Permutationally-invariant Fitting of Many-body, Non-covalent Interactions with Application to Three-body Methane-water-water

Riccardo Conte,* Chen Qu, and Joel M. Bowman*

Department of Chemistry, Emory University, Atlanta, GA 30322, United States

E-mail: riccardo.conte@emory.edu; jmbowma@emory.edu

Abstract

A modified, computationally efficient method to provide permutationally-invariant polynomial bases for molecular energy surface fitting via monomial symmetrization (Xie Z., Bowman J. M. *J. Chem. Theory Comput.* **2010**, *6*, 26-34) is reported for applications to complex systems characterized by many-body, non-covalent interactions. Two approaches, each able to ensure the asymptotic zero-interaction limit of intrinsic potentials, are presented. They are both based on the tailored selection of a subset of the polynomials of the original basis. A computationally-efficient approach exploits reduced permutational invariance and provides a compact fitting basis dependent only on intermolecular distances. We apply the original and new techniques to obtain a number of full-dimensional potentials for the intrinsic three-body methane-water-water interaction by fitting a database made of 22,592 *ab initio* energies calculated at the MP2-F12 level of theory with haTZ basis set. An investigation of the effects of permutational

*To whom correspondence should be addressed

symmetry on fitting accuracy and computational costs is reported. Several of the fitted potentials are then employed to evaluate with high accuracy the three-body contribution to the $CH_4-H_2O-H_2O$ binding energy and the three-body energy of three conformers of the $CH_4@(H_2O)_{20}$ cluster.

Keywords

Many-body Interactions; Methane-water-water; Methane Hydrate; monomial symmetrization; purified bases; non-covalent interactions.

Introduction

Calculations of molecular properties and molecular dynamics simulations rely on the availability of realistic potential energy surfaces (PESs). In the case of simple systems, like diatomic molecules, model potentials, e.g., harmonic, Morse or Lennard-Jones, have been successfully employed in many applications. For triatomic systems, numerical fitting techniques, e.g., 3-d splines, are suitable and often adopted to obtain a precise analytical potential from calculated *ab initio* energies.[1]

However, the task becomes very challenging as the dimensionality of the problem increases, since simple functions are unable to provide accurate fits. One possibility to tackle this issue is to calculate the electronic energy at the nuclear geometries of interest whenever needed during the simulation. This approach is at the heart of *ab initio* molecular dynamics and related techniques. They are all based on the calculation of electronic energies "on-the-fly", i.e. step-by-step along the trajectory evolution (see, for instance, refs. 2–9). However, to be practically feasible, these techniques are usually limited to computationally cheap levels of electronic theory associated with small basis sets. Furthermore, every time a new simulation is performed, the electronic energies must be re-calculated, and the computational overhead often becomes too expensive for long-time dynamics or rare-event investigations. A different and alternative type of approach consists in the fitting of a large number of *ab initio* energies

to analytical expressions. These energies are calculated once even at high level of theory and with large basis sets. The outcome is represented by accurate full-dimensional analytical PESs which allows for fast potential calls.

Several approaches have been undertaken to precisely represent multi-dimensional PESs in an analytical, fast-to-compute way. A viable route to fit *ab initio* energies is provided by neural networks.[10] They have been applied to high-dimensional PESs of isolated molecules and molecule-surface systems.[11–17] For instance, a recent application of neural networks has provided a high-level PES for the HOCO radical and its dissociation channels.[18] Another way to treat the fitting of multidimensional PESs lies in the n-mode representation of the potential.[19–21] The potential is written as a series of intrinsic potentials that depend on normal coordinates. A variation of it, the so-called potfit potential,[22] which has been developed for applications of the Multi-configuration Time-dependent Hartree technique, is a product of one-mode potentials. A final representation worth mentioning is the modified Shepard approach.[23,24] It is based on force fields centered at several reference geometries. These force fields are low-order series representations of the potential, dependent on the inverse of the internuclear distances, weighted and combined to represent the PES.

