Improving the accuracy of interpolated potential energy surfaces by using an analytical zeroth-order potential function

Akio Kawano, Yin Guo, and Donald L. Thompson

Department of Chemistry, Oklahoma State University, Stillwater, OK

Albert F. Wagner
Chemistry Division, Argonne National Laboratory, Argonne, IL

Michael Minkoff
Mathematics and computer Science, Division,
Argonne National Laboratory, Argonne, IL

Argonne National Laboratory, a U.S. Department of Energy Office of Science laboratory, is operated by The University of Chicago under contract W-31-109-Eng-38.

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor The University of Chicago, nor any of their employees or officers, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of document authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Available electronically at http://www.doe.gov/bridge

Available for a processing fee to U.S. Department of Energy and its contractors, in paper, from:

U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831-0062
phone: (865) 576-8401
fax: (865) 576-5728
email: reports@adonis.osti.gov

Improving the accuracy of interpolated potential energy surfaces by using an analytical zeroth-order potential function

Akio Kawano, Yin Guo, and Donald L. Thompson

Department of Chemistry, Oklahoma State University, Stillwater, Oklahoma 74078-0444

Albert F. Wagner

Chemistry Division, Argonne National Laboratory, Argonne, Illinois 60439

Michael Minkoff

Mathematics and Computer Science Division,

Argonne National Laboratory, Argonne, Illinois 60439

(Dated: December 17, 2003)

Abstract

We present a method for improving the accuracy and efficiency of interpolation methods, in which an analytical zeroth-order potential energy surface is employed as a reference surface. To investigate and test the method, we apply it to HOOH where there exists an accurate analytical surface which we take as the "exact" surface for obtaining the energies and derivatives for fitting and assessing the accuracy. Examples are given for four-dimensional and six-dimensional surfaces interpolated by using either the modified Shepard or second-degree interpolating moving least squares (IMLS) approach, with comparisons for cases with and without using the zeroth-order potential.

I. INTRODUCTION

In spite of rapid advances in quantum electronic structure theory and computational power, generating a reliable potential energy surface (PES) entirely using *ab initio* calculations is still not feasible for most practical applications, especially for studying reaction dynamics. One alternative approach is to perform *ab initio* calculations at a set of points in the configuration space and then interpolate them using some sort of fitting scheme. Examples of such fitting methods include the modified Shepard method proposed by Ischtwan and Collins[1] and the IMLS method recently investigated by us[2,3]. However, even with the use of a fitting method, the computational cost becomes increasingly demanding with the increase in the size of the system. The number of data points needed to achieve a given accuracy increases rapidly with dimension, leading to higher cost in both *ab initio* calculations and the fitting scheme. Thus, applications of these interpolating methods to large systems are hindered by the high computational cost and hence difficulty in achieving good accuracy.

To reduce the computational cost and improve the accuracy, a dual-level approach has been proposed by Nguyen *at al*,[5] in which a correction function is defined as the difference between a high-level and a lower-level electronic structure calculations and interpolated using the modified Shepard method. The potential advantage of the approach is that a fewer number of higher-level points may be needed than that required by using the single-level approach to achieve the same level of accuracy, thus providing a cost saving means for constructing a PES using high-level *ab initio* data.

Inspired by the idea of this dual-level approach, we present here a computational scheme that has the potential to significantly reduce the computational cost and improve the accuracy of an interpolated surface. The basic idea, as illustrated in Fig. 1, is as follows. Suppose V is the potential function to be fitted. If an analytical zeroth-order potential V_0 can be easily found which is a good approximation to V, then the difference

$$\Delta V = V - V_0 \tag{1}$$

should be much easier to fit than the potential V itself and can be interpolated by using

any fitting method of choice. The total interpolated potential is then

$$V_{\text{fitted}} = V_0 + \Delta V_{\text{fitted}}. \tag{2}$$

This analytical zeroth-order potential can be a simple form such as a valence force field potential or a more sophisticated one. Clearly, introducing V_0 adds negligible computational time to the overall cost of computing an interpolated surface. In addition, since fitting ΔV may require less data points than fitting the potential V itself for a given accuracy, the method could provide cost savings in both *ab initio* calculations at the data points and the use of the interpolation method for constructing the PES. Furthermore, since *ab initio* and experimental information can be incorporated in the zeroth-order PES in regions such as near the equilibrium, around the barrier, and the asymptotic regions, the method may provide the potential energies in these chemically important regions with much greater accuracy. For regions we do not wish to fit, such as regions where the zeroth-order potential is already accurate enough or regions far away from one's interest, ΔV_{fitted} can be damped to zero and the total potential is just equal to the zeroth-order one.

To test this computational scheme and assess its accuracy, we apply it to hydrogen peroxide (HOOH), which is a prototype system for studying spectroscopic and kinetic properties of small polyatomic molecules. Since our purpose is to investigate the method, instead of performing *ab initio* calculations to generate data points and assess accuracy, we choose to use the 6D analytical PCPSDE potential developed by Kuhn *et al*[4] with one modification (detailed in Sec. III) as the "exact" potential to fit. Numerical examples are given for cases with and without using the zeroth-order potential, with ΔV or V fitted by using the modified Shepard and second-degree IMLS methods.

II. INTERPOLATING METHODS

In principle, the difference function ΔV defined in Eq. (1) can be fitted by using any interpolating method. Because the modified Shepard or second-degree IMLS interpolating methods are used here, brief descriptions of them are given. Details of the two methods are in the literature.[1–3]

A. The modified Shepard method

Consider a data set consisting of N symmetry distinct points $\{\boldsymbol{Z}^{(1)}, \boldsymbol{Z}^{(2)}, \ldots, \boldsymbol{Z}^{(N)}\}$. In the modified Shepard method, the potential function $V_{\text{fitted}}(\boldsymbol{Z})$ is represented by a weighted sum of Taylor expansion about each data point

$$V_{\text{fitted}}(\boldsymbol{Z}) = \sum_{g \in G} \sum_{i=1}^{N} w_{g \circ i}(\boldsymbol{Z}) T_{g \circ i}(\boldsymbol{Z}). \tag{3}$$

Here $G = \{g_1, g_2 \ldots, g_{n_{\text{G}}}\}$ denotes the symmetry group of the molecule, $g \circ i$ denotes the data point transformed from the ith data point by the group element g, $\boldsymbol{Z}$ is a vector of internal coordinates, $T_i(\boldsymbol{Z})$ is the Taylor expansion of the potential about each data point

$$\begin{aligned} T_i(\boldsymbol{Z}) = V(\boldsymbol{Z}^{(i)}) &+ \sum_j (Z_j - Z_j^{(i)}) \left. \frac{\partial V}{\partial Z_j} \right|_{\boldsymbol{Z}=\boldsymbol{Z}^{(i)}} \\ &+ \frac{1}{2} \sum_j \sum_k (Z_j - Z_j^{(i)})(Z_k - Z_k^{(i)}) \left. \frac{\partial^2 V}{\partial Z_j \, \partial Z_k} \right|_{\boldsymbol{Z}=\boldsymbol{Z}^{(i)}} + \ldots, \end{aligned} \tag{4}$$

and $w_i(\boldsymbol{Z})$ is the normalized weight function,

$$w_i(\boldsymbol{Z}) = \frac{v_i(\boldsymbol{Z})}{\sum_{g \in G} \sum_{j=1}^{N} v_{g \circ j}(\boldsymbol{Z})}, \tag{5}$$

where $v_i(\boldsymbol{Z})$ is the un-normalized weight function with behavior such that it has a larger value if $\boldsymbol{Z}$ is closer to $\boldsymbol{Z}^{(i)}$.

In practice, the energies and first and second derivatives at the data points are obtained from *ab initio* calculations. The expansion in Eq. (4) is confined to second order because higher than second order analytical derivatives are difficult to obtain routinely for any level of *ab initio* calculations. In this study, since we use a modified version of the 6D analytical PES from Ref. 4 as the "exact" PES, the energies and derivatives used in the Taylor expansion are computed from the "exact" PES.

B. The IMLS method

In the IMLS method, the fitted potential $V_{\text{fitted}}(\boldsymbol{Z})$ is represented by a linear combination of the basis functions:

$$V_{\text{fitted}}(\boldsymbol{Z}) = \boldsymbol{a}^T \boldsymbol{b}, \tag{6}$$

where $\boldsymbol{b}(\boldsymbol{Z}) = (b_1(\boldsymbol{Z}), b_2(\boldsymbol{Z}), \ldots, b_m(\boldsymbol{Z}))^T$ is the vector of basis functions with m being the total number of basis functions, and $\boldsymbol{a} = (a_1, a_2, \ldots, a_m)^T$ is the coefficient vector which is determined by minimizing the sum of the weighted squared-deviations

$$\begin{aligned} D[V_{\text{fitted}}(\boldsymbol{Z})] &= \sum_{g \in G} \sum_{i=1}^{N} w_{g \circ i}(\boldsymbol{Z}) [V_{\text{fitted}}(g\boldsymbol{Z}^{(i)}) - V(g\boldsymbol{Z}^{(i)})]^2 \\ &= \sum_{g \in G} \sum_{i=1}^{N} w_{g \circ i}(\boldsymbol{Z}) [V_{\text{fitted}}(g\boldsymbol{Z}^{(i)}) - V(\boldsymbol{Z}^{(i)})]^2 \end{aligned} \tag{7}$$

that is, by the conditions $\partial D / \partial a_k = 0$. This yields the following weighted normal equation

$$\boldsymbol{B}^T \cdot \boldsymbol{W} \cdot \boldsymbol{B} \cdot \boldsymbol{a} = \boldsymbol{B}^T \cdot \boldsymbol{W} \cdot \boldsymbol{V} \tag{8}$$

