  $\mathbf{s} \equiv \mathbf{r} \equiv \{r_{\text{OH}_a}, r_{\text{OH}_b}, r_{\text{OH}_c}, r_{\text{H}_a\text{H}_b}, r_{\text{H}_b\text{H}_c}, r_{\text{H}_a\text{H}_c}\}$ .

#### III.A. The target potential energy surface and DFT calculations

We will assess the performance of the MCMM procedure presented above by testing an interpolated energy at dynamically important nuclear configurations against the “accurate” or target potential that MCMM is trying to reproduce. The target results for the present study are obtained using an affordable and efficient density functional, in particular, the MPWB1K (Ref. 21) functional, which is a hybrid metafunctional designed for kinetics. The zero-point-exclusive reaction barrier height and reaction energy of 4.9 and  $-13.6$  kcal/mol, respectively, obtained by MPWB1K calculations with the standard 6-31+G( $d,p$ ) (Ref. 22) basis set, are in reasonable agreement with the best available estimates<sup>23</sup> of 5.7 and  $-16.3$  kcal/mol, respectively. The frequencies (3898, 2614, 1070, 634, 589, and 1108i  $\text{cm}^{-1}$ ), breaking and forming bond distances (0.82 and 1.32 Å), and bond angle (164°) at the transferring H atom at the saddle point are also in reasonable agreement with best available estimates<sup>6(e)</sup> (3771, 2622, 1051, 598, 501, 1192i, 0.82, 1.36, and 165, respectively). Since our goal here is to test the accuracy of the MCMM interpolation scheme rather than to compare the results derived from the potential energy surface to experimental data, the use of this computational level is adequate for the present purposes. The MPWB1K/6-31+G( $d,p$ ) results are simply called the density functional theory (DFT) results in the rest of this article.

#### III.B. Parameters

MCMM is an interpolation procedure rather than a fitting procedure, but we did introduce five nonlinear parameters ( $\alpha$ ,  $\Delta$ ,  $n$ ,  $A$ , and  $\mu$ ) as components of the interpolation; these are set to practical working values rather than fully optimized. In particular, for  $\alpha$  of Eq. (7) we used 15.0 kcal/mol, for  $\Delta$  and  $n$  of Eq. (28) we used  $\sqrt{2.56} \times 10^{-8} E_h$  and 2, respectively, and for  $A$  and  $\mu$  parameters of Eq. (31) we used  $\sqrt{6} \times 10^{-4} E_h$  and 4, respectively.

#### III.C. $\tilde{V}_{11}$ and $\tilde{V}_{22}$

As noted above, MCMM takes advantage of pre-existing molecular mechanics potentials to represent the diagonal elements  $\tilde{V}_{11}$  and  $\tilde{V}_{22}$  of the Hamiltonian. Due to the availability of such potentials for a wide variety of systems in previously optimized molecular force fields, for example, MM3,<sup>24</sup> CHARMM,<sup>25</sup> etc., no design of new functional forms or optimization of parameters is required for constructing of a potential energy surface with MCMM (although, some refinement may be desirable for the best results<sup>15</sup>). Previous work<sup>9,12,15</sup> has demonstrated that the use of the MM3 (Ref. 24) force field leads to satisfactory results for the representation of reactive potential energy surfaces for a diverse set of reactions in the vicinity of the reaction paths. In case of the reaction  $\text{OH} + \text{H}_2$ , however, all necessary molecular mechanics parameters are not available; therefore, we have performed an additional step of obtaining these parameters, in particular, the parameters to describe the van der Waals interaction energy. The remaining part of this subsection describes the molecular mechanics functions used in the present work.

In the diagonal elements of Eq. (19), bond stretches are represented by Morse potentials, the H–O–H angle bending potential energy is described by a sixth-order polynomial, and the van der Waals interaction energy for each pair of nonbonded, non-geminal atoms is represented by

$$V_{\text{vdw}}(r) = A' e^{-B'r} + C' r^{-6} + D' r^{-12}, \quad (39)$$

where  $r$  is the distance between the atoms. Note that even though  $\text{H} \cdots \text{H}$  and  $\text{O} \cdots \text{H}$  interactions are present in both reactants and products, we used different parameters for these two cases.

Two sets of molecular mechanics parameters have been used; these parameters are collected in Tables I and II. In one set (denoted as parameter set 1, abbreviated as p1), all force

TABLE I. Force field parameters (set p1).

	Morse parameters			Angle bending parameters	
	H-H in H <sub>2</sub>	O-H in OH	O-H in H <sub>2</sub> O	H-O-H	
$r_e$ (Å)	0.7414 <sup>a</sup>	0.9707 <sup>a</sup>	0.9470 <sup>a</sup>		
$f$ (mdyn/Å)	5.752 <sup>b</sup>	7.8 <sup>b</sup>	8.45 <sup>b</sup>		
$D_e$ (kcal/mol)	150.0	150.0	150.0		
$\angle$ HOH (deg)				105.0 <sup>c</sup>	
$f$ (mdyn/rad)				0.63 <sup>c</sup>	
$c, q, p, s$ <sup>d</sup>				MM3 <sup>e</sup>	
van der Waals parameters (atomic) <sup>f</sup>					
	atomic H	H in H <sub>2</sub>	H in H <sub>2</sub> O	O in OH	O in H <sub>2</sub> O
$r_m$ (Å)	1.32 <sup>c</sup>	1.20 <sup>c</sup>	1.32 <sup>c</sup>	1.62 <sup>c</sup>	1.82 <sup>b</sup>
$\epsilon$ (kcal/mol)	0.016	0.016 <sup>c</sup>	0.008 <sup>c</sup>	0.059 <sup>c</sup>	0.036 <sup>c</sup>

<sup>a</sup>Reference 27.<sup>b</sup>Reference 26.<sup>c</sup>Reference 19.<sup>d</sup>Cubic, quartic, pentic, and sextic coefficients in the Taylor series.<sup>e</sup>Reference 24.<sup>f</sup>The coefficients in Eq. (40) are the same as in the MM3 (Ref. 24) force field:  $A=18\,400$ ,  $B=12.0$ ,  $C=2.5$ , and  $D$  is the same as in the previous work ( $D=0.01$ ) (Ref. 19).

constants and the  $r_e$  values were taken from Refs. 26 and 27, respectively. The bond dissociation energies ( $D_e$ ) were set to 150 kcal/mol for each bond type. This value is considerably larger than the best available estimates of  $D_e$  in each case in order to make it less likely that negative values of  $V_{12}(\mathbf{r}, k, i)^2$  are obtained. The angle bending energy parameters and the van der Waals parameters (except  $\epsilon$  for atomic H) were taken from the previous work.<sup>19</sup> Following the original notation of the MM3 (Ref. 24) force field, the van der Waals parameters in this case (set p1) are given as individual parameters  $r'_m$  and  $\epsilon'$  for each molecular mechanics atom type. Then  $r_m$  and  $\epsilon$  values for the interaction of an atom type of 1 with an atom of type 2 are obtained from a pair of atomic  $r'_m$  and  $\epsilon'$  as the arithmetic mean and the geometric mean, respectively. Equation (39) is written in terms of these parameters as<sup>15</sup>

$$V_{\text{vdW}}(r) = \epsilon \left[ A e^{-Br/r_m} - C \left( \frac{r_m}{r} \right)^6 \right] + DE \left( \frac{r_m}{r} \right)^{12}, \quad (40)$$

where

$$E = \frac{V_{6\text{-exp}}(r)}{\left( \frac{r_m}{r} \right)^{12}} \bigg|_{r=(1/2)r_m}, \quad (41)$$

and  $V_{\text{exp-6}}(r)$  is the exp-6 potential in the original MM3 force field

$$V_{\text{exp-6}}(r) = \epsilon \left[ A e^{-Br/r_m} - C \left( \frac{r_m}{r} \right)^6 \right]. \quad (42)$$

Note that Eq. (40) is a modified form of the exp-6 potential [Eq. (42)] used in the original MM3 force field.

The second set of parameters (abbreviated p2) uses the  $r_e$  values that correspond to equilibrium bond lengths for reactants and products at the target (i.e., DFT) computational level. The force constants are chosen such that the calculated harmonic vibrational frequencies for reactants and products derived from the interpolated potential energy surface are equal to the harmonic vibrational frequencies obtained by the DFT calculations. The four coefficients of Eq. (39) for each

TABLE II. Force field parameters (set p2).