A drawback of all these methods (at least in their original versions) is that they do not account for permutational invariance. Invariance of the PES under permutations of identical atoms is needed to undertake successfully some dynamical applications. For instance, isomerization studies[25] or investigations of unimolecular dissociations[26–28] necessitate a permutationally-invariant surface. Early work achieved this numerically by replicating data, see e.g. ref. 25. Ideally, the mathematical description of the PES should have permutational invariance built-in, then *ab initio* energies do not have to be replicated for equivalent configurations, and the analytical potential is based on a smaller number of functions.

There has been substantial progress in fitting approaches that exploit the invariance of an electronically adiabatic PES with respect to permutation of like atoms, to obtain precise mathematical fits to ca 10^4 to 10^5 electronic energies for molecular potentials with as many

as 10 atoms and numerous minima, saddle points and fragments channels.[29–35] Examples where this has been done are nitromethane (CH_3NO_2),[30] acetaldehyde (CH_3CHO),[36] and the allyl radical (C_3H_5).[27] The two approaches that have been developed by our group are based on using a fitting basis of invariant polynomials. In the more sophisticated and efficient approach the polynomials are represented as products of so-called invariant primary and secondary polynomials.[37] The generation of these polynomials is not a trivial task, and it is accomplished with the MAGMA software.[38] An extensive library of such polynomials for as many as 10 atoms has been generated by Braams and one of the authors.[37] The second approach is a straightforward monomial symmetrization one, in which monomials are made invariant by application of all permutations of all like atoms to generate multi-term polynomials.[39,40] Software to perform this symmetrization using an efficient iterative method to obtain higher-order symmetrized monomials from lower-order ones was developed by Xie and one of the authors.[41] This software is general and user "transparent". It requires as input the permutational symmetry of the molecular system and the maximum order of the polynomials to generate. The resulting representation is equivalent to the more sophisticated and efficient one based on factored primary and secondary invariants. The monomial symmetrization approach has been combined by Guo and co-workers with the multi-level neural network approach to create a hybrid method to fit electronic energies with application to a number of 4 and 5 atom systems.[42,43]

Roughly 50 PESs have been generated using the primary and secondary invariant approach. Several variants of this approach have also been applied. These include using a single polynomial representation as well as a "many-body" one. In the most recent applications, mainly to reactive systems, the single polynomial representation has been used. As pointed out in a review of this method,[37] the single polymonial representation does not rigorously separate into non-interacting fragments. However, by incorporating non-interacting fragment data into the data set for fitting (an essential component of the procedure) the fits do numerically incorporate this separation, with of course a (small) fitting error. The lack of

rigorous separation is more easily seen in the monomial symmetrization method as will be shown in detail below. This was pointed out and then remedied by Truhlar and co-workers, following a suggestion of Xie and Bowman, in their application of monomial symmetrization using the Xie-Bowman software for N_4.[44] The remedy was to remove the small number of basis functions that do not rigorously separate and then to perform the fit with what we will refer to as a "purified" basis.

The need to describe correctly the interaction energy of separated species becomes crucial when dealing with non-covalent systems made of several molecular monomers. The potential energy of such systems can often be expressed by means of a rapidly convergent many-body representation. Thereby, the challenging problem of fitting a global full-dimensional energy surface of a high-dimensional system with a single-polynomial representation is reduced to the easier problem of fitting a number of lower-dimensional potentials. The two and higher-body terms in the representation are so-called intrinsic potentials, and have the property that a generic intrinsic p-body potential vanishes when a single monomer is separated from the other p-1 ones. This property can be built-in as in the case of two-body sum-of-pairs potentials,[45] while the issue of describing three-body (or higher) interactions in a general and easy way remains open. Analytical expressions based on a limited number of parameters are available for two-body interactions. Among them are Lennard-Jones, Buckingham exp-6, Varandas,[46,47] and Tang-Toennies[48] potentials, to name a few. Three-body potentials are more difficult to model. Analytical expressions for atomic long-range three-body interactions based on atomic multipoles have been reported.[49] Recently, the E3B model that involves explicit three-body interactions has been applied to study the water hexamer,[50] and a force field method has been employed to fit the three-body water interaction, even if on the basis of non-linear parameters and rigid monomers.[51] A many-body force field for CO_2 with explicit three-body interactions has also been reported.[52]