where $\boldsymbol{B}$ is a $(n_G\, N) \times m$ matrix

$$\boldsymbol{B} = \begin{pmatrix} b_1(g_1\boldsymbol{Z}^{(1)}) & b_2(g_1\boldsymbol{Z}^{(1)}) & \cdots & b_m(g_1\boldsymbol{Z}^{(1)}) \\ b_1(g_2\boldsymbol{Z}^{(1)}) & b_2(g_2\boldsymbol{Z}^{(1)}) & \cdots & b_m(g_2\boldsymbol{Z}^{(1)}) \\ \vdots & \vdots & \ddots & \vdots \\ b_1(g_{n_G}\boldsymbol{Z}^{(1)}) & b_2(g_{n_G}\boldsymbol{Z}^{(1)}) & \cdots & b_m(g_{n_G}\boldsymbol{Z}^{(1)}) \\ b_1(g_1\boldsymbol{Z}^{(2)}) & b_2(g_1\boldsymbol{Z}^{(2)}) & \cdots & b_m(g_1\boldsymbol{Z}^{(2)}) \\ b_1(g_2\boldsymbol{Z}^{(2)}) & b_2(g_2\boldsymbol{Z}^{(2)}) & \cdots & b_m(g_2\boldsymbol{Z}^{(2)}) \\ \vdots & \vdots & \ddots & \vdots \\ b_1(g_{n_G}\boldsymbol{Z}^{(2)}) & b_2(g_{n_G}\boldsymbol{Z}^{(2)}) & \cdots & b_m(g_{n_G}\boldsymbol{Z}^{(2)}) \\ \vdots & \vdots & \ddots & \vdots \\ b_1(g_1\boldsymbol{Z}^{(N)}) & b_2(g_1\boldsymbol{Z}^{(N)}) & \cdots & b_m(g_1\boldsymbol{Z}^{(N)}) \\ b_1(g_2\boldsymbol{Z}^{(N)}) & b_2(g_2\boldsymbol{Z}^{(N)}) & \cdots & b_m(g_2\boldsymbol{Z}^{(N)}) \\ \vdots & \vdots & \ddots & \vdots \\ b_1(g_{n_G}\boldsymbol{Z}^{(N)}) & b_2(g_{n_G}\boldsymbol{Z}^{(N)}) & \cdots & b_m(g_{n_G}\boldsymbol{Z}^{(N)}) \end{pmatrix}, \tag{9}$$

$\boldsymbol{W}$ is a diagonal matrix:

$$\begin{aligned}\boldsymbol{W} = \mathrm{diag}\big(&w_{g_1\circ 1}(\boldsymbol{Z}), w_{g_2\circ 1}(\boldsymbol{Z}), \ldots, w_{g_{n_G}\circ 1}(\boldsymbol{Z}),\\ &w_{g_1\circ 2}(\boldsymbol{Z}), w_{g_2\circ 2}(\boldsymbol{Z}), \ldots, w_{g_{n_G}\circ 2}(\boldsymbol{Z}), \ldots,\\ &w_{g_1\circ N}(\boldsymbol{Z}), w_{g_2\circ N}(\boldsymbol{Z}), \ldots, w_{g_{n_G}\circ N}(\boldsymbol{Z})\big),\end{aligned} \tag{10}$$

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

$$\begin{aligned}\boldsymbol{V} = \big(&\underbrace{V(Z^{(1)}), \ldots, V(Z^{(1)})}_{n_G\text{'s}},\\ &\underbrace{V(Z^{(2)}), \ldots, V(Z^{(2)})}_{n_G\text{'s}}, \ldots,\\ &\underbrace{V(Z^{(N)}), \ldots, V(Z^{(N)})}_{n_G\text{'s}}\big)^T.\end{aligned} \tag{11}$$

For the second-degree IMLS method employed in this study, the basis functions b_k's are of the form $\prod_i Z_i^{l_i}$, where $\{l_i\}$ are non-negative integers satisfying $\sum_i l_i \leq 2$. For the 4D and 6D systems considered here, the total number of basis functions m is 15 and 28, respectively.

Once the coefficient vector $\boldsymbol{a}$ is obtained by solving Eq. (8), the interpolated energy at a given point $\boldsymbol{Z}$ is evaluated via Eq. (6). The QR factorization method was used to solve Eq. (8).

III. APPLICATION TO HOOH

For the "exact" surface to be fitted, we use the analytical PCPSDE potential developed by Kuhn *et al.*[4] with one modification. The original potential, denoted by V_{orig}, is not invariant under the permutation of the two hydrogens or two oxygens. Therefore, we employ a symmetrized function as the "exact" function:

$$\begin{aligned}V_{\mathrm{exact}}(\boldsymbol{Z}) = s(\boldsymbol{Z})&\frac{V_{\mathrm{orig}}(\boldsymbol{Z}) + V_{\mathrm{orig}}(g\boldsymbol{Z})}{2}\\ &+ [1 - s(\boldsymbol{Z})]\min[\mathrm{V}_{\mathrm{orig}}(\boldsymbol{Z}), \mathrm{V}_{\mathrm{orig}}(\mathrm{g}\boldsymbol{Z})]\end{aligned} \tag{12}$$

with

$$s(\boldsymbol{Z}) = \exp\left[-\left(\frac{V_{\text{orig}}(\boldsymbol{Z}) - V_{\text{orig}}(g\boldsymbol{Z})}{b}\right)^2\right], \tag{13}$$

where g is the permutation operator and b is a parameter determining the smoothness of the connection between the two surfaces. b is chosen as 0.5 kcal mol^{-1}. This function has the following property:

$$V_{\text{exact}}(\boldsymbol{Z}) \to \min[V_{\text{orig}}(\boldsymbol{Z}), V_{\text{orig}}(g\boldsymbol{Z})] \qquad \text{as } |V_{\text{orig}}(\boldsymbol{Z}) - V_{\text{orig}}(g\boldsymbol{Z})| \to \infty, \tag{14a}$$

$$V_{\text{exact}}(\boldsymbol{Z}) \to \frac{V_{\text{orig}}(\boldsymbol{Z}) + V_{\text{orig}}(g\boldsymbol{Z})}{2} \qquad \text{as } |V_{\text{orig}}(\boldsymbol{Z}) - V_{\text{orig}}(g\boldsymbol{Z})| \to 0. \tag{14b}$$

The accuracy of fitted potentials strongly depends on the locations of the data points. In practical applications, one should use available *ab initio* and experimental information to place some data points in the regions of interest, such as around the equilibrium and along the reaction coordinate(s). For the 6D calculations presented below, a set of 89 symmetry distinct predetermined data points were used. They were placed along the six internal coordinates $(r_{\text{OH1}}, r_{\text{OH2}}, r_{\text{OO}}, \alpha_1, \alpha_2, \tau)$, where α_1 and α_2 are the HOO angles and τ is the dihedral angle. The 89 points are determined by finding the symmetry distinct points in the following data sets

$$\{\ (r^0_{\text{OH}}, r^0_{\text{OH}}, r^0_{\text{OO}}, \alpha^0, \alpha^0, \tau^0) + (b_1\Delta_1, b_2\Delta_2, b_3\Delta_3, c_1\Delta_4, c_2\Delta_5, c_3\Delta_6) \mid \Delta_i \in \{-1, 0, 1\},\ \sum |\Delta_i| \le 2\}, \tag{15a}$$

$$\{\ (r^0_{\text{OH}}\, a_1^n,\ r^0_{\text{OH}},\ r^0_{\text{OO}},\ \alpha^0,\ \alpha^0,\ \tau^0) \mid n = -3, \ldots, 10\}, \tag{15b}$$

$$\{\ (r^0_{\text{OH}},\ r^0_{\text{OH}}\, a_1^n,\ r^0_{\text{OO}},\ \alpha^0,\ \alpha^0,\ \tau^0) \mid n = -3, \ldots, 10\}, \tag{15c}$$

$$\{\ (r^0_{\text{OH}},\ r^0_{\text{OH}},\ r^0_{\text{OO}},\ \alpha^0 + n\, a_3,\ \alpha^0,\ \tau^0) \mid n = -5, \ldots, 9\}, \tag{15d}$$

$$\{\ (r^0_{\text{OH}},\ r^0_{\text{OH}},\ r^0_{\text{OO}},\ \alpha^0,\ \alpha^0 + n\, a_3,\ \tau^0) \mid n = -5, \ldots, 9\}, \tag{15e}$$

$$\{\ (r^0_{\text{OH}},\ r^0_{\text{OH}},\ r^0_{\text{OO}}\, a_2^n,\ \alpha^0,\ \alpha^0,\ \tau^0) \mid n = -3, \ldots, 8\}, \tag{15f}$$

$$\{\ (r^0_{\text{OH}},\ r^0_{\text{OH}},\ r^0_{\text{OO}},\ \alpha^0,\ \alpha^0, n\, a_4) \mid n = 0, \ldots, 9\}, \tag{15g}$$

where the equilibrium values of the coordinates (denoted by the superscript "0") are taken from Ref. 4 and where $a_1 = a_2 = 1.1$, $a_3 = 10°$, $a_4 = 20°$, $b_1 = b_2 = b_3 = 0.05a_0$, and $c_1 = c_2 = c_3 = 3°$. Equation (15a) is a small amplitude expansion about the equilibrium configuration. Equations (15b)-(15e) explore large amplitude deviations from equilibrium in each of four coordinates separately. Equations (15f) and (15g) are highly approximate representations of the two reaction paths on the HOOH surface below 100 kcal/mol, namely the O-O bond fission (Eq. (15f)) and the hindered rotation that connects the two rotomers of HOOH (Eq. (15g)). Equation (15) is arbitrarily designed to fit the whole surface for energies below a given limit. Obviously, more intelligent chioces could be found with some effort to achieve greater accuracy in fitting. Since our main purpose here was to demonstrate the benefits of employing a zeroth-order reference surface, we did not make great effort on the initial point selection. More efficient fitting would be possible if the point selection is tailored to a particular application of the PES. If there is a specific application of interest, one should place initial data points in the regions of greatest influence on that application, such as low-energy regions for computing low-level bound states or vicinity of the reaction path for obtaining rate constants.