	Morse parameters			Angle bending parameters	
	H-H in H <sub>2</sub>	O-H in OH	O-H in H <sub>2</sub> O	H-O-H	
$r_e$ (Å)	0.738	0.967	0.9470		
$f$ (mdyn/Å)	6.1	8.34	9.1		
$D_e$ (kcal/mol)	150.0	150.0	150.0		
$\angle$ HOH (deg)				106.1	
$f$ (mdyn/rad)				0.68	
van der Waals parameters (pairwise)					
	H···H in reactant	O···H in reactant	H···H in product	O···H in product	
$A'$	1638.355	248.169	259.226	4746.517	
$B'$	4.500	2.300	3.400	3.000	
$C'$	-2.474	-11.096	-0.339	-391.536	
$D'$	0.0882	384.428	0.269	772.927	

TABLE III. MUEs [averaged over 480–1613 geometries, depending on the case, for the energy range from  $-13.6$  kcal/mol (product asymptote) up to  $24.9$  kcal/mol] (kcal/mol) for interpolated potential energy surfaces with different parameter sets and numbers of the electronic structure Shepard points ( $N$ ).

Parameters		$N$	MUE
van der Waals	Bond stretching and angle bending		
Set p1	Set p1	13	3.3
Set p1	Set p2	13	2.3
Set p2	Set p2	13	1.7
Set p2	Set p2	20	1.3
Set p2	Set p2	31	1.3
Set p2	Set p2	37	1.4
Set p2	Set p2	3	4.0

van der Waals interaction ( $O\cdots H$  and  $H\cdots H$  in  $OH+H_2$  and  $O\cdots H$  and  $H\cdots H$  in  $H_2O+H$ ) are obtained by the following procedure: (i) The “van der Waals” interaction energies for reactants and products were calculated at the DFT level for several types of approaches of the reactants and products. These types of geometries (three for reactants and two for products) are shown in the supplementary material.<sup>28</sup> (ii) Eight coefficients [ $A'$ ,  $B'$ ,  $C'$ , and  $D'$  of Eq. (39) for  $O\cdots H$  and  $H\cdots H$  interactions] were obtained for reactants and products by least-squares fits to the calculated energies. More details of the fitting procedure are given in the supplementary material.<sup>28</sup>

### III.D. $V_{12}$ , locations of Shepard points, and quasiclassical dynamics calculations

The coupling term  $V_{12}^S(\mathbf{r})$  of Eq. (19) is obtained via Shepard interpolation as given in Eq. (26). Quasiclassical trajectories<sup>17</sup> were propagated on the interpolated potential energy surfaces that were obtained with various numbers of Shepard points. A Bulirsch–Stoer algorithm<sup>29</sup> with an adaptive time step size was used in trajectory calculations. The trajectory initial conditions were calculated on the interpolated surface using quasiclassical sampling. The vibrational and rotational quantum numbers for OH and  $H_2$  were set to 0 and 2, respectively, and the relative translational energy was set to 9.9 kcal/mol; the total energy for each trajectory was  $\sim 22.5$  kcal/mol. The initial impact parameter was set to zero. All trajectories were started at a distance between the oxygen atom and the farthest hydrogen atom of the  $H_2$  molecule corresponding to approximately 4.0 Å, and all trajectories were terminated when the oxygen atom in the water molecule and atomic hydrogen were separated by approximately the same distance.

We started with three surfaces (with different parameters—see Table III), each based on 13 electronic structure Shepard points. These points are placed at the following locations: at three stationary structures specified below, and along the intrinsic reaction coordinate path for reaction  $OH+H_2\rightarrow H_2O+H$  calculated at the target level. The three stationary points are (i) the reaction saddle point,

(ii) the ammonialike  $OH_3$  shallow local minimum at 15.3 kcal/mol above the global minimum, and (iii) the saddle point that connects the ammonialike local minimum with the van der Waals complex formed by  $H_2O$  and H, at 16.0 kcal/mol above the global minimum. Note that structures (ii) and (iii) are considerably higher in energy than the saddle point for the main reaction channel, and they do not contribute significantly to the interpolated potential energy surface near the reaction path of reaction  $OH+H_2\rightarrow H_2O+H$ . We also note that although the initial conditions in trajectory calculations were set so that the  $OH_3$  well is energetically accessible, we have never observed trajectories visiting this well. The Shepard points along the intrinsic reaction coordinate (in mass-weighted Cartesian coordinates) are approximately equally spaced. In addition, we also place two molecular mechanics Shepard points at locations corresponding to the van der Waals minima of the reactant and product complexes optimized at the MM level;  $V_{12}^S$  and its gradient and Hessian are taken as zero at these points. Thus,  $N=13$  in Eqs. (26) and (29). Following the notation introduced in Ref. 19, we denote this surface MCMM(13) where the number in parentheses is the number  $N$  of electronic structure Shepard points. Furthermore, we denote a MCMM( $N$ ) potential energy surface constructed using set pX of molecular mechanics parameters as MCMM( $N$ )/pX.

We also briefly discuss a surface based on electronic structure Hessians at only the three stationary structures (i)–(iii); this is called MCMM(3).

Three additional surfaces, with  $N$  equal to 20, 31, and 37, were generated by adding additional electronic structure Shepard points to the third MCMM(13) potential energy surface following a scheme analogous to the Grow algorithm of Collins and co-workers,<sup>6</sup> i.e., additional Shepard points were added at locations with large differences between the interpolated and accurate energies in a trial trajectory with the previous value of  $N$ . These differences were obtained by monitoring a trajectory by calculating an accurate energy every tenth integration step. A large number of trajectories, about 100–200 per surface, were then run on each interpolated potential energy surface. Statistics regarding conservation of energy and angular momentum were gathered for all trajectories but only the reactive trajectories, which are more relevant, were taken into account to generate statistics for mean unsigned errors (MUEs) in the potential.

### III.E. Properties of the stationary points

Since we placed the electronic structure Hessians at the reaction saddle point and at the two other relevant structures mentioned above at their exact location at the target level, the properties of these structures, such as geometries, energies, and vibrational frequencies, are precisely the same as at the DFT level.

### III.F. Software

MPWB1K/6-31+G( $d,p$ ) energies, gradients, and Hessians were obtained with the GAUSSIAN (Ref. 30) code, and all MM calculations were performed with the modified TINKER (Ref. 31) code, which is part of the MC-TINKER-2008–2

(Ref. 32) code. The latter was used to obtain MCMM energies and gradients. Semiclassical dynamics calculations were performed with the ANT (Ref. 33) code.

#### IV. RESULTS AND DISCUSSION

First, we will discuss the improvements of the MCMM scheme that made the full dynamics calculations possible. In the present implementation, we introduced a scaling function  $Y_{ki}$  [Eq. (31)], which is used to calculate weights of Eq. (29) for Shepard interpolation. The scaling was necessary to avoid the large values of the coupling term at geometries far from the interaction region where  $V_{12}^S$  is supposed to be negligible. In such regions, one could either scale the weighting function based on the geometry or based on the energy. Because the first option would be system dependent, we have written  $Y_{ki}$  as a function of the energy difference, in particular, the difference of  $V'_{12}(\mathbf{r}, k, i)$  at the current geometry  $\mathbf{r}$  and at a Shepard point  $(k, i)$ . If the difference is small,  $Y_{ki}$  is close to unity (no scaling), and if the difference is large, the contribution of the  $V'_{12}(\mathbf{r}, k, i)$  at a point  $\mathbf{r}$  is scaled down. This ensures the correct behavior of the interpolated surface at large internuclear separation between the fragments, including the reaction asymptotic regions. Using such a function also eliminates the need of placing electronic structure Shepard points in the far out regions even if our goal is the global potential energy surface valid at both large and small intermolecular separations.

Another improvement consists in replacing the old (exponential)  $u$  function with the function given in Eq. (28). The  $u$  function is used to avoid imaginary values of  $V_{12}$ . We did not get acceptable results for trajectories propagated on interpolated surfaces with  $N \geq 1$  with the old  $u$  function given in Eq. 19 of Ref. 9. The new  $u$  function given in Eqs. (28) and (A15) performed better.

These improvements (along with correcting an error in gradients introduced in the original MCMM formulation<sup>9</sup>) lead to the very stable results and to good conservation of energy and angular momentum in trajectory calculations on MCMM surfaces with various numbers of Shepard points and/or different choices of the force fields. For example, the total energy for 113 trajectories based on the MCMM(31)/p2 potential energy surface was conserved on the average to  $3.8 \times 10^{-4}$  kcal/mol, and the angular momentum for the same trajectories was conserved on the average to  $1.8 \times 10^{-7} \hbar$ . Similar results are obtained for all other MCMM potential energy surfaces that we have considered.