There have been numerous applications of permutationally-invariant fitting to non-covalent interactions, e.g., the water dimer[53] and trimer,[54] $(HCl)_2$,[55] $(HCl)_3$,[56] and mixed $HCl-H_2O$

clusters.[57] In these applications the full permutational symmetry was used, even though some of the permutations are unfeasible. This paper is focused on permutationally fitting non-covalent interactions without using the full permutational invariance, with emphasis on three-body interactions in general and considering the 11-atom complex CH_4−H_2O−H_2O as a specific and challenging application. Two approaches will be employed and both are based on what we term "purified" fitting bases. By that we mean eliminating symmetrized monomials that do not rigorously separate as fragments separate. The first approach does this by starting from the output of the monomial symmetrization software, and retaining only those polynomials with the correct fragment limit. In the second approach, encouraged by our results in previous studies of two-body Ar−HOCO[58] and CH_4−H_2O[59] interactions, the purified basis is further reduced by keeping only polynomials that depend exclusively on inter-monomer distances.

The rest of the paper is organized as follows. The next section presents the methods to obtain the permutationally invariant fitting of many-body interactions with purified bases and a discussion of the role of permutational symmetry. Several new, fully-flexible intrinsic three-body CH_4−H_2O−H_2O potentials are then obtained based on fitting roughly 23,000 *ab initio* electronic energies. Section 3 reports an analysis of accuracy and efficiency of the potentials depending on the permutational symmetry and maximum polynomial order employed, with applications of the intrinsic three-body CH_4−H_2O−H_2O potentials to the trimer and to the $CH_4@(H_2O)_{20}$ cluster. This $CH_4@(H_2O)_{20}$ cluster can be used as a model system for methane hydrates. Methane hydrates are crystalline water-based solids similar to ice, in which methane molecules are trapped inside cages formed by hydrogen-bonded water molecules.[60] Conclusions and summarizing remarks are in Section 4.

Theoretical Details

Derivation of purified invariant polynomial bases

The many-body representation of the total potential energy of a molecular cluster characterized by non-covalent interactions and made of N monomers can be expressed in a general way as

$$V^{(I,J,K,\ldots,N)}(\mathbf{r}) = \sum_{I}^{N} V^{(I)}(\mathbf{r}_I) + \sum_{I<J}^{N} V_{2b}^{(I,J)}(\mathbf{r}_I, \mathbf{r}_J) + \sum_{I<J<K}^{N} V_{3b}^{(I,J,K)}(\mathbf{r}_I, \mathbf{r}_J, \mathbf{r}_K) + \ldots + \\ + V_{Nb}^{(I,J,K,\ldots,N)}(\mathbf{r}_I, \mathbf{r}_J, \mathbf{r}_K, \ldots, \mathbf{r}_N), \tag{1}$$

where I, J, K etc. are shorthand notations for labeling the monomers, $\mathbf{r}$ is the collection of all nuclear coordinates of the cluster, and $\mathbf{r}_I$, $\mathbf{r}_J$ etc. are the sets of nuclear coordinates respectively of monomer I, J, etc. Eq. (1) applies to both homogenous clusters, where all the monomers are of the same kind, and to heterogeneous systems, made of monomers of different species. The one-body terms ($V^{(I)}$) are the potential(s) of the isolated monomers. The other higher-order terms in the representation are the so-called intrinsic potentials. The p-body intrinsic energy ($V_{pb}^{(I,\ldots,P)}$) for a given cluster geometry is equal to the difference between the total energy of the same p-body system ($V^{(I,\ldots,P)}$) and the sum of all terms in the representation up to the intrinsic (p-1)-body potentials, calculated at same geometry. Clearly, as complexes separate to monomers these intrinsic potentials go to zero.