A. The zeroth-order PES

The zeroth-order PES employed for HOOH is a valence force field potential with the bending motions described by trigonometric functions to better describe the regions where the bend angles are far away from their equilibrium values and also to properly account for the periodicity. It has the following form:

$$\begin{aligned} V_0(r_{\mathrm{OH1}}, r_{\mathrm{OH2}}, r_{\mathrm{OO}}, \alpha_1, \alpha_2, \tau) = {} & V_{\mathrm{OH}}(r_{\mathrm{OH1}}) + V_{\mathrm{OH}}(r_{\mathrm{OH2}}) + V_{\mathrm{OO}}(r_{\mathrm{OO}}) \\ & + S_1(r_{\mathrm{OO}})\,[V_{\mathrm{bend}}(\alpha_1) + V_{\mathrm{bend}}(\alpha_2)] \\ & + S_2(r_{\mathrm{OO}}, \alpha_1, \alpha_2)\, V_{\mathrm{dihedral}}(\tau), \end{aligned} \tag{16}$$

where

$$V_{\mathrm{OH}}(r_{\mathrm{OH}}) = C_1\{1 - \exp[-C_2(r_{\mathrm{OH}} - r^0_{\mathrm{OH}})]\}^2, \tag{17}$$

$$V_{\mathrm{OO}}(r_{\mathrm{OO}}) = C_3\{1 - \exp[-C_4(r_{\mathrm{OO}} - r^0_{\mathrm{OO}})]\}^2, \tag{18}$$

$$V_{\mathrm{bend}}(\alpha) = C_5[\cos(\alpha) - \cos(\alpha^0)]^3 + C_6[\cos(\alpha) - \cos(\alpha^0)]^2, \tag{19}$$

$$V_{\mathrm{dihedral}}(\tau) = C_7[\cos(\tau) - \cos(\tau^0)]^3 + C_8[\cos(\tau) - \cos(\tau^0)]^2, \tag{20}$$

$$S_1(r_{\mathrm{OO}}) = \exp[-C_9\,(r_{\mathrm{OO}} - r^0_{\mathrm{OO}})], \tag{21}$$

$$S_2(r_{\mathrm{OO}}, \alpha_1, \alpha_2) = \frac{\{1 - \tanh[C_{10}(r_{\mathrm{OO}} - C_{11})]\}\sin^2(\alpha_1)\sin^2(\alpha_2)}{\{1 - \tanh[C_{10}(r^0_{\mathrm{OO}} - C_{11})]\}\sin^2(\alpha^0)\sin^2(\alpha^0)}. \tag{22}$$

The switching function S_1 is employed to gradually reduce the bending terms as the O-O bond distance increases and S_2 is employed to gradually reduce the dihedral term as the O-O bond distance increases or the bending angles approach 0 or π.

In practice, the constants $\{C_i\}$ are usually determined by *ab initio* and experimental information on dissociation energies and the quatities (geometries, energies, and frequencies) at the equilibrium and transition state structures. For HOOH, the data points given by Eq. (15) and a set of 100 randomly selected points are used. Except C_7 and C_8, the $\{C_i\}$ are determined by least-squares fitting of the zeroth-order potential to the "exact" one; C_1 and C_2 are determined by least-squares fitting at the points given by Eq. (15b), C_3 and C_4 by Eq. (15f), C_5 and C_6 by Eq. (15d), and C_9, C_{10}, and C_{11} on randomly selected 100 points. C_7 is determined by the condition $V_{\mathrm{dihedral}}(0^\circ) = V_{\mathrm{exact}}(r^0_{\mathrm{OH}}, r^0_{\mathrm{OH}}, r^0_{\mathrm{OO}}, \alpha^0, \alpha^0, 0^\circ)$ and C_8 by $V_{\mathrm{dihedral}}(180^\circ) = V_{\mathrm{exact}}(r^0_{\mathrm{OH}}, r^0_{\mathrm{OH}}, r^0_{\mathrm{OO}}, \alpha^0, \alpha^0, 180^\circ)$. The values of the $\{C_i\}$ are: $C_1 = 106.9$ kcal mol $^{-1}$, $C_2 = 1.257\ a_0^{-1}$, $C_3 = 55.54$ kcal mol $^{-1}$, $C_4 = 1.261\ a_0^{-1}$, $C_5 = 7.117$ kcal mol $^{-1}$, $C_6 = 86.62$ kcal mol $^{-1}$, $C_7 = 0.5895$ kcal mol $^{-1}$, $C_8 = 3.628$ kcal mol $^{-1}$, $C_9 = 0.82\ a_0^{-1}$, $C_{10} = 0.28\ a_0^{-1}$, and $C_{11} = 2.5\ a_0$.

In later sections of this paper, the global accuracy of the fitted surface with respect to the number of *ab initio* points used in the fit will be displayed. In every case where a contrast is drawn in fitting accuracy between using V_0 or not, the cost of the additional 100 *ab initio* points used to construct this V_0 will not be explicitly included in the discussion. As these results will show, the *ab initio* cost of determining V_0 is essentially inconsequential. To the degree that this V_0 is representative of the many ways of determining V_0 functions, the

general conclusion will stand that the cost of predetermining a V_0 is marginal in the overall fitting process.

The zeroth-order PES given by the above equations is also not invariant under the permutation of the two hydrogens or two oxygens. Thus, the same symmetrization technique used to generate the "exact" PES is employed here.

Representative contour plots showing the comparison of the zeroth-order PES (solid curves) with the "exact" one (dotted curves) are given in Fig. 2. The potential is plotted as a function of (a) the OH and OO bond lengths and (b) the HOO bend angle and OO bond length. The other coordinates are fixed at their equilibrium values. The interval of the contour lines is 10 kcal/mol and the maximum energy is 100 kcal/mol. It is seen that the zeroth-order PES is a poor approximation to the "exact" one as a function of the HOO bending angles. This is because the bending angles are highly coupled to the other coordinates, thus their contributions to the PES cannot be well represented by a simple form like Eq. (16) which does not have proper coupling terms between the coordinates. Therefore, to demonstrate how much improvement can be obtained in the fitted PES if a good zeroth-order PES is available, we first carried out calculations for a 4D system by fixing the two HOO bending angles at their equilibrium values. We then performed calculations for the full 6D system.

B. Four-dimensional calculations

For the 4D calculations, a set of 53 symmetry distinct predetermined data points were selected the same way as in the 6D case described in the preceding section. Their locations are determined from the data sets of Eqs. (15b), (15c) and (15f) along the three bond lengths, Eq. (15g) along the dihedral angle with $n = 1$ to 8, and Eq. (15a) with the additional restriction that $\Delta_4 = \Delta_5 = 0$.

The rest of the data points were selected randomly. The two OH distances, the OO distance, and the torsional angle were chosen from four independent random numbers distributing uniformly between $1.0a_0$ to $5.0a_0$, $1.5a_0$ to $6.0a_0$, and $5°$ to $175°$, respectively. The planar or near-planar geometries were excluded to avoid a singularity problem[8–10] associated

with the angular coordinate Z_4 defined below in Eq. (23d). If the "exact" potential energy at the point is lower than the selected upper limit of energy E_{upper}, then the point was accepted and included in the data set. Since a complete random sampling will yield a distribution of points concentrated in the high-energy regions, we chose to pick the randomly selected $N - 53$ data points from the following two sets: $2(N - 53)/3$ data points were sampled with $E_{\text{upper}} = 100$ kcal mol^{-1}, and $(N - 53)/3$ data points with $E_{\text{upper}} = 50$ kcal mol^{-1}. The second ensemble was used to enhance the density of the data point distribution in the low-energy regions.

The internal coordinate $\boldsymbol{Z} = (Z_1, Z_2, Z_3, Z_4)$ for the 4D system is chosen as

$$Z_1 = \frac{1}{r_{\text{OH1}}/a_0}, \tag{23a}$$

$$Z_2 = \frac{1}{r_{\text{OH2}}/a_0}, \tag{23b}$$

$$Z_3 = \frac{1}{r_{\text{OO}}/a_0}, \tag{23c}$$

$$Z_4 = \cos(\tau). \tag{23d}$$

The un-normalized weight function is chosen similarly as in the work of Collins and coworkers[1,6]

$$v_i(\boldsymbol{Z}) = \frac{1}{(\|\boldsymbol{\rho} - \boldsymbol{\rho}^{(i)}\|^2 + \epsilon^2)^p}, \tag{24}$$

where $\epsilon = 0.001$ for both modified Shepard and IMLS methods, and $\boldsymbol{\rho} = (r_{\text{OH1}}/a_0,\ r_{\text{OH2}}/a_0,\ r_{\text{OO}}/a_0,\ b\ \cos(\tau))$. The fitting accuracy, especially for the modified Shepard, is quite sensitive to the coefficient b. After a few tests, $b = 0.5$ is seen to yield the least rms errors in fitting and is thus used. The exponent p is chosen as 7 and 4 for the modified Shepard and IMLS methods, respectively. Our tests show that the rms errors of the fitted energies using IMLS are weakly dependent of p for $p \geq 4$ for both 4D and 6D systems considered here, thus $p = 4$ is used in both cases. Our tests also show that the rms errors are insensitive to p for $p = 9 - 12$ for the 6D system using the modified Shepard, thus $p = 9$ is used for the 6D system. Since the smallest p one can use for the modified Shepard is linearly proportional to the number of degrees of freedom, $p = 7$ is used for the 4D system.