In the present work, we judge the accuracy of the interpolated potential energy surface solely on the basis of energies (not gradients or Hessians). Table III gives the MUEs of MCMM surfaces constructed with various molecular mechanics parameters and with various numbers of electronic structure Shepard points. An unsigned error is defined as an absolute value of the difference between interpolated and accurate energies. These errors are averaged over a large number of nuclear configurations sampled by semiclassical trajectories, and are considered to be reasonably well “converged” with respect to the number of nuclear configurations. The sampled nuclear configurations are spaced by

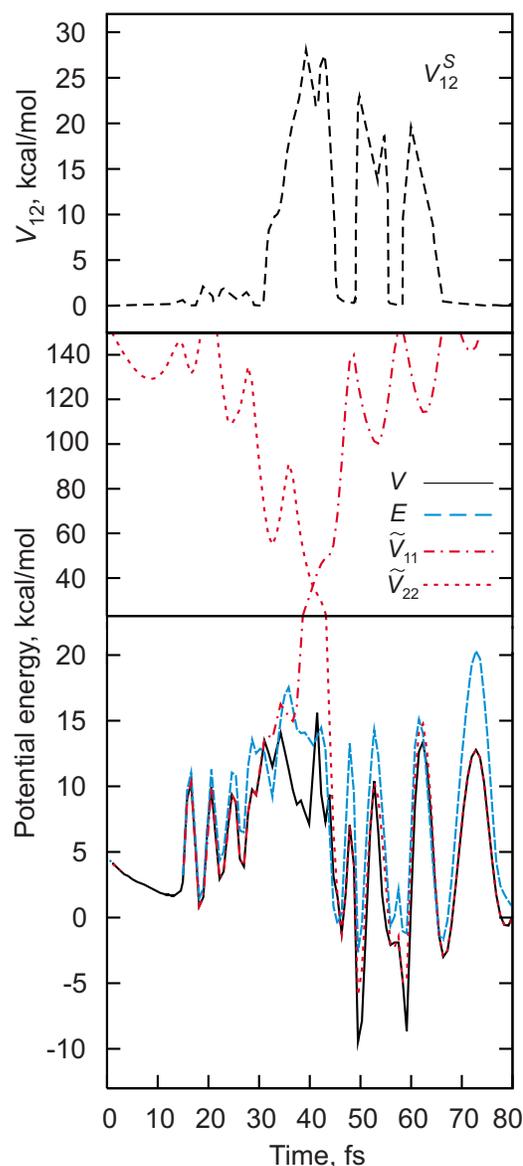


FIG. 1. (Color online) MCMM energy  $V$  [solid (black) line], diagonal elements  $\tilde{V}_{11}$  and  $\tilde{V}_{22}$  of matrix  $\mathbf{V}$  [dashed-dotted-dashed (red) line and short dashed (red) line, respectively], and accurate energy  $E$  [long dashed (blue) line] monitored along a representative trajectory on the MCMM(13) potential energy surface constructed using MM parameters p1. The upper panel shows the corresponding coupling term  $V_{12}$  [long dashed (black) line].

every ten integration steps. The subsequent discussion focuses on the analysis of the sources of errors and discusses a possible strategy for constructing an optimal global MCMM potential.

First, we consider the MCMM(3)/p2 surface in the last row of Table III. (Similar results could be obtained with  $N = 1$ .) Although the MUE is higher than for the other surfaces considered, it is interesting that even with so few Hessians we can obtain a *qualitatively* correct surface.

Figures 1 and 2 show interpolated ( $V$ ) and accurate ( $E$ ) potential energies along with the  $\tilde{V}_{11}$ ,  $\tilde{V}_{22}$ , and  $V_{12}^S$  matrix elements as functions of time for a representative trajectory on two different MCMM(13) potential energy surfaces constructed with set p1 and set p2 of molecular mechanics parameters [we will denote these potential energy surfaces as

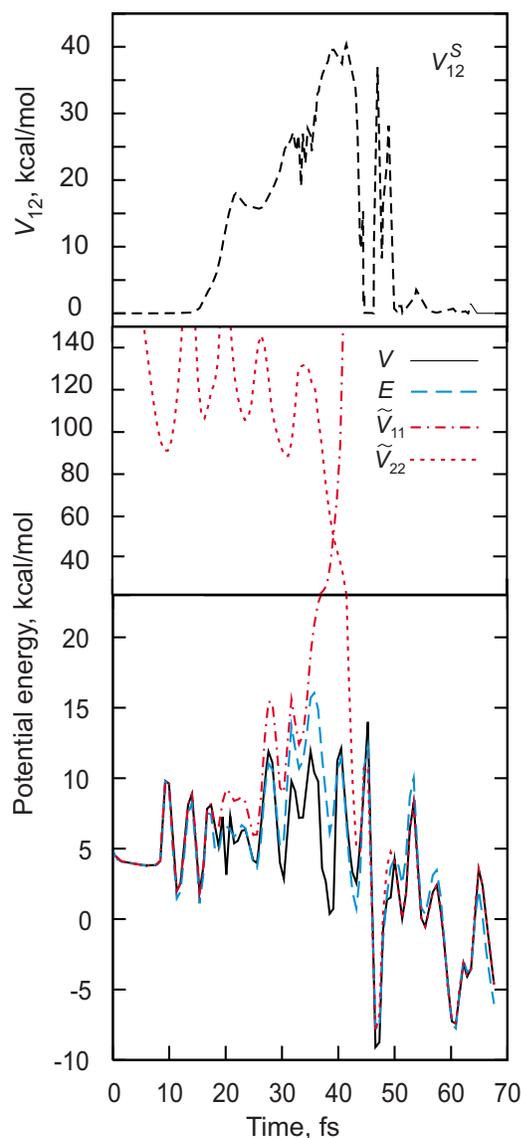


FIG. 2. (Color online) Same as Fig. 1 but for a trajectory propagated on a MCMM(13) potential energy surface constructed using MM parameters p2.

MCMM(13)/p1 and MCMM(13)/p2, respectively]. At large separations of the reactants (or products)  $V_{12}^S$  is essentially zero; therefore, the MCMM eigenvalue reduces to the diagonal matrix element and the corresponding curves merge. In the interaction region, the molecular mechanics and MCMM curves can be seen separately due to nonzero values of  $V_{12}^S$ .

Both MCMM(13)/p1 and MCMM(13)/p2 are constructed with the same set of the electronic structure Shepard points; the only differences in the generation of these potentials are the molecular mechanics parameters that underlie  $\tilde{V}_{11}$ ,  $V_{12}^S$ , and  $\tilde{V}_{22}$ . The first and third rows of Table III show that the MUE for MCMM(13)/p2 is nearly a factor of 2 smaller than the MUE for MCMM(13)/p1. The curves shown in Figs. 1 and 2 provide insight as to why MCMM(13)/p2 is considerably more accurate. As one can see from Fig. 1, the energy of the dominant molecular mechanics configuration and, consequentially, the MCMM energy, are often below the accurate energy (along a representative trajectory); this is especially noticeable on the product side. This implies that

$V_{12}(\mathbf{r}, k, i)$  becomes imaginary at all these geometries; therefore one cannot further improve the fit at these and nearby geometries by adding more electronic structure Shepard points. At nuclear configurations sampled by a trajectory propagated on the MCMM(13)/p2 potential energy surface (Fig. 2), the dominant MM term can also be seen to sometimes lie below the target energy, but at fewer locations than with parameter set 1; this results in a better fit. The better molecular mechanics potentials in the latter case are due both to better Morse/bending parameters and to different van der Waals parameters. We now consider these issues in detail. Set p1 of molecular mechanics force constants is primarily derived from the experiment, whereas set p2 is approximately corresponds to the DFT calculations. The latter overestimate the experimental harmonic vibrational frequencies in the present case, which is the typical situation. As a consequence, using the experimentally derived force fields to fit a DFT potential leads to an underestimation of the interpolated energy and to imaginary values of  $V_{12}$  in certain regions of the potential energy surface. In the present implementation, the imaginary values of  $V_{12}$  are replaced with zero  $V'_{12}$  according to Eq. (27), and this reduces the accuracy of the resulting potential energy surface. An alternative choice of parameters that yields vibrational frequencies closer to those obtained with the same method as used for the target potential energy surface will lead to a more successful fit. The results in the second row of Table III indicate that switching from set p1 of Morse/bending parameters to the corresponding set p2 reduces the MUE by approximately 30%.