Standard generation of permutationally-invariant fitting bases via monomial symmetrization needs the permutational group symmetry and the maximum order of polynomials as inputs. Software[41] is available for this purpose. We will refer in the following to these polynomial bases as full (F) bases, in contrast to the purified ones we describe below. F bases are used to represent the intrinsic p-body potentials, via least-squares fitting to databases of

corresponding *ab initio* energies. That is, in generic notation

$$V_{pb}^{\mathrm{F}(I,\ldots,P)} = \sum_{m=0}^{M} D_{\underline{b}}^{\mathrm{F}} \mathcal{S} \left[\prod_{i<j}^{N_{at}} y_{ij}^{b_{ij}} \right] \quad (m = \sum b_{ij}). \tag{2}$$

$D_{\underline{b}}^{\mathrm{F}}$ is a set of linear coefficients which are determined by means of a least-square fit. $\underline{b}$ stands for the ordered collection of exponents b_{ij}. N_{at} is the number of atoms in the system, and $y_{ij} = \exp(-r_{ij}/\alpha)$ are Morse functions of the inter-nuclear distances r_{ij} between atoms i and j. The α parameter is usually in the range between 2 and 3 au ($\alpha = 2$ au for the potentials presented in this work). $\mathcal{S}$ is the formal operator that symmetrizes the monomials according to the chosen permutational group. Eq. (2) is invariant for translation, rotation, and the allowed permutations of like atoms. Hereafter, we will denote permutational groups by means of a short-hand notation based on their indexes. For instance, 422111 indicates an 11-atom system with a permutational group made of a subgroup of four identical atoms, two subgroups of two identical atoms, and three single atoms. Permutation is allowed between atoms belonging to the same subgroup, but not between atoms of different subgroups. The three single atoms cannot permute at all. $CH_4{-}H_2O{-}H_2O$ is a relevant example of such a system.

Several permutational groups of different order can be employed to describe the potential surface of molecular clusters characterized by non-covalent interactions and for which inter-monomer atom-exchange does not occur. In calculations where permutationally-invariant potentials are adopted, input atomic coordinates are rigorously ordered according to the symmetry, and atoms belonging to different monomers can be easily identified. However, when two or more monomers are of the same kind, we require an additional permutational invariance with respect to this monomer interchange. This particular symmetry is of course included in full permutational groups, but it must be added to low-order groups. In the case where this is done, our notation is to label the group with an additional * symbol. For instance, in this paper we will present fitted intrinsic three-body $CH_4{-}H_2O{-}H_2O$ potentials with full (821) or partial (4421 or 422111*) permutational symmetry. Operationally,

since a modification of the existing software to include the star symmetry into low-order symmetry groups is not straightforward, we incorporate it by duplicating the database of energies upon collective permutation of atoms of like monomers. For the intrinsic three-body $CH_4-H_2O-H_2O$ potential the database needs only to be doubled. In general, for a system with n monomers of same type, the database must be replicated n! times. The effect of the replication of the database is that part of the linear coefficients on which the potential is based have identical values and can be factored. This replication of data harkens back to earlier work where this was done, e.g., the PES of C_2H_2.[25] Progress is being made to incorporate this additional symmetry and will be reported later. Figure 1 clarifies the case of the $CH_4-H_2O-H_2O$ system and 422111^* symmetry.