We have done some tests using the bond lengths $\{R_i\}$, instead of the inverse bond lengths, as the first three coordinates Z_1, Z_2 and Z_3 for fitting. Tests were performed for using either the bond lengths or inverse bond lengths in the expression for the weight function Eq. (24), and the results are all worse; the rms errors are increased by about a factor of two to five for the Modified Shepard and 10% to a factor of four for the IMLS. This is consistent with the conclusion of Collins and coworkers that using the inverse bond lengths as the fitting coordinates yields more accurate fits (for example, see Ref. 6). We have also done tests using Eq. (23) as the fitting coordinates but the inverse bond lengths in the weight function, and contrary to the conclusion in Ref. 6, the results are worse; the rms errors are increased by about a factor of 2 for the modified Shepard and 10% to a factor of two for the IMLS.

Table I shows comparisons of the root-mean-square (rms) errors of the fitted energies obtained with and without using the zeroth-order PES for the IMLS and modified Shepard methods. The rms errors were obtained using ten independent ensembles. In each ensemble a new set of $N - 53$ randomly selected data points was used for fitting, and a new set of 3000 randomly sampled points, with 2000 points sampled with $E_{\text{upper}} = 100$ kcal mol^{-1} and 1000 points with $E_{\text{upper}} = 50$ kcal mol^{-1}, was used for computing the rms error. Thus the results should reflect the fitting quality of the whole surface with $E < 100$ kcal/mol for a given N. The same sets of points were used for calculations using the modified Shepard and the IMLS methods. The average values and the rms deviations of the ten ensembles are listed in Table I. It is clear that the rms errors of energy when the zeroth-order PES is used are smaller than those computed without using the zeroth-order PES. Depending on the number of data points used in fitting, the rms errors are reduced by 7% to 41% for the modified Shepard method and 40% to 68% for the IMLS method. Generally, the improvement gradually decreases with increasing N. This is likely because as direct fitting of the potential gets more accurate with increasing N, using a zeroth-order potential becomes less effective.

To illustrate the fitting quality with respect to energy, we plot in Fig. 3 the rms errors as a function of energy obtained with (filled diamonds) and without (squares) employing the zeroth-order PES. The data were taken from the same sets of calculations with the number

of data points $N = 113$ as listed in Table I. The calculations were performed by binning the energies at an interval of 5 kcal/mol. The results in panels (a) and (b) are for the modified Shepard and IMLS methods, respectively. Clearly, using the zeroth-order PES provides more accurate fits for the entire energy range considered here. Again, the improvement is more significant for the IMLS method.

Representative contour plots showing the comparison of the fitted 4D surface (solid curves) with the "exact" one (dotted curve) are given in Fig. 4 for the modified Shepard and Fig. 5 for the IMLS scheme. The potentials are plotted as a function of the OH and OO bond lengths with the other OH bond length fixed at $2.0a_0$ and the torsional angle at 150°. The results in panels (a) and (b) are for the fits without and with using the zeroth-order PES, respectively. The intervals of the contour lines is 10 kcal/mol. The number of unique data points used for fitting is 113. The "+" signs in the plots represent the projection of all data points whose OO and OH bond lengths fall within the range of the figures. This includes those connected by symmetry to the unique points. The rms errors for the surface sections shown in Figs. 4 and 5 with and without using the zeroth-order PES are 0.621 kcal/mol and 1.76 kcal/mol for the IMLS and 0.311 kcal/mol and 0.551 kcal/mol for the modified Shepard, respectively, which are smaller than the results for $N = 113$ listed in Table I.

C. Six-dimensional calculations

A set of 89 symmetry distinct predetermined data points given by Eq. (15) was used for the 6D calculations. The rest of the data points were sampled by using the efficient microcanonical sampling (EMS) method[7] for $E = 100$ kcal mol $^{-1}$ with no restriction to the total angular momentum. All the atoms were moved in the Cartesian coordinates in each Markov step with a step size of $0.5a_0$, which gives an acceptance/rejection ratio of approximately unity. To reduce the correlation of the sampled points, one point was picked from the Markov sequence in every one hundred steps. The OO distance was restricted to $r_{\mathrm{OO}} < 6a_0$ during the walk.

As in the 4D case, planar or near-planar geometries were excluded in selecting the data points to avoid a singularity problem.[8–10] The "distortion" technique proposed by Yonehara

et al.[10] was used to move the location of the data points in near-singular regions.

Similar as in the 4D case, the reciprocal interatomic distances, $\{Z_k = 1/R_k\}$, are used as the fitting coordinates, and the un-normalized weight function is chosen as

$$v_i(\boldsymbol{Z}) = \frac{1}{(\|\boldsymbol{R} - \boldsymbol{R}^{(i)}\|^2 + \epsilon^2)^p}, \tag{25}$$

where $\boldsymbol{R}$ is a vector of interatomic distances, p is 9 for the modified Shepard and 4 for the IMLS, and $\epsilon = 0.001a_0$ for both methods.

We have also performed same tests as in the 4D case using the bond lengths as the fitting coordinates or inverse bond lengths in the expression for the weight function, the results are either similar or worse.

We list in Table II the rms errors for energy with and without using the zeroth-order PES for the IMLS and modified Shepard methods. The results were obtained using ten independent ensembles. As in the 4D calculations, each ensemble consists of a new set of data points for fitting and a new set of 5000 points selected by using the EMS method for calculating the rms error. The average values and the corresponding rms deviations are given in Table II. Again, using the zeroth-order PES gives better results for both the modified Shepard and IMLS methods, and the improvement is more significant for the IMLS method. The rms errors are reduced by approximately 20% for the modified Shepard and 50% for the IMLS. The improvement is not as significant as in the 4D case for the same level of accuracy because the 6D zeroth-order PES employed here is a more crude representation of the real potential.

Figure 6 shows the rms errors as a function of energy obtained from 6D calculations with (filled diamonds) and without (squares) employing the zeroth-order PES. The number of unique data points used is 889. As in the 4D case, the calculations were performed by binning energies at intervals of 5 kcal/mol. The results shown in panels (a) and (b) are for the modified Shepard and IMLS methods, respectively. Again, using the zeroth-order PES gives more accurate fits for the entire energy range for both methods, and the improvement is greater for the IMLS.

Representative contour plots showing the comparison of the fitted 6D surface (solid curves) with the "exact" one (dotted curves) are given in Fig. 7 for the modified Shep-

ard and Fig. 8 for the IMLS. The potentials are plotted as functions of the OO bond length and one of the HOO bend angles with the other coordinates fixed at their equilibrium values. The interval of the contour lines is again 10 kcal/mol. The number of unique data points is 889. The "+" signs in the plots are the projection of all data points whose OO and OH bond lengths fall within the range of the figures. This includes those connected by symmetry to the unique data points. The results shown in panels (a) and (b) are for the fits without and with using the zeroth-order PES, respectively. The rms errors for the surface sections shown in Figs. 7 and 8 with and without using the zeroth-order PES are 2.35 kcal/mol and 4.12 kcal/mol for the IMLS and 3.36 kcal/mol and 3.46 kcal/mol for the modified Shepard, respectively, which are comparable to the results listed in Table II.

Table III shows the rms errors for the IMLS and modified Shepard methods in the vicinity of paths along the O-O bond and the torsional coordinate with the other coordinates fixed at their equilibrium values. These paths approximately correspond to the minimum energy paths for the O-O bond breaking and the hindered torsional rotation, respectively. The number of unique data points used is 889. The rms errors were estimated using points selected as follows. We first pick a set of points from the two paths along the O-O bond and the torsional angle according to $\{r_{\rm OO} = (r^0_{\rm OO} + n\,\Delta_{\rm OO})a_0 \mid \Delta_{\rm OO} = (6 - r^0_{\rm OO})/1000,\ n = 0, 1, \ldots, 999\}$ and $\{\tau = n\,\Delta_\tau \text{ (degree)} \mid \Delta_\tau = 180/1000,\ n = 0, 1, \ldots, 999\}$. Each Cartesian coordinate of the point is then perturbed to add a Gaussian-distributed random number with the standard deviation of $0.1a_0$. A narrow tube along each path was generated this way. For calculations along the O-O bond breaking path, the maximum value of the O-O distance considered is 6 bohr, which is already near the minimum of the hydrogen bonded hydroxyl dimer. For the modified Shepard method, using the zeroth-order PES provides improvement for the region around the torsional path, but not for that close to the O-O path. On the other hand, the method provides about 50% improvement along both paths for the IMLS method. The modified Shepard works very well along the torsional path because a Taylor expansion to the second order is quite accurate in this region. Overall, the rms errors in the vicinity of the two reaction paths are smaller than the corresponding ones listed in Table II which are for the entire energy range considered. Note that the fitting quality

along the two reaction paths can be improved by placing more data points in the region. This will be addressed in detail in a subsequent work.

IV. SUMMARY AND CONCLUSIONS

We have presented a computational scheme for improving the accuracy of interpolation methods for constructing a PES. The basic idea is to employ a zeroth-order analytical potential V_0 which is a reasonably good approximation to the real potential V, and fit the difference $V - V_0$ instead of V. We have tested the method on HOOH for which an accurate analytical PES is available and used as the "exact" PES for obtaining the energies and derivatives for fitting and assessing the accuracy. Calculations were performed for two interpolating methods: the modified Shepard and second-degree IMLS.