The other important contributor to the MUEs collected in Table III is related to the description of the van der Waals energy. This type of interaction is generally harder to describe with an analytical function as compared to the one-dimensional bond stretching or angle bending potentials.<sup>34</sup> As discussed in Sec. III.C, we used an exp-6-12 functional form parametrized in two different ways. In set p1, the  $A$ ,  $B$ , and  $C$  parameters are the same as in the MM3 force field, and  $D$ ,  $r_m$ , and  $\epsilon$  are roughly adjusted to make the contours of the interpolated potential energy near the minimum energy path for a MCMM(3) potential energy surface smooth.<sup>19</sup> (Note that this set of parameters was not fitted to any data, but only approximately adjusted to demonstrate the symmetry properties of a MCMM fit, which was the goal of Ref. 19 in which they were originally presented.) This set of parameters yields a reasonably accurate global fit in the low energy range; however, it results in considerable underestimation of the energy at geometries corresponding to the repulsive van der Waals walls. At such geometries, the MM energy with the p1 parameters becomes lower than the accurate energy, and thus one cannot further improve the MCMM energy by adding more Shepard points. If one wished to reduce an error in this fit while keeping the  $A$ ,  $B$ , and  $C$  coefficients of Eq. (40) as they are in the MM3 force field,<sup>24</sup> one would have to readjust individual atomic parameters. Using an alternative set of the (pairwise) van der Waals parameters (set p2) obtained as described in Sec. III.C reduces the MUE from 2.3 to 1.7 kcal/mol. Even though these parameters are still far from being optimal (they are fitted to DFT potential energy curves that in addition to the van der Waals energy include

valence interactions), one can see that one obtains a better-quality MCMM fit by including a better description of the van der Waals energy.

The results shown in Table III demonstrate that, starting with MCMM(13)/p2, the MUE goes down from 1.7 to 1.3 kcal/mol as the number of electronic structure Shepard points increases from 13 to 20. The reason why the MUE does not decrease further when  $N$  is increased further is that the molecular mechanics with p2 still underestimates the energy at geometries far from equilibrium; these geometries primarily correspond to the repulsive van der Waals walls and to significantly stretched bonds in water in the product valley (as one can see from Fig. 2, the MM potential used in set p2 for the product configuration is below the accurate energy at some geometries). Ideally, one would have a MM energy that always exceeds the accurate energy, but the difference should not be large; the latter restriction is imposed by using the scaling function of Eq. (31) for the weights in Eq. (29). This scaling function is used to reduce the contribution of  $V_{12}(\mathbf{r}, k, i)$  at a geometry that is far from the geometry  $(k, i)$ . If one is only interested in the potential energy surface in a restricted region, e.g., near the minimum energy path (for example, in VTST calculations<sup>10</sup>), one need not scale the normalized weights as in Eq. (29), and one need not be particularly careful in adjusting MM parameters. Instead, one could set these parameters so that the MM energy is higher than the accurate energy at all or most geometries, and one could get a good MCMM fit by placing an electronic structure Shepard point with an unscaled weight so that it “corrects” an interpolated energy in a wider region of nuclear configurations. In contrast, scaling the weights does ensure the correct asymptotic behavior of the interpolated surface and the correct behavior of the surface everywhere at large internuclear separations of reactants and products, but it “restricts” the effect of an electronic structure Shepard point to a more localized region and thus is most effective when appropriate adjustments are made to the MM force field.

Figure 3 illustrates the  $\tilde{V}_{11}$ ,  $\tilde{V}_{22}$ , and  $V_{12}^S$  terms along with the interpolated energy  $V$  and accurate energy  $E$  for a representative trajectory propagated on the MCMM(37)/p2 surface. The MUE in the interpolated energy for this surface (averaged over 973 nuclear configurations from reactive trajectories) is 1.4 kcal/mol, and the MUE for this particular trajectory (averaged over 176 nuclear configurations) is 1.0 kcal/mol. The MCMM(37)/p2 surface is thus more accurate as compared to the MCMM(13)/p2 surface, but the error does not go down by more than 18% due to the limitation discussed above.  $V_{12}^S$  of Eq. (18) is negative at some geometries in the product valley so that the interpolated quantity  $V_{12}^S$  of Eq. (26) is set to zero in the present MCMM algorithm [Eq. (28)]. This reduces the quality of the fit. To try to avoid this problem one could fine-tune the MM parameters, especially, the Morse and angle bending force constants for water so that the MM energy of the product configuration exceeds accurate energy in this region of nuclear configurations, but we did not do this.

It is clear that the MUEs shown in Table III depend on regions of the potential energy surfaces that we take into account, in terms of both molecular geometries and vertical

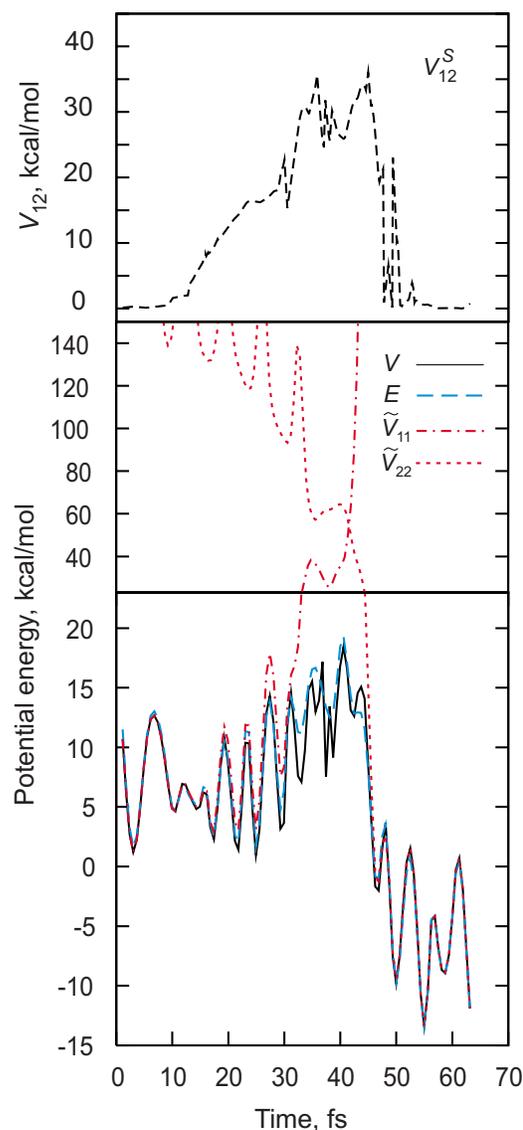


FIG. 3. (Color online) Same as Fig. 1 but for a trajectory propagated on a MCMM(37) potential energy surface constructed using MM parameters p2.

range of potential energy; these regions are completely specified by the trajectory initial conditions. We consider the dependence on molecular geometries first. Tables IV and V present MUEs for several regions of the potential energy surface defined in terms of an approximate reaction coordi-

TABLE IV. MUE and largest error (kcal/mol) for a representative trajectory on the MCMM(13)/p2 potential energy surface as a function of geometry along the approximate reaction coordinate  $z$  (the reaction coordinate  $z$  is defined as  $z = r_{\text{HH}} - r_{\text{OH}}$ , where  $r_{\text{HH}}$  is the shortest H–H distance, and  $r_{\text{OH}}$  is the second longest distance between oxygen and a hydrogen atom; negative values of  $z$  correspond to the reactants and positive values of  $z$  correspond to products; at the reaction saddle point,  $z \approx -0.5$  Å).

$z$	MUE	Largest error
-2.9 to -2.0	0.2	0.5
-2.0 to -1.0	1.1	4.1
-1.0 to 0.0	3.2	12.9
0.0 to +1.0	1.3	2.1
+1.0 to +2.3	0.9	2.3

TABLE V. MUE and largest error (kcal/mol) for a representative trajectory on the MCMM(37)/p2 potential energy surface as a function of geometry along the approximate reaction coordinate (see Table IV caption)  $z$  (Å).

$z$	MUE	Largest error
-3.3 to -2.0	0.4	0.6
-2.0 to -1.0	0.4	0.0
-1.0 to 0.0	1.9	5.2
0.0 to +1.0	0.6	1.1
+1.0 to +3.0	0.9	2.3

nate  $z$ , which is a function of the two key internuclear distances,  $r_{\text{HH}}$  and  $r_{\text{OH}}$ . Note that  $z$  equals  $-\infty$  at the reactant asymptote, and  $z$  equals  $+\infty$  at the product asymptote. Results for the MCMM(13)/p2 potential energy surface are given in Table IV, and results for MCMM(37)/p2 potential energy surface are given in Table V. The largest errors occur at geometries close (in terms of the two distances) to the reaction saddle point, which is at  $z \approx -0.5$  Å, that is, in the strong interaction region. The largest magnitudes of the MUE correspond to geometries at the repulsive walls, where a small deviation of a fitted curve from the accurate curve results in a large energetic error; these geometries are usually far from the minimum energy path. Changing the initial conditions of the trajectory, for example, by starting a trajectory at a larger separation between the reactants and/or terminating a trajectory at a larger separation of the products as compared to those distances used in the present work, would result in a smaller magnitude of the total MUE for a particular potential energy surface due to the larger contribution of the regions of the reactant and product valleys and regions beyond these valleys, where the MCMM energy is quite accurate.