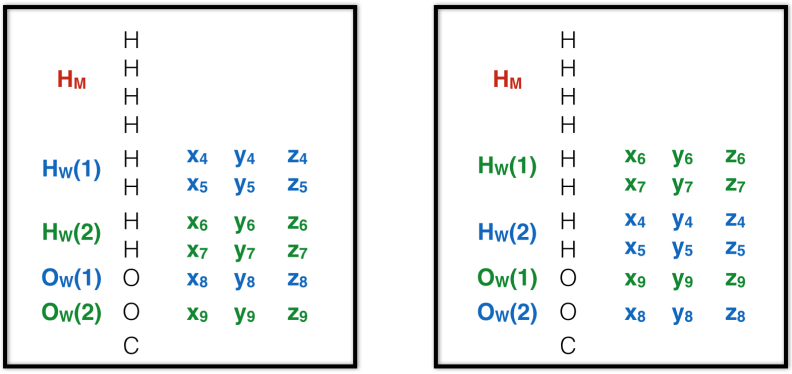


Figure 1: Input coordinates for the $CH_4-H_2O-H_2O$ system with 422111^* symmetry, as explained in the text. Atomic order follows the permutational group and atoms can be grouped according to the monomer they belong to. H_M indicates the hydrogen atoms of the methane monomer, while H_W labels the hydrogen atoms of the two water monomers. The two inputs differ for the collective order of the two water monomers (blue and green). To point out this aspect only water coordinates have been explicitly reported. For both inputs the potential must return the same value.

One fundamental property of intrinsic p-body potentials, as noted already, is that they approach zero as one of the p monomers is separated at large distance from the other p -1 ones. A fitted potential must be able to reproduce this feature. As noted previously in the literature,[37] this is not strictly enforced by F potentials since not all the polynomials in the

F basis separate to vanishing interactions in this limit. A simple way to tackle the issue is to purify the full basis as noted. This is practically accomplished by generating a set of cluster configurations where each monomer (one at a time) is set far away from the others which are kept close together. For each configuration, inter-monomer Morse variables involving atoms of the isolated monomer are set equal to zero, while the remaining Morse variables are given different, non-zero values with the care to avoid (unlikely) accidental cancellation. All monomials and polynomials of the F basis are then evaluated and those returning a value different from zero are discarded. The final outcome is a smaller polynomial basis able to ensure the correct zero-interaction limit. This procedure can be undertaken independently from the permutational symmetry employed. We will label the fitting bases and potentials obtained in this way as purified (P) bases and potentials.

The purification technique, which reduces the size of the fitting basis, can be further advanced by restricting the P basis to polynomials dependent exclusively on inter-monomer variables. Since this further reduces the size of the fitting basis, it is expected to speed up substantially potential calls. In practice, we set equal to zero only intra-monomer Morse variables and then calculate all monomials and polynomials in the P basis. Only polynomials returning a value different from zero are maintained. The final outcome of the technique is a very compact fitting basis, that we term a pruned purified (PP) basis. A generic intrinsic p-body PP potential is analytically expressed as

$$V_{pb}^{\mathrm{PP}(I,\ldots,P)} = \sum_{m=p-1}^{M} D_{\underline{b}}^{\mathrm{PP}} \, \mathcal{S}' \left[\prod_{i<j}^{N_{at}} \left(y_{ij}^{\mathrm{PP}} \right)^{b_{ij}} \right] \quad (m = \sum b_{ij}), \tag{3}$$

where $D_{\underline{b}}^{\mathrm{PP}}$ are the coefficients to fit, $\mathcal{S}'$ is the formal operator that symmetrizes the monomials and returns polynomials able to reproduce correctly the zero-interaction limit, and $\{y_{ij}^{\mathrm{PP}}\}$ is the set of Morse variables dependent on inter-monomer distances. We note that the sum in Eq. (3) starts from $m = p$-1. The reason is that polynomials of order p-2 or less are not suitable to describe the zero asymptotic limit.

Next, we apply these various fitting approaches to the demanding, eleven-atom intrinsic

three-body $CH_4-H_2O-H_2O$ (MWW) potential.