The zeroth-order PES employed in this study is a simple form which does not contain the proper coupling terms, especially between the bending angles and the rest of the coordinates. To see the effectiveness of the method if a good zeroth-order PES is available, we first carried out 4D calculations with the two bending angles fixed at their equilibrium values. We then carried out full 6D calculations. The rms errors for energy are considerably reduced by employing the zeroth-order PES. For energies below 100 kcal/mol, depending on the number of data points used for fitting, the rms errors are reduced by 7% to 41% for the modified Shepard and 40% to 68% for the IMLS for the 4D calculations, and by approximately 20% for the modified Shepard and 50% for the IMLS for the 6D calculations. In general, the method works better for the IMLS than for the modified Shepard. This is likely due to the fact that the modified Shepard scheme also requires good quality of the first and second derivatives of the zeroth-order PES, while the IMLS explicitly uses only the energies.

The accuracy of a fitted surface depends on the selections of the coordinate system, the weight function, and the locations of the data points. The optimized choices of these quantities are often different for different fitting methods and different systems. For the calculations presented here, the accuracies are generally better for the modified Shepard for a given N. Comparison of efficiency of the two methods depends on the type of applications. If frequent calculations of the PES are not required, IMLS is generally more efficient than

modified Shepard for most electronic structure methods because differences in accuracy are not substantial but calculating the gradients and Hessians generally requires an order of magnitude more cpu time than calculating only the energy. However, if one wants to do trajectory calculations, the time required to repeatedly generate the derivatives also needs to be taken into account. The main purpose of this paper is to demonstrate the effectiveness of this technique of employing a zeroth-order function. In a subsequent work, we will provide more detailed studies on improving the accuracy of fitting, the accuracy of the trajectory calculations, and the comparison of the IMLS and modified Shepard methods.

In summary, the results presented here demonstrate the effectiveness of this approach for fitting PES's. Since a reasonably accurate analytical PES can often be written, especially in the chemically important regions, the scheme may provide a useful tool toward the goal of developing an automated procedure to construct accurate PES for practical applications.

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).

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

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

[3] G. G. Maisuradze, D. L. Thompson, A. F. Wagner, and M. Minkoff, J. Chem. Phys. **119**, 10002 (2003).

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

[5] K. A. Nguyen, I. Rossi, and D. G. Truhlar, J. Chem. Phys. **103**, 5522 (1995).

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

[7] (a) H. W. Schrantz, S. Nordholm, and G. J. Nyman, J. Chem. Phys. **94**, 1487 (1991). (b) G. Nyman, S. Nordholm, and H. W. Schrantz, J. Chem. Phys. **93**, 6767 (1990).

[8] M. J. T. Jordan, K. C. Thompson, and M. A. Collins, J. Chem. Phys. **103**, 9669 (1995).

[9] R. P. A. Bettens and M. A. Collins, J. Chem. Phys. **108**, 2424 (1998).

[10] T. Yonehara and S. Kato, J. Chem. Phys. **117**, 11131 (2002).

TABLE I: Comparisons of the rms errors for energy from the 4D calculations with and without using the zeroth-order PES for the modified Shepard and IMLS methods[a].

N^b	rms(E) (kcal mol^{-1}) fitting ΔV	rms(E) (kcal mol^{-1}) fitting V
Modified Shepard method		
83	0.750 $\pm$0.159	1.28 $\pm$0.08
113	0.545 $\pm$0.082	0.919 $\pm$0.079
173	0.376 $\pm$0.036	0.611 $\pm$0.093
293	0.246 $\pm$0.017	0.337 $\pm$0.026
533	0.157 $\pm$0.010	0.200 $\pm$0.012
1013	0.103 $\pm$0.008	0.118 $\pm$0.007
1973	0.0611 $\pm$0.0045	0.0659 $\pm$0.0042
IMLS method		
83	1.35 $\pm$0.22	4.24 $\pm$0.84
113	0.916 $\pm$0.090	2.38 $\pm$0.31
173	0.612 $\pm$0.053	1.52 $\pm$0.14
293	0.421 $\pm$0.038	0.965 $\pm$0.072
533	0.281 $\pm$0.014	0.598 $\pm$0.059
1013	0.199 $\pm$0.015	0.361 $\pm$0.034
1973	0.138 $\pm$0.013	0.229 $\pm$0.016

[a]The results are obtained using ten different ensembles and the average values and their rms deviations are listed here.

[b] N is the number of symmetry distinct data points.

TABLE II: Comparisons of the rms errors for energy from the 6D calculations with and without using the zeroth-order PES for the modified Shepard and IMLS methods[a].

N^b	rms(E) (kcal mol^{-1}) fitting ΔV	rms(E) (kcal mol^{-1}) fitting V
Modified Shepard method		
189	3.15 $\pm$ 0.52	3.71 $\pm$ 0.58
289	2.10 $\pm$ 0.32	2.54 $\pm$ 0.36
489	1.44 $\pm$ 0.18	1.88 $\pm$ 0.29
889	1.04 $\pm$ 0.13	1.31 $\pm$ 0.18
1689	0.727 $\pm$ 0.11	0.918 $\pm$ 0.061
3289	0.508 $\pm$ 0.037	0.634 $\pm$ 0.030
6489	0.372 $\pm$ 0.036	0.466 $\pm$ 0.031
IMLS method		
189	3.40 $\pm$ 0.68	6.88 $\pm$ 1.17
289	2.19 $\pm$ 0.54	4.49 $\pm$ 0.61
489	1.64 $\pm$ 0.22	3.21 $\pm$ 0.17
889	1.19 $\pm$ 0.06	2.42 $\pm$ 0.09
1689	0.897 $\pm$ 0.058	1.78 $\pm$ 0.07
3289	0.695 $\pm$ 0.047	1.43 $\pm$ 0.15
6489	0.546 $\pm$ 0.031	1.08 $\pm$ 0.05

[a]The results are obtained using ten different ensembles and the average values and their rms deviations are listed here.

[b] N is the number of symmetry distinct data points.

TABLE III: The rms errors in the vicinity of the IRC for the O-O bond fission and torsional isomerization from the 6D calculations using 889 symmetry distinct points.

	rms error of energy (kcal mol^{-1})	
	fitting ΔV	fitting V
Modified Shepard method		
Along O-O	0.697	0.664
Torsional	0.069	0.127
IMLS method		
Along O-O	0.382	0.742
Torsional	0.620	1.38

FIGURE CAPTIONS

Fig. 1: A 1D illustration of the basic idea of the fitting scheme. The solid curve V is the potential to be fit and the long-short dashed curve V_0 is the zeroth-order approximation to V. The lower dashed curve $\Delta V = V - V_0$ is the curve that is actually fit.

Fig. 2: Representative contour plots showing the comparisons of the zeroth-order PES (solid curves) with the "exact" one (dotted curves). The potentials are plotted as a function of (a) OO and OH bond lengths and (b) OO bond length and HOO bend angle. The other coordinates are fixed at their equilibrium values. The interval of the contour lines is 10 kcal/mol.

Fig. 3: The rms errors as a function of energy obtained from 4D calculations with (filled diamonds) and without (squares) employing the zeroth-order PES. The results in panels (a) and (b) are for the modified Shepard and IMLS methods, respectively.

Fig. 4: Representative Contour plots showing the comparison of the fitted 4D surface (solid curves) with the "exact" one (dotted curves) by using the Modified Shepard. The results in panels (a) and (b) are for the fits without and with using the zeroth-order PES, respectively. The "+" signs represent the projection of all the data points in the region. The interval of the contour lines is 10 kcal/mol.

Fig. 5: Same as figure 5 except for the IMLS method.

Fig. 6: The rms errors as a function of energy obtained from 6D calculations with (filled diamonds) and without (squares) employing the zeroth-order PES. The results in panels (a) and (b) are for the modified Shepard and IMLS methods, respectively.

Fig. 7: Representative Contour plots showing the comparison of the fitted 6D surface (solid curves) with the "exact" one (dotted curves) by using the Modified Shepard. The results in panels (a) and (b) are for the fits without and with using the zeroth-order PES, respectively. The "+" signs represent the projection of all the data points in the region. The interval of the contour lines is 10 kcal/mol.

Fig. 8: Same as figure 7 except for the IMLS method.