Next we consider the dependence of the MUE on the magnitude of the potential energy, using as examples the MCMM(13)/p2 and MCMM(37)/p2 potential energy surfaces. The range of potential energies sampled by trajectories on these surfaces corresponds to approximately 34 kcal/mol. The MUEs for different ranges of potential energy along with the largest error for each energy range for these surfaces are shown in Tables VI and VII. For both MCMM(13)/p2 and MCMM(37)/p2, in the energy range from -13.6 to 6 kcal/mol that approximately corresponds to the reverse intrinsic barrier height plus 1 kcal/mol, the MUE is only  $\leq 1.0$  kcal/mol, that is, within chemical accuracy. The interpolated energies are more accurate in the lower energy range, as expected.

TABLE VI. MUE and largest error (kcal/mol) for MCMM(13)/p2 for different ranges of potential energy (the zero of potential energy corresponds to the reactant asymptote).

Energy range	MUE	Largest error
$< -6$	0.5	1.7
$< 3$	1.0	5.3
$< 6$	1.0	5.3
6-12	1.9	6.9
12-18	4.6	12.9

TABLE VII. MUE and largest error (kcal/mol) for MCMM(37)/p2 for different ranges of potential energy (the zero of potential energy corresponds to the reactant asymptote).

Energy range	MUE	Largest error
$< -6$	0.3	1.6
$< 3$	0.8	3.7
$< 6$	0.9	4.3
6-12	1.7	6.8
12-20	2.3	8.2

## V. CONCLUDING REMARKS

MCMM is an interpolation method for generating reactive potential energy surfaces with the aid of nonreactive molecular mechanics. Good performance of this method with a small amount of data was demonstrated for thermal rate coefficients using variational transition state theory in previous work.<sup>9,12,15</sup> In this work, we introduced some refinements to the MCMM algorithm to make the interpolated potential energy surfaces suitable for full-dynamics calculations. Although the method is designed primarily for treating more complex systems, in this work we have tested the enhanced algorithm by quasiclassical trajectory calculations for the model reaction  $\text{OH} + \text{H}_2 \rightarrow \text{H}_2\text{O} + \text{H}$ . The interpolated potential energy surfaces for this reaction were obtained with various numbers of Shepard points and with two different molecular mechanics force fields. Very stable results and good conservation of energy and angular momentum were obtained for each interpolated potential energy surface that we considered, regardless of the number and locations of the electronic structure Hessians and regardless of the nonreactive force fields employed.

In contrast to the variational transition state theory calculations that rely on the shape of potential in the vicinity of the reaction path and in the corner-cutting-tunneling region of the reaction swath, the full-dynamics calculations depend on the potential energy surface more globally. It was therefore of interest to evaluate the accuracy of the interpolated potential energies for a wide range of dynamically relevant geometries. This was accomplished by comparing these energies to the target results. The magnitude and sources of the deviations are analyzed as functions of molecular geometries sampled by trajectories and as functions of the potential energy range. For a potential energy surface constructed with 13 sparsely placed electronic structure Hessians the MUE averaged over the wide range of potential energies is only 1.7 kcal/mol, and the MUE for the lower portion of the surface, in particular, from 0 at the product asymptote to  $\sim 18$  kcal/mol, is only  $\sim 1.0$  kcal/mol.

For the trajectories we studied we found that it is difficult to converge the potential energy to better than  $\sim 1$  kcal/mol. We examined the origin of this finding, and we showed that it is primarily a consequence of the replacement of the imaginary values of  $V_{12}$  with zero, and we propose that it can be avoided in future work by allowing  $V_{12}$  to be imaginary. The errors in potential energy can also be reduced by designing a better  $Y(\mathbf{r})$  function or better force fields (parameters and/or functional forms), especially for the

description of the van der Waals interactions. In general, fitting molecular mechanics parameters is not strictly required (unless the required nonreactive molecular mechanics parameters are missing or unreliable); even if molecular mechanics parameters are optimized, which involves, as in most fitting procedures, some human “art” and judgment, such a task is still easier to carry out than fitting a *reactive* potential energy surface with analytical functions, which has only been implemented for systems with no more than about nine atoms<sup>1,35</sup> (and usually less than or equal to six). While highly accurate potentials for low-dimensional systems can be obtained with other methods (such as fitting<sup>1,35</sup> or interpolation<sup>6</sup>), the MCMM method in its present form can be applied to high-dimensional systems to get a reasonably accurate potential for a single reaction channel, possibly with one or more intermediates, with a small number of electronic structure Hessians. It has already been applied to systems<sup>9,12–14</sup> with as many as 13 atoms and is designed to make it possible to also treat bigger systems. In principle, when the spectator degrees of freedom are well fitted by nonreactive molecular mechanics, because one can transfer the molecular mechanics force fields for the nonreactive degrees of freedom from one reactive system to another, constructing a full-dimensional MCMM potential energy surface for a large system is no harder than constructing such a surface for a small system.

## ACKNOWLEDGMENTS

This work was supported in part by the NSF under Grant No. CHE07-04974.

## APPENDIX: MCMM ALGORITHM (NONSYMMETRIZED CASE)

The procedure for constructing a potential energy surface using the nonsymmetrized MCMM algorithm involves the following steps.

(i) Select  $k=1, 2, \dots, N$  molecular geometries to be used as electronic structure Shepard points, and calculate the electronic structure energies  $V^{(k)}$ , gradients  $\mathbf{G}^{(k)}$ , and Hessians  $\mathbf{F}^{(k)}$  and molecular mechanics energies  $V_n^{(k)}$ , gradients  $\mathbf{G}_n^{(k)}$ , and Hessians  $\mathbf{F}_n^{(k)}$  at these geometries for  $n=1, 2$  and  $k=1, 2, \dots, N$ .

(ii) Transform  $\mathbf{G}^{(k)}$ ,  $\mathbf{F}^{(k)}$ ,  $\mathbf{G}_n^{(k)}$ , and  $\mathbf{F}_n^{(k)}$  to the set of internal coordinates  $\mathbf{r}$  by the Wilson  $\mathbf{B}$  matrix and  $\mathbf{C}$  tensor. This yields

$$\mathbf{g}^{(k)} \equiv \frac{\partial}{\partial \mathbf{r}} V \Big|_{\mathbf{r}=\mathbf{r}(\mathbf{x}^{(k)})}, \quad (\text{A1})$$

$$\mathbf{f}^{(k)} \equiv \frac{\partial^2}{\partial \mathbf{r}^2} V \Big|_{\mathbf{r}=\mathbf{r}(\mathbf{x}^{(k)})}, \quad (\text{A2})$$

$$\mathbf{g}_n^{(k)} \equiv \frac{\partial}{\partial \mathbf{r}} V_n \Big|_{\mathbf{r}=\mathbf{r}(\mathbf{x}^{(k)})}, \quad (\text{A3})$$

$$\mathbf{f}_n^{(k)} \equiv \frac{\partial^2}{\partial \mathbf{r}^2} V_n \Big|_{\mathbf{r}=\mathbf{r}(\mathbf{x}^{(k)})}. \quad (\text{A4})$$

Throughout this paper, we use capital  $\mathbf{G}$  and  $\mathbf{F}$  to denote the gradients and Hessians with respect to Cartesian coordinates, and lower case  $\mathbf{g}$  and  $\mathbf{f}$  to denote the corresponding derivatives with respect to internal coordinates.