Ab initio calculations

For the applications reported in this paper, each intrinsic three-body energy was calculated from *ab initio* data as

$$V_{3b}^{\mathrm{MWW}} = V^{\mathrm{MWW}} - V^{\mathrm{M\text{-}W_1}} - V^{\mathrm{M\text{-}W_2}} - V^{\mathrm{W\text{-}W}} + V^{\mathrm{M}} + V^{\mathrm{W_1}} + V^{\mathrm{W_2}}, \tag{4}$$

where the two water monomers are labeled as W_1 and W_2 in terms where only one of the two is involved. Eq. (4) is a rearrangement of Eq. (1) to express the *ab initio* intrinsic three-body MWW energy. All electronic energies were obtained using MP2-F12/haTZ **(aug-cc-pVTZ for C and O, cc-pVTZ for H)**[61,62] theory computed with MOLPRO 2010.[63] In total, 22,592 $CH_4-H_2O-H_2O$ configurations were obtained as follows. A first set of 3,226 points was obtained. 2,000 points were chosen to cover various O-C-O angles and C-O distances; 380 points were selected from molecular dynamics simulations of $CH_4-H_2O-H_2O$ employing preliminary PES fits; 846 points were sampled from $CH_4@(H_2O)_{20}$ and $CH_4@(H_2O)_{24}$ geometries for future application of this intrinsic potential to methane clathrates. The monomers were kept almost rigid in these 3,226 points. Finally, the remaining 19,366 points were randomly sampled around the 3,226 points to cover the distortion of monomers.

Figure 2 reports the distribution of energies in the database. Most of them are in the range -100 cm^{-1} to 100 cm^{-1}. This small range in energies is mainly a consequence of the weak three-body interaction, but a reasonable fraction of energies is sampled at short monomer distances.

Results and Discussion

This database of intrinsic three-body energies was fit by 9 PESs. These are given along with a variety of performance metrics in Table 1. The notation is best explained by a couple of

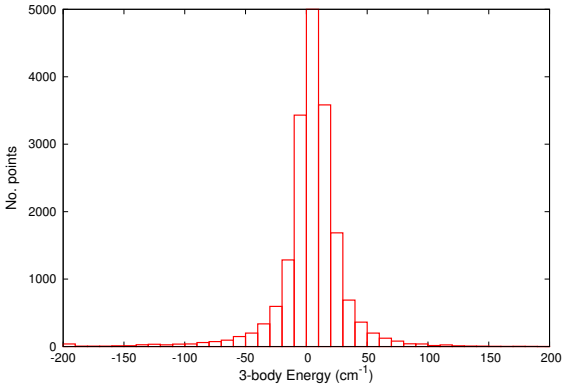


Figure 2: Distribution of the 22,592 MP2-F12/haTZ energies in the database. Bin width is 10 cm^{-1}. The most populated bin is truncated to better appreciate the population in the high-energy bins. Bins labeled as 200 and -200 contain also all populations from energies above and below those values.

examples. "821/4" indicates a full permutationally invariant basis of maximum polynomial order 4 and "4221111*/3" indicates a partial permutationally invariant basis, as described already, with maximum polynomial order, M, of 3. In addition, we label the fitted potentials according to their basis set (F, P, or PP) followed by the permutational group and maximum polynomial order.

As seen, and as expected, the number of coefficients (i.e., polynomials) increases with M for given basis type. Furthermore, and also as expected, there are fewer coefficients for PESs of higher permutational order. The purified potentials have a substantially lower number of polynomials than the corresponding full ones. The impact of this reduction increases with the maximum polynomial order within a group. Root mean square errors (rmse) are small, partially due to the fact that three-body interactions are themselves small. However, the variation of the rmse is significant. Within a given permutational group, the fit precision increases with the number of polynomials, as expected. A more complicated dependence concerns computational time. We have averaged over batches of ten repetitions the time needed by the different potentials in evaluating 50,000 potential calls. For comparison purposes, we have set equal to 100.0 the time required by PP-422111*/4. We note that within the same permutational group the computational effort grows with the number of polynomials to eval-