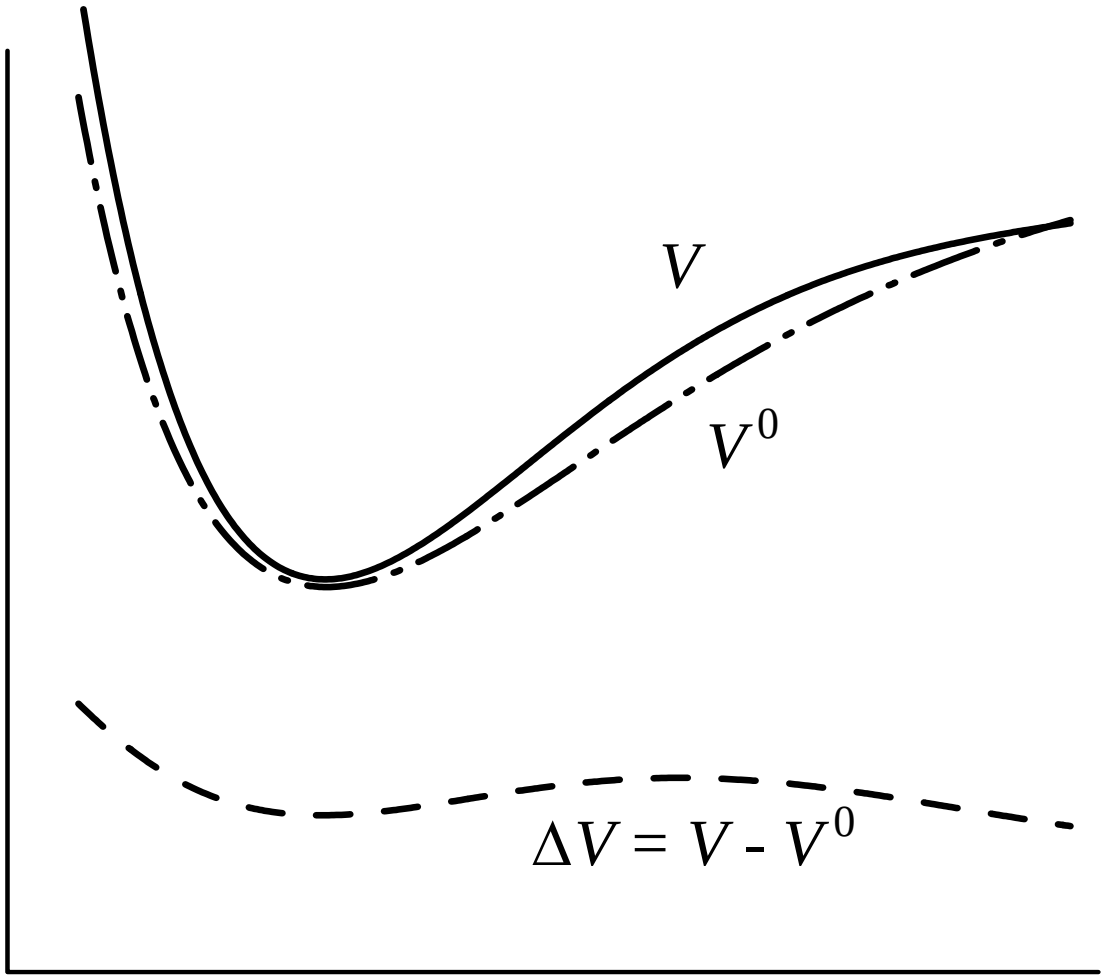

Fig. 1

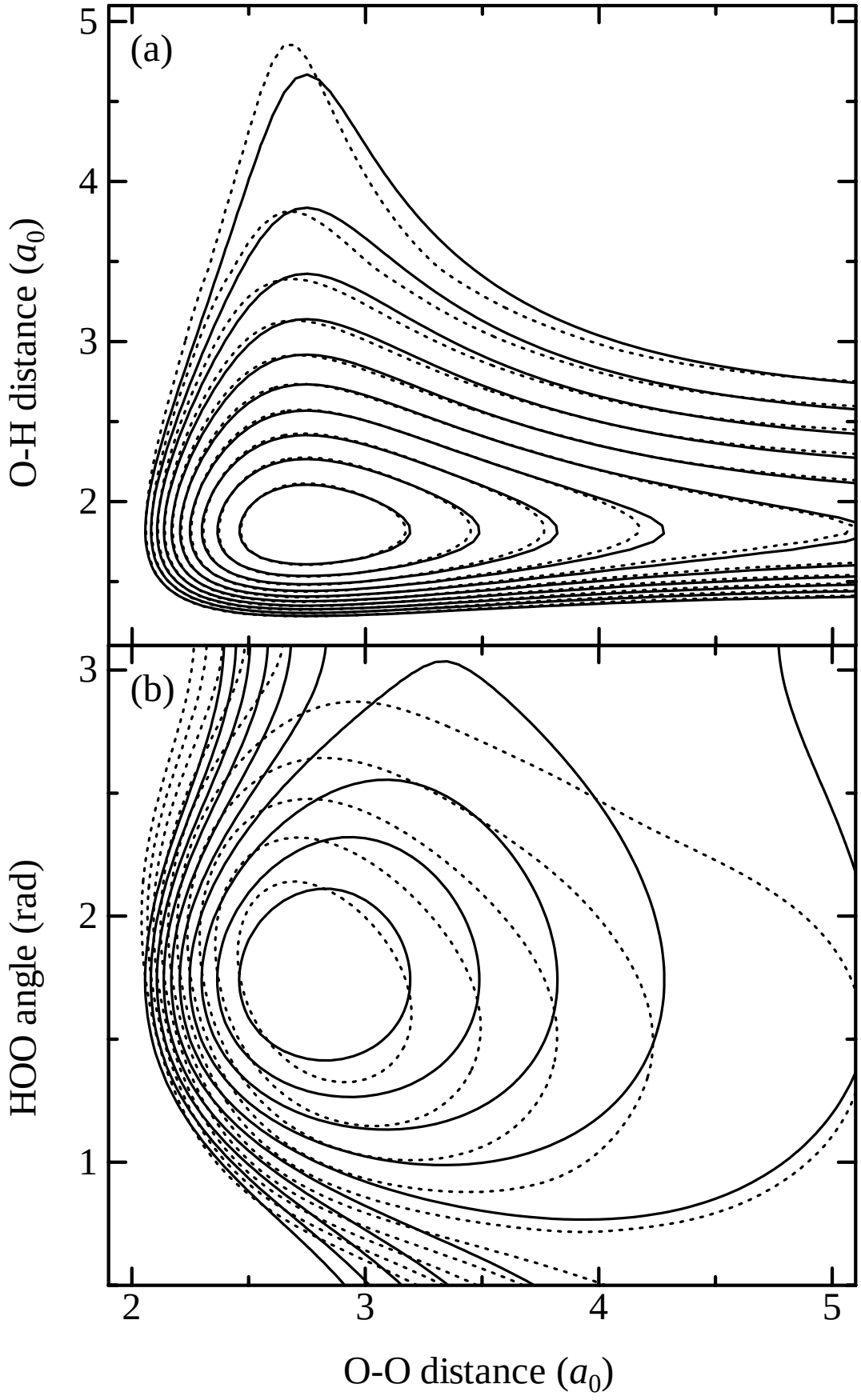

Fig. 2

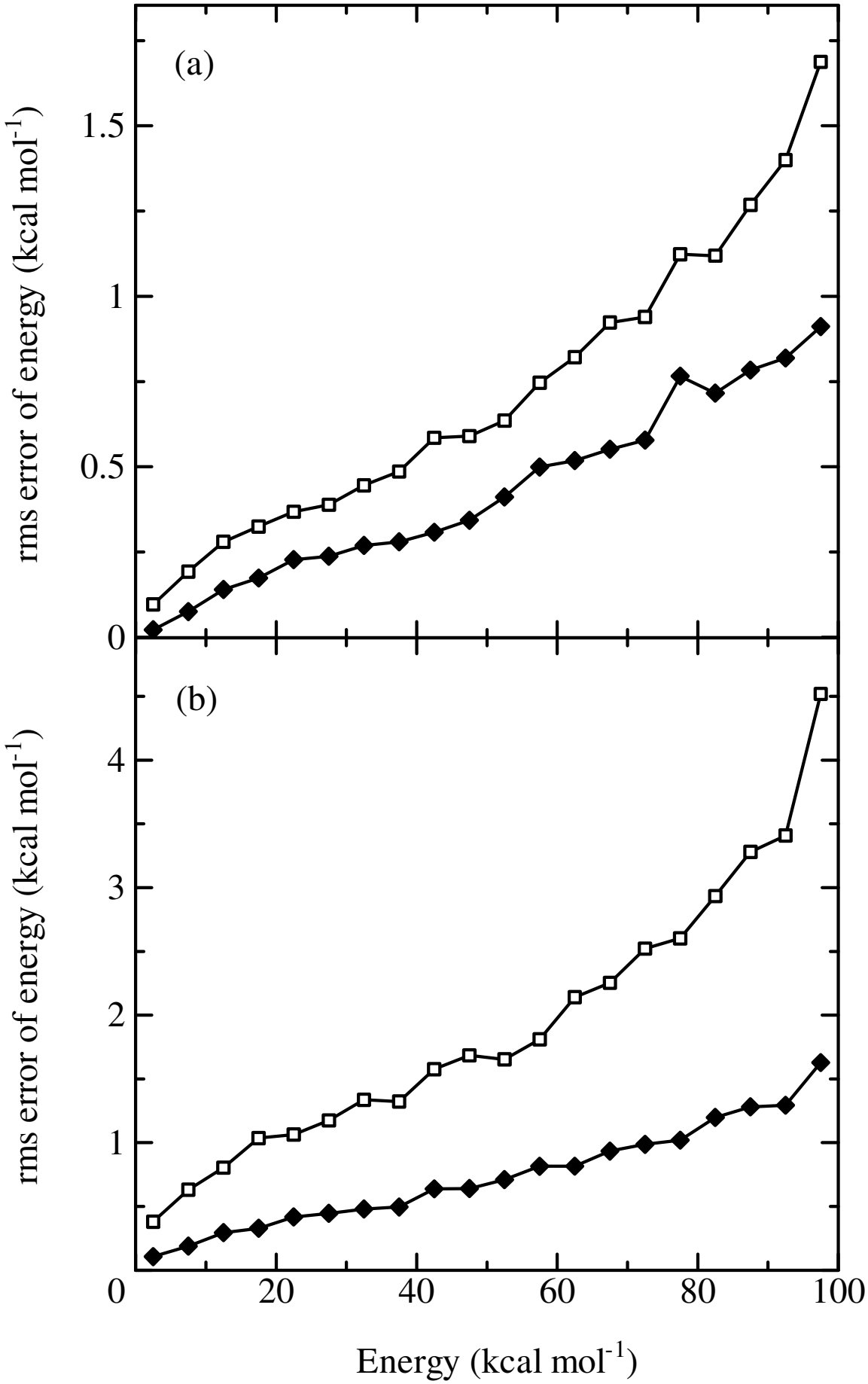
(a)
(b)
rms error of energy (kcal mol-1)
rms error of energy (kcal mol-1)
0
0.5
1
1.5
0
1
2
3
4
0
20
40
60
80
100
Energy (kcal mol-1)