(iii) Define a matrix  $\mathbf{V}^{(k)}$  at each geometry ( $k$ ) by

$$\mathbf{V}^{(k)}(\mathbf{r}) = \begin{pmatrix} V_n^{(k)}(\mathbf{r}) & V_{12}(\mathbf{r}, k) \\ V_{12}(\mathbf{r}, k) & V_n^{(k)}(\mathbf{r}) \end{pmatrix} \quad (\text{A5})$$

and construct Taylor series expansions of  $V_{12}$  around each data point ( $k$ ) by

$$\begin{aligned} V_{12}(\mathbf{r}, k)^2 \approx & (V_1^{(k)} - V^{(k)})(V_2^{(k)} - V^{(k)}) + (V_2^{(k)} - V^{(k)})(\mathbf{g}_1^{(k)} \\ & - \mathbf{g}^{(k)})^T \Delta \mathbf{r}^{(k)} + (V_1^{(k)} - V^{(k)})(\mathbf{g}_2^{(k)} - \mathbf{g}^{(k)})^T \Delta \mathbf{r}^{(k)} \\ & + \frac{1}{2}(V_2^{(k)} - V^{(k)}) \Delta \mathbf{r}^{(k)T} (\mathbf{f}_1^{(k)} - \mathbf{f}^{(k)}) \Delta \mathbf{r}^{(k)} \\ & + \frac{1}{2}(V_1^{(k)} - V^{(k)}) \Delta \mathbf{r}^{(k)T} (\mathbf{f}_2^{(k)} - \mathbf{f}^{(k)}) \Delta \mathbf{r}^{(k)} \\ & + [(\mathbf{g}_1^{(k)} - \mathbf{g}^{(k)})^T \Delta \mathbf{r}^{(k)} (\mathbf{g}_2^{(k)} - \mathbf{g}^{(k)})^T \Delta \mathbf{r}^{(k)}], \end{aligned} \quad (\text{A6})$$

where

$$\Delta \mathbf{r}^{(k)} = \mathbf{r}(\mathbf{x}) - \mathbf{r}(\mathbf{x}^{(k)}). \quad (\text{A7})$$

This step uses the Taylor series reversion<sup>20</sup> of  $V_{12}^2$ .

(iv) For each geometry ( $k$ ) calculate Taylor coefficients  $D^{(k)}$ ,  $\mathbf{b}^{(k)}$ , and  $\mathbf{C}^{(k)}$ :

$$D^{(k)} = (V_1^{(k)} - V^{(k)})(V_2^{(k)} - V^{(k)}), \quad (\text{A8})$$

$$\mathbf{b}^{(k)} = \frac{\mathbf{g}_1^{(k)} - \mathbf{g}^{(k)}}{V_1^{(k)} - V^{(k)}} + \frac{\mathbf{g}_2^{(k)} - \mathbf{g}^{(k)}}{V_2^{(k)} - V^{(k)}}, \quad (\text{A9})$$

$$\begin{aligned} \mathbf{C}^{(k)} = & (1/D^{(k)})(\mathbf{g}_1^{(k)} - \mathbf{g}^{(k)})(\mathbf{g}_2^{(k)} - \mathbf{g}^{(k)})^T + (\mathbf{g}_2^{(k)} - \mathbf{g}^{(k)})(\mathbf{g}_1^{(k)} \\ & - \mathbf{g}^{(k)})^T + \frac{\mathbf{f}_1^{(k)} - \mathbf{f}^{(k)}}{V_1^{(k)} - V^{(k)}} + \frac{\mathbf{f}_2^{(k)} - \mathbf{f}^{(k)}}{V_2^{(k)} - V^{(k)}}. \end{aligned} \quad (\text{A10})$$

With these coefficients, Eq. (A6) can be rewritten as

$$[V_{12}(\mathbf{r}, k)]^2 = D^{(k)} \left( 1 + \mathbf{b}^{(k)T} \Delta \mathbf{r}^{(k)} + \frac{1}{2} \Delta \mathbf{r}^{(k)T} \mathbf{C}^{(k)} \Delta \mathbf{r}^{(k)} \right). \quad (\text{A11})$$

Note that steps (i)–(iv) are performed once at the beginning. Then steps (v) and (vi) are carried out every time that the dynamics algorithm needs the energy, gradient, and/or Hessian.

(v) Define matrix  $\mathbf{V}$  at the input geometry  $\mathbf{x}$  by

$$\mathbf{V}(\mathbf{x}) = \begin{pmatrix} V_n(\mathbf{x}) & V_{12}^S(\mathbf{x}) \\ V_{12}^S(\mathbf{x}) & V_n(\mathbf{x}) \end{pmatrix}. \quad (\text{A12})$$

The lowest-energy eigenvalue of this matrix is the MCMM PEF. The diagonal matrix elements  $V_n(\mathbf{x})$  and their deriva

tives  $\mathbf{G}_n(\mathbf{x})$  and  $\mathbf{F}_n(\mathbf{x})$  are defined by molecular mechanics. The off-diagonal matrix elements  $V_{12}$  are obtained via Shepard interpolation in internal coordinates  $\mathbf{r}(\mathbf{x})$  as follows:

$$V_{12}^S(\mathbf{r}) = \sum_{k=1}^N w_k(\mathbf{r}) V'_{12}(\mathbf{r}, k), \quad (\text{A13})$$

where  $w_k$  are *normalized weights*, and  $V'_{12}$  is defined by

$$V'_{12}(\mathbf{r}, k) = \sqrt{V_{12}(\mathbf{r}, k)^2 u(\mathbf{r}, k)}, \quad (\text{A14})$$

where  $V_{12}(\mathbf{r}, k)^2$  is given in Eq. (A11), and

$$u(\mathbf{r}, k) = \begin{cases} \frac{1}{1 + (\Delta/V_{12}(\mathbf{r}, k))^{2n}}, & V_{12}(\mathbf{r}, k)^2 > 0 \\ 0 & \text{otherwise.} \end{cases} \quad (\text{A15})$$

The normalized weight function is

$$w_k(\mathbf{s}) = \frac{\frac{Y_k}{d_k(\mathbf{s})^4}}{\sum_{k=1}^{N+2} \frac{Y_k}{d_k(\mathbf{s})^4}}, \quad (\text{A16})$$

where  $Y_k$  can be either unity (recommended for variational transition state theory calculations) or a function like Eq. (31) of the main text (recommended for trajectory calculations); the variable  $d_k$  is the generalized distance between  $\mathbf{s}$

and  $\mathbf{s}^{(k)}$  defined as

$$d_k(\mathbf{s}) = \sqrt{\sum_{\gamma=1}^{\Gamma} (s_{\gamma} - s_{\gamma}^{(k)})^2}, \quad (\text{A17})$$

where  $\mathbf{s} \equiv \{s_1, s_2, \dots, s_{\gamma}, \dots, s_{\Gamma}\}$  is a set of internal coordinates that is generally different from the set  $\mathbf{r}$ . The current implementation only supports the following cases: (a) set  $\mathbf{s}$  is

the same as set  $\mathbf{r}$ , and (b) set  $\mathbf{s}$  is a subset of  $\mathbf{r}$ . Note that the sum in Eq. (A16) has two more points than the sum in Eq. (A13). The two extra points consist of one point in the region where  $V$  is assumed to be well approximated by  $V_1$  and another in the region where  $V$  is assumed to be well approximated by  $V_2$ . At both of these points  $V_{12}$  and  $V'_{12}$  are assumed to be zero, so these points do not occur in Eq. (A13). Thus Eq. (A13) actually corresponds to an  $(N+2)$ -point interpolation with  $N$  terms.

The first and second derivatives of  $V_{12}^S(\mathbf{r})$  of Eq. (A13) with respect to internal coordinates are

$$\mathbf{g}^S(\mathbf{r}) \equiv \frac{\partial V_{12}^S(\mathbf{r})}{\partial \mathbf{r}} = \sum_{k=1}^N \left[ \frac{\partial w_k}{\partial \mathbf{r}} V'_{12}(\mathbf{r}, k) + w_k \mathbf{g}_{12}(\mathbf{r}, k) \right], \quad (\text{A18})$$

$$\mathbf{f}^S(\mathbf{r}) \equiv \frac{\partial^2 V_{12}^S(\mathbf{r})}{\partial \mathbf{r}^2} = \sum_{k=1}^N \left( \frac{\partial^2 w_k}{\partial \mathbf{r}^2} V'_{12}(\mathbf{r}, k) + \frac{\partial w_k}{\partial \mathbf{r}} \mathbf{g}_{12}(\mathbf{r}, k)^{\mathbf{T}} + \mathbf{g}_{12}(\mathbf{r}, k) \left( \frac{\partial w_k}{\partial \mathbf{r}} \right)^{\mathbf{T}} + w_k \mathbf{f}_{12}(\mathbf{r}, k) \right), \quad (\text{A19})$$

where

$$\begin{aligned} \mathbf{g}_{12}(\mathbf{r}, k) &\equiv \frac{\partial V'_{12}(\mathbf{r}, k)}{\partial \mathbf{r}} \\ &= \frac{1}{2V'_{12}(\mathbf{r}, k)} D^{(k)}(\mathbf{b}^{(k)} + \mathbf{C}^{(k)} \Delta \mathbf{r}^{(k)}) u(\mathbf{r}, k) \\ &\quad \times \left( 1 + \left( \frac{\Delta}{V_{12}(\mathbf{r}, k)^2} \right)^n nu(\mathbf{r}, k) \right) \end{aligned} \quad (\text{A20})$$