Fig. 3

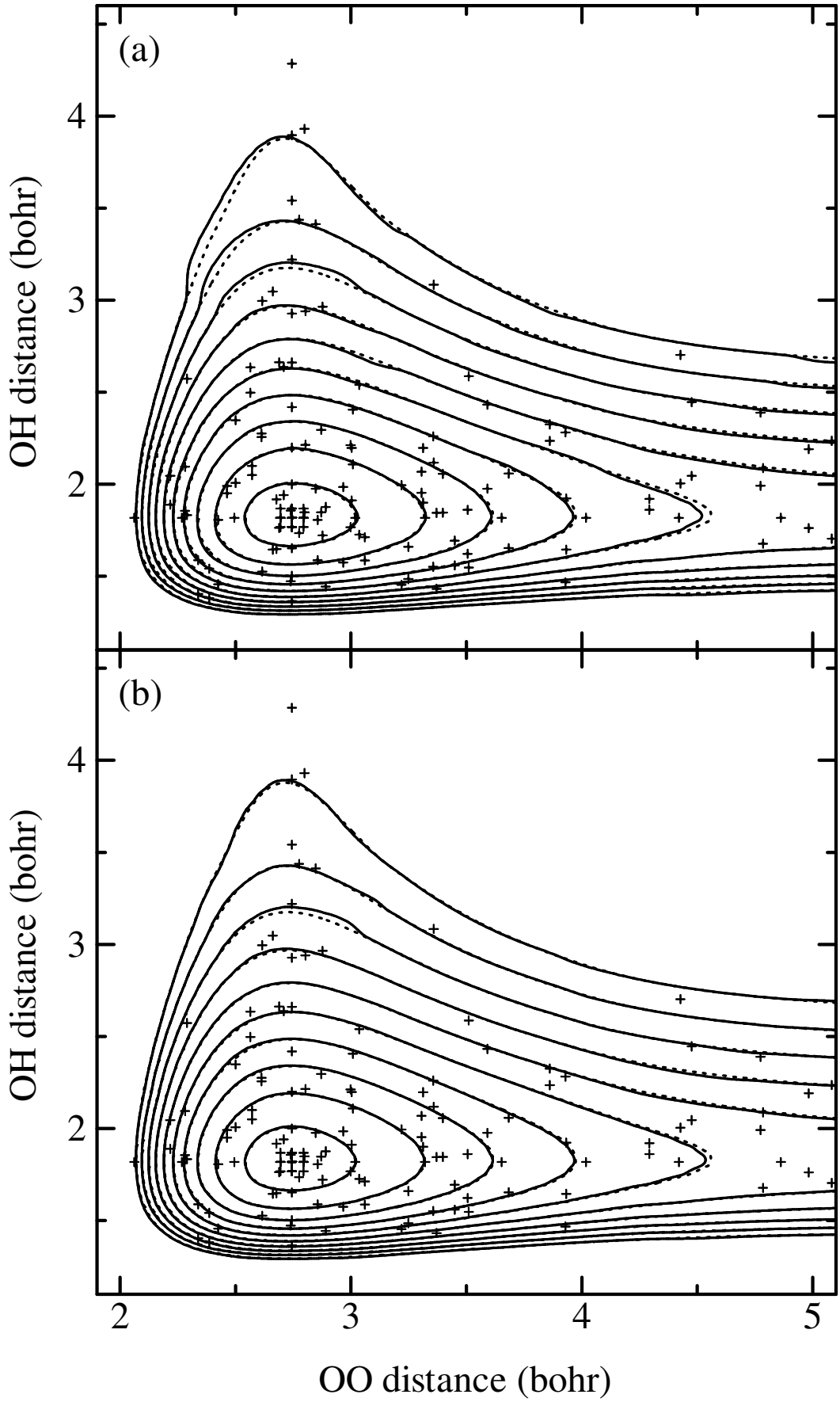

Fig. 4

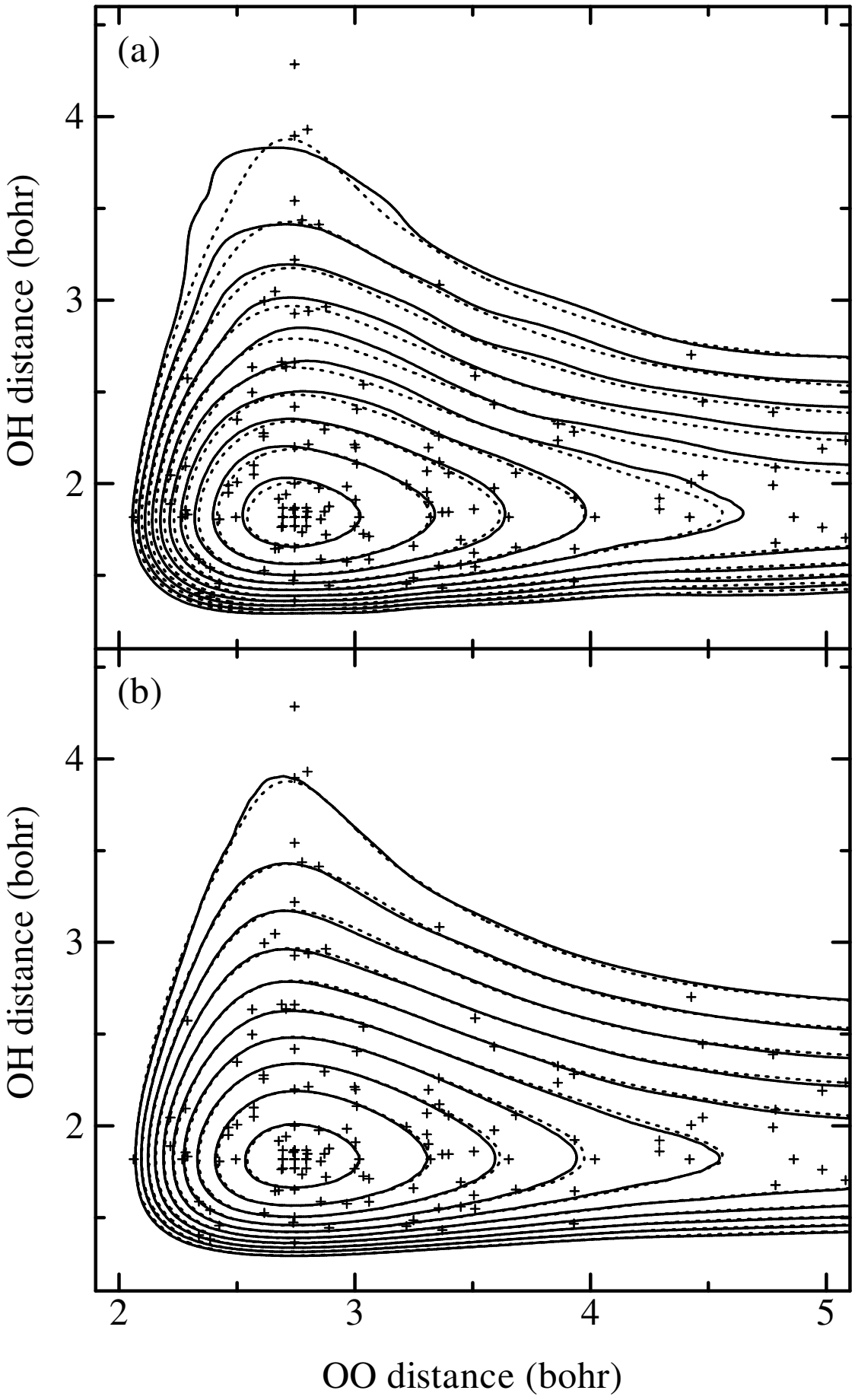
(a)
(b)
OH distance (bohr)
OH distance (bohr)
2
3
4
2
3
4
5
OO distance (bohr)

Fig. 5

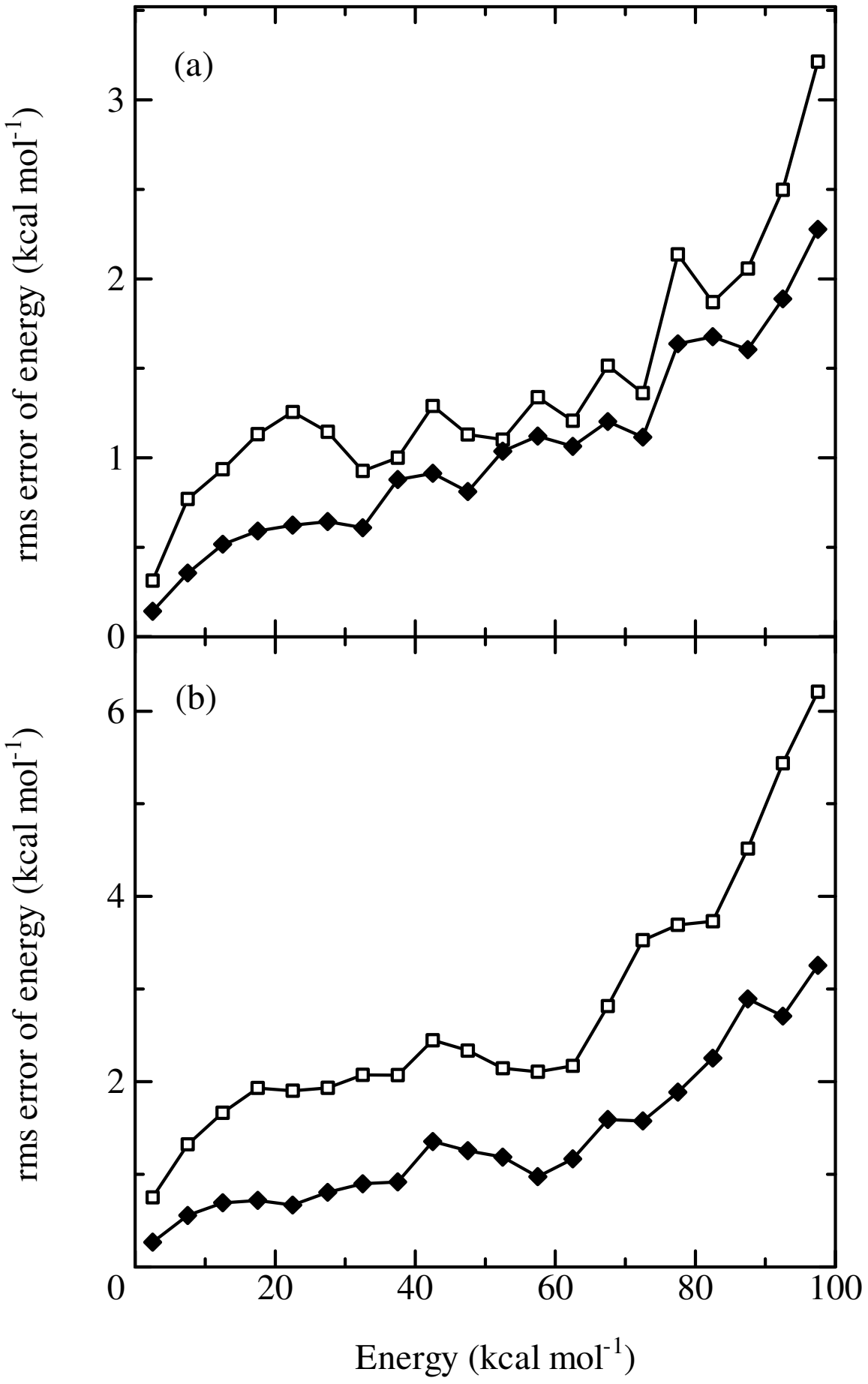
(a)
(b)
rms error of energy (kcal mol-1)
rms error of energy (kcal mol-1)
0
1
2
3
0
2
4
6
0
20
40
60
80
100
Energy (kcal mol-1)