and

$$\begin{aligned} \mathbf{f}_{12}(\mathbf{r}, k) &\equiv \frac{\partial^2 V'_{12}(\mathbf{r}, k)}{\partial \mathbf{r}^2} = \frac{1}{V'_{12}(\mathbf{r}, k)} \left( -\mathbf{g}_{12}(\mathbf{r}, k) \mathbf{g}_{12}(\mathbf{r}, k)^{\mathbf{T}} + \frac{D^{(k)2} n^2 (\mathbf{b}^{(k)} + \mathbf{C}^{(k)} \Delta \mathbf{r}^{(k)})^2 u(\mathbf{r}, k)^2}{(V_{12}(\mathbf{r}, k))^2} \times \left( \frac{\Delta}{(V_{12}(\mathbf{r}, k))^2} \right)^n \right. \\ &\quad \left. \times \left\{ nu(\mathbf{r}, k) \left( \frac{\Delta}{(V_{12}(\mathbf{r}, k))^2} \right)^n - (n-1) \right\} + \frac{1}{2} D^{(k)} \mathbf{C}^{(k)} u(\mathbf{r}, k) \left( nu(\mathbf{r}, k) \left( \frac{\Delta}{(V_{12}(\mathbf{r}, k))^2} \right)^n + 1 \right) \right), \end{aligned} \quad (\text{A21})$$

where the coefficients  $D^{(k)}$ ,  $\mathbf{b}^{(k)}$ , and  $\mathbf{C}^{(k)}$  are given in Eqs. (A8)–(A10).

The first and second derivatives with respect to  $\mathbf{r}$  of the weight function given in Eq. (16) are obtained numerically. Since all operations except for this numerical intermediate step are analytic, the final MCMM derivatives may be called semianalytical.

(vi) Find the eigenvalue  $V$  of Eq. (A12) and its derivatives in Cartesian coordinates. The lowest eigenvalue of Eq. (A12) is given by

$$V(\mathbf{x}) = \frac{1}{2} (V_1(\mathbf{x}) + V_2(\mathbf{x})) - [(V_1(\mathbf{x}) - V_2(\mathbf{x}))^2 + 4(V_{12}^S(\mathbf{x}))^2]^{1/2}, \quad (\text{A22})$$

where  $V_n$  are uninterpolated MM potentials, and  $V_{12}^S$  is the resonance integral obtained via the  $(N+2)$ -point Shepard interpolation, Eq. (A13). The gradient and Hessian components of  $V$  with respect to Cartesian coordinates are given by

$$G_i = \frac{\partial V}{\partial x_i} = \frac{1}{2} \left( G_{1i} + G_{2i} - \left( \frac{4V_{12}^S \left( \frac{\partial V_{12}^S}{\partial x_i} \right) + (V_1 - V_2)(G_{1i} - G_{2i})}{((V_1 - V_2)^2 + 4(V_{12}^S)^2)^{1/2}} \right) \right) \quad (\text{A23})$$

and

$$F_{ij} = \frac{\partial^2 V}{\partial x_i \partial x_j} = \frac{1}{2} \left( F_{1ij} + F_{2ij} + \frac{\left( 4V_{12}^S \left( \frac{\partial V_{12}^S}{\partial x_i} \right) + (V_1 - V_2)(G_{1i} - G_{2i}) \right) \left( 4V_{12}^S \left( \frac{\partial V_{12}^S}{\partial x_j} \right) + (V_1 - V_2)(G_{1j} - G_{2j}) \right)}{((V_1 - V_2)^2 + 4(V_{12}^S)^2)^{3/2}} \right. \\ \left. - \frac{4 \left( \frac{\partial V_{12}^S}{\partial x_i} \right) \left( \frac{\partial V_{12}^S}{\partial x_j} \right) + (G_{1i} - G_{2i})(G_{1j} - G_{2j})}{((V_1 - V_2)^2 + 4(V_{12}^S)^2)^{1/2}} - \frac{4 \left( \frac{\partial^2 V_{12}^S}{\partial x_i \partial x_j} \right) + (V_1 - V_2)(F_{1ij} - F_{2ij})}{((V_1 - V_2)^2 + 4(V_{12}^S)^2)^{1/2}} \right). \quad (\text{A24})$$