Fig. 6

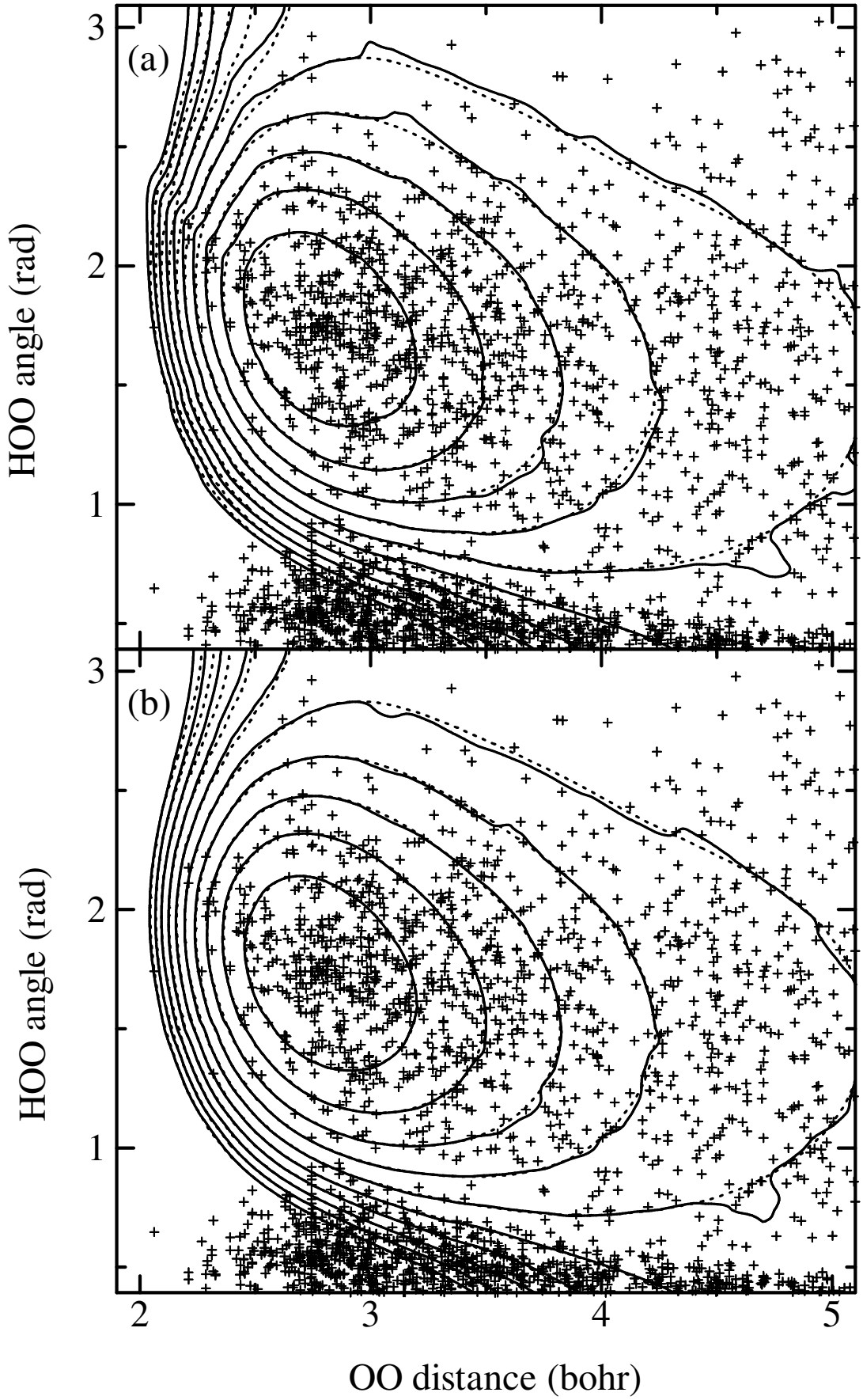

Fig. 7

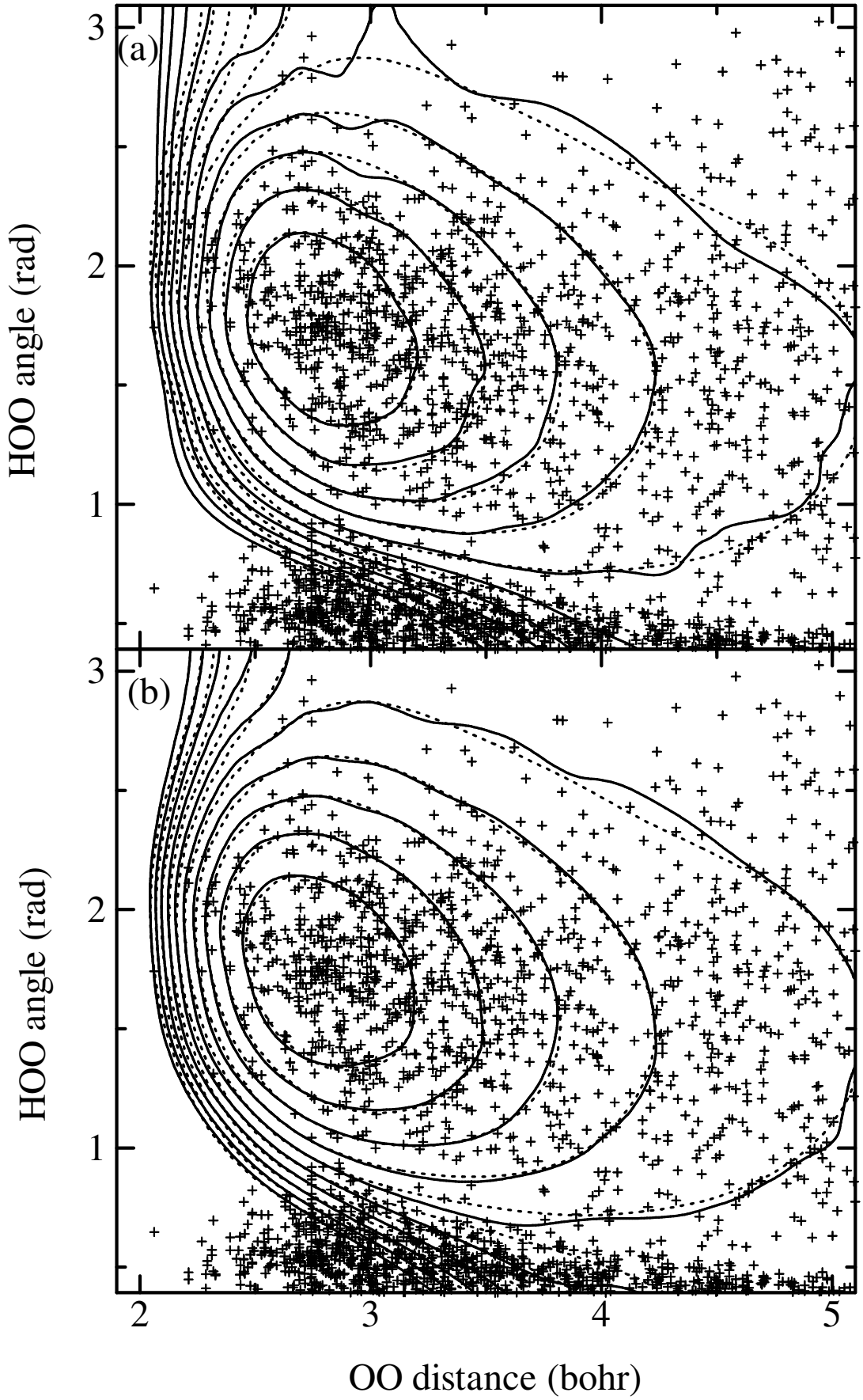

Fig. 8