- <sup>1</sup>J. N. Murrell, S. Carter, S. C. Farantos, P. Huxley, and A. J. C. Varandas, *Molecular Potential Energy Functions* (Wiley, Chichester, 1984); D. M. Hirst, *Potential Energy Surfaces* (Taylor & Francis, London, 1985); D. G. Truhlar, R. Steckler, and M. S. Gordon, *Chem. Rev. (Washington, D.C.)* **87**, 217 (1987); G. C. Schatz, *Rev. Mod. Phys.* **61**, 669 (1989); T. V. Albu, J. Espinosa-Garcia, and D. G. Truhlar, *Chem. Rev. (Washington, D.C.)* **107**, 5109 (2007).
- <sup>2</sup>See, for example, D. W. Brenner, O. A. Shenderova, J. A. Harrison, S. J. Stuart, B. Ni, and S. B. Sinnott, *J. Phys.: Condens. Matter* **14**, 407 (2002); K. D. Nielsen, A. C. T. V. Duin, J. Oxgaard, W.-Q. Deng, and W. A. Goddard III, *J. Phys. Chem. A* **109**, 493 (2005).
- <sup>3</sup>For examples of early work in this field, see I. S. Y. Wang and M. Karplus, *J. Am. Chem. Soc.* **95**, 8160 (1973); A. Warshel and M. Karplus, *Chem. Phys. Lett.* **32**, 11 (1975); D. J. Malcome-Lawes, *J. Chem. Soc., Faraday Trans. 2* **71**, 1183 (1975); C. Leforestier, *J. Chem. Phys.* **68**, 4406 (1978); D. G. Truhlar, J. W. Duff, N. C. Blais, J. C. Tully, and B. C. Garrett, *ibid.* **77**, 764 (1982); R. Car and M. Parrinello, *Phys. Rev. Lett.* **55**, 2471 (1985); C. Doubleday and J. W. McIver, Jr., *J. Phys. Chem.* **92**, 4367 (1988); K. K. Baldrige, M. S. Gordon, R. Steckler, and D. G. Truhlar, *ibid.* **93**, 5107 (1989); B. C. Garrett, M. L. Kozlowski, C. F. Melius, and M. Page, *ibid.* **94**, 7096 (1990); G.-X. Qian, M. Weinert, G. W. Fernando, and J. W. Davenport, *Phys. Rev. Lett.* **64**, 1146 (1990); C. Z. Wang, C. T. Chan, and K. M. Ho, *Proc.-Electrochem. Soc.* **91**, 463 (1991); W. Andreoni, *Z. Phys. D: At., Mol. Clusters* **19**, 31 (1991); R. Wentzcovich and J. L. Martins, *Solid State Commun.* **78**, 831 (1991); A. González-Lafont, T. N. Truong, and D. G. Truhlar, *J. Phys. Chem.* **95**, 4618 (1991).
- <sup>4</sup>Y.-P. Liu, D.-h. Lu, A. González-Lafont, D. G. Truhlar, and B. C. Garrett, *J. Am. Chem. Soc.* **115**, 7806 (1993).
- <sup>5</sup>See, for example, P. Blöchl and M. Parrinello, *Phys. Rev. B* **45**, 9413 (1992); D. A. Gibson and E. A. Carter, *J. Phys. Chem.* **97**, 13429 (1993); M. E. Tuckerman and M. Parrinello, *J. Chem. Phys.* **101**, 1301 (1994); G. H. Peslherbe and W. L. Hase, *ibid.* **104**, 7882 (1996); Y.-Y. Chuang and D. G. Truhlar, *J. Phys. Chem. A* **101**, 3808 (1997); K. Bolton, W. L. Hase, and G. H. Peslherbe, in *Modern Methods for Multidimensional Dynamics Computations in Chemistry*, edited by D. L. Thompson (World Scientific, Singapore, 1998), p. 143; J. C. Corchado, E. L. Coitiño, Y. Y. Chuang, P. L. Fast, and D. G. Truhlar, *J. Phys. Chem.* **102**, 2424 (1998); A. Fernandez-Ramos, Z. Smedarchina, W. Siebrand, M. Z. Zgiersle, and M. A. Rios, *J. Am. Chem. Soc.* **121**, 6280 (1999); M. E. Tuckerman, D. A. Yarne, S. O. Samuelson, A. L. Hughes, and G. Martyna, *Comp. Phys. Comm.* **128**, 333 (2000); S. Carter, J. M. Bowman, and B. J. Braams, *Chem. Phys. Lett.* **342**, 636 (2001); E. Kelly, M. Seth, and T. Ziegler, *J. Phys. Chem. A* **108**, 2167 (2004); X. Li, J. C. Tully, H. B. Schlegel, and M. J. Frisch, *J. Chem. Phys.* **123**, 084106 (2005); J. Vandervondele, M. Krack, F. Mohamed, T. Chassaing, and J. Hutter, *Comp. Phys. Comm.* **167**, 103 (2005).
- <sup>6</sup>(a) M. J. T. Jordan, K. C. Thompson, and M. A. Collins, *J. Chem. Phys.* **102**, 5647 (1995); (b) K. C. Thompson, M. J. T. Jordan, and M. A. Collins, *ibid.* **108**, 8302 (1998); (c) R. P. Bettens and M. A. Collins, *ibid.* **111**, 816 (1999); (d) D. H. Zhang, M. A. Collins, and S.-Y. Lee, *Science* **290**, 961 (2000); (e) M. Yang, D. G. Zhang, M. A. Collins, and S.-Y. Lee, *J. Chem. Phys.* **115**, 174 (2001); (f) M. A. Collins, *Theor. Chem. Acc.* **108**, 313 (2002); (g) J. F. Castillo, M. A. Collins, F. J. Aoiz, and L. Banares, *J. Chem. Phys.* **118**, 7303 (2003).
- <sup>7</sup>F. London, *Z. Elektrochem.* **35**, 551 (1929).
- <sup>8</sup>A. Warshel, *J. Am. Chem. Soc.* **102**, 6218 (1980); A. Warshel, F. Sussman, and J.-K. Hwang, *J. Mol. Biol.* **201**, 139 (1988); A. Yadav, R. M. Jackson, J. J. Holbrook, and A. Warshel, *J. Am. Chem. Soc.* **113**, 4800 (1991); A. Warshel, *Computer Modeling of Chemical Reactions in Enzymes and Solutions* (Wiley-Interscience, New York, 1991); J. Åqvist and A. Warshel, *Chem. Rev. (Washington, D.C.)* **93**, 2523 (1993).
- <sup>9</sup>Y. Kim, J. C. Corchado, J. Villa, J. Xing, and D. G. Truhlar, *J. Chem. Phys.* **112**, 2718 (2000).
- <sup>10</sup>A. Fernandez-Ramos, B. A. Ellingson, B. C. Garrett, and D. G. Truhlar, in *Reviews in Computational Chemistry*, edited by K. B. Lipkowitz and T. R. Cundari (Wiley-VCH, Hoboken, 2007), Vol. 23, pp. 125–232.
- <sup>11</sup>A. Fernandez-Ramos, B. A. Ellingson, R. Meana-Paneda, J. M. C. Marques, and D. G. Truhlar, *Theor. Chem. Acc.* **118**, 813 (2007).
- <sup>12</sup>T. V. Albu, J. C. Corchado, and D. G. Truhlar, *J. Phys. Chem. A* **105**, 8465 (2001).
- <sup>13</sup>H. Lin, J. Pu, T. V. Albu, and D. G. Truhlar, *J. Phys. Chem. A* **108**, 4112 (2004).
- <sup>14</sup>H. Lin, Y. Zhao, O. Tishchenko, and D. G. Truhlar, *J. Chem. Theory Comput.* **2**, 1237 (2006).
- <sup>15</sup>O. Tishchenko and D. G. Truhlar, *J. Phys. Chem. A* **110**, 13530 (2006).
- <sup>16</sup>M. Higashi and D. G. Truhlar, *J. Chem. Theory Comput.* **4**, 790 (2008).
- <sup>17</sup>D. G. Truhlar and J. T. Muckerman, in *Atom-Molecule Collision Theory*, edited by R. B. Bernstein (Plenum, New York, 1979), pp. 505–566.
- <sup>18</sup>See, for example, L. M. Raff, *J. Chem. Phys.* **60**, 2220 (1974); A. Schmelzer and J. N. Murrell, *Int. J. Quantum Chem.* **18**, 287 (1985); T. C. Thompson, D. G. Truhlar, and C. A. Mead, *J. Chem. Phys.* **82**, 2392 (1985); R. Steckler, K. J. Dykema, F. B. Brown, G. C. Hancock, D. G. Truhlar, and T. Valencich, *ibid.* **87**, 7024 (1987); H. Koizumi, *Chem. Phys. Lett.* **194**, 472 (1992); M. A. Collins and D. F. Parsons, *J. Chem. Phys.* **99**, 6756 (1993); M. J. T. Jordan and R. G. Gilbert, *ibid.* **102**, 5669 (1995); T. Hollebeek, T.-S. Ho, and H. Rabitz, *Annu. Rev. Phys. Chem.* **50**, 537 (1999); J. C. Corchado, D. G. Truhlar, and J. Espinosa-Garcia, *J. Chem. Phys.* **112**, 9375 (2000); A. Brown, A. B. McCoy, B. J. Braams, Z. Jin, and J. M. Bowman, *ibid.* **121**, 4105 (2004).
- <sup>19</sup>O. Tishchenko and D. G. Truhlar, *J. Chem. Theory Comput.* **3**, 938 (2007).
- <sup>20</sup>Y. T. Chang and W. H. Miller, *J. Phys. Chem.* **94**, 5884 (1990).
- <sup>21</sup>Y. Zhao and D. G. Truhlar, *J. Phys. Chem. A* **108**, 6908 (2004).
- <sup>22</sup>W. J. Hehre, L. Radom, P. v. R. Schleyer, and J. A. Pople, *Ab Initio Molecular Orbital Theory* (Wiley, New York, 1986).
- <sup>23</sup>B. J. Lynch, P. L. Fast, M. Harris, and D. G. Truhlar, *J. Phys. Chem. A* **104**, 4811 (2000).
- <sup>24</sup>N. C. Allinger, Y. H. Yuh, and J.-H. Lii, *J. Am. Chem. Soc.* **111**, 8551 (1989); J.-H. Lii and N. L. Allinger, *ibid.* **111**, 8566 (1989); **111**, 8576 (1989).
- <sup>25</sup>A. D. MacKerell, Jr., D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorcikiewicz-Kuczera, D. Yin, and M. J. Karplus, *J. Phys. Chem. B* **102**, 3586 (1998); A. D. MacKerell, Jr., *J. Comput. Chem.* **25**, 1584 (2004).
- <sup>26</sup>*Handbook of Chemistry and Physics*, edited by D. R. Lide (CRC, New York, 1997).
- <sup>27</sup>G. Herzberg, *Molecular Spectra and Molecular Structure. I. Spectra of Diatomic Molecules* (Van Nostrand Reinhold, New York, 1950).
- <sup>28</sup>See EPAPS Document No. E-JCPA6-129-020848 for supplementary information including details of the fitting procedure of the van der Waals parameters set p2. For more information on EPAPS, see <http://www.aip.org/pubservs/epaps.html>.

- <sup>29</sup>R. Bulirsch and J. Stoer, *Numer. Math.* **8**, 1 (1966); J. Stoer and R. Bulirsch, *Introduction to Numerical Analysis* (Springer-Verlag, Berlin, 1991).
- <sup>30</sup>M. J. Frisch, G. W. Trucks, H. G. Schlegel *et al.*, GAUSSIAN 03, Revision C.01, Gaussian, Inc., Pittsburgh, PA, 1998.
- <sup>31</sup>J. W. Ponder, TINKER, Version 4.2, Washington University, St. Louis, MO, 2004.
- <sup>32</sup>O. Tishchenko, M. Higashi, T. V. Albu, J. C. Corchado, Y. Kim, J. Villà, J. Xing, H. Lin, and D. G. Truhlar, MC-TINKER-2008-2, University of Minnesota, Minneapolis, MN, 2008.
- <sup>33</sup>Z. H. Li, A. W. Jasper, D. A. Bonhommeau, R. Valero, and D. G. Truhlar, ANT 08, University of Minnesota, Minneapolis, 2008.
- <sup>34</sup>See, e.g., D. W. Schwenke and D. G. Truhlar, *Chem. Phys. Lett.* **98**, 217 (1983); F. B. Brown, D. W. Schwenke, and D. G. Truhlar, *Theor. Chim. Acta* **68**, 23 (1985).
- <sup>35</sup>A. Brown, B. J. Braams, K. Cristoffel, Z. Jin, and J. M. Bowman, *J. Chem. Phys.* **119**, 8790 (2003); A. Chakraborty, Y. Zhao, H. Lin, and D. G. Truhlar, *ibid.* **124**, 044315 (2006); A. Fernandez-Ramos, J. A. Miller, S. J. K. Klippenstein, and D. G. Truhlar, *Chem. Rev. (Washington, D.C.)* **106**, 4518 (2006); Y. Wang, S. Carter, B. J. Braams, and J. M. Bowman, *J. Chem. Phys.* **128**, 071101 (2008).