uate, of course. However, this is no longer true when comparing between different groups. For instance, PP-422111*/4 with 5809 terms is more than 9 times faster than F-4421/4 with 4698 terms. The reason is that monomial symmetrization requires evaluation of a much larger number of monomials for groups of higher permutational order. In other words, the fewer polynomials in the F basis with high symmetry are more expensive to evaluate because they contain more monomials. For example, F-4421/4 with 4698 terms requires calculation of about 121 K monomials against the 16 K required by PP-422111*/4 with 5809 terms. **The two potentials that fit more accurately the database are P-422111*/4 (2.2 cm^{-1}) and PP-422111*/4 (4.6 cm^{-1}). The former maintains some dependence on intramolecular variables, thus providing a better rms but at the price of around three-time slower potential calls. Both potentials keep the permutational symmetry of their group (422111*) and describe correctly the asymptotic interaction limit**.

Rmse results do not necessarily translate into accuracy of the potentials in energy calculations. To investigate this, we have employed our potentials to evaluate the intrinsic three-body contribution to the energy of the CH_4−H_2O−H_2O trimer. The trimer equilibrium configuration has been optimized at the high CCSD(T)-F12a/haDZ **(aug-cc-pVDZ for C and O, cc-pVDZ for H)**[64,65] level of theory to best evaluate the *ab initio* trimer binding energy (see below). The last row of Table 1 reports the *ab initio* intrinsic three-body energy calculated with MP2-F12/haTZ, the same level of theory employed for the database, thus allowing for a direct comparison to the results of our potentials. The last column of Table 1 presents the intrinsic three-body contribution of the potentials to the CH_4−H_2O−H_2O trimer energy, with the exception of F-821/4 due to the huge computational overhead of this potential which makes it not suitable for more complex calculations. F-821/3 and F-4421/3 are far off the *ab initio* value and will not be further considered. As for the remaining six potentials, those with maximum polynomial order M=4 provide excellent approximations, while F-422111*/3 is less accurate but still acceptable. Remarkably, the two very fast puri-

fied potentials with M=3 (P-422111*/3 and PP-422111*/3) yield much better energies than the corresponding F potential even if based on a lower number of polynomials. We will employ all six potentials in the final and more complex energy calculations before drawing our conclusions about the accuracy of the fitted potentials.

Table 1: Number of coefficients, fitting root mean square error, computational times, and intrinsic three-body energy of the trimer for different analytical CH_4−H_2O−H_2O potentials. M is the maximum polynomial order. The type of polynomial basis is full (F), purified (P), or pruned purified (PP). Time is arbitrarily set equal to 100 for PP-422111*/4. The *ab initio* value has been calculated with MP2-F12/haTZ.

Group Symmetry/ M	Basis	No. Coeff.	rmse (cm^{-1})	time (arb. units)	V_{3b}^{MWW} (cm^{-1})
821/4	F	716	15.9	1944.6	-
821/3	F	153	23.0	83.1	-11.3
4421/4	F	4,698	5.1	927.8	-134.8
4421/3	F	654	13.8	49.5	-102.8
422111*/4	P	15,551	2.2	301.0	-131.7
422111*/4	PP	5,809	4.6	100.0	-133.5
422111*/3	F	2,553	7.9	33.4	-137.1
422111*/3	P	1,245	8.8	27.8	-130.9
422111*/3	PP	729	10.5	13.6	-135.1
ab initio					**-130.7**

By summing up the pre-existing WHBB water potential[66] (which includes the water intramolecular potential, intrinsic two-body H_2O−H_2O and three-body $(H_2O)_3$ potentials), the methane intramolecular potential,[67] and a new intrinsic two-body CH_4−H_2O[59] and the present three-body CH_4−H_2O−H_2O potentials, we are able to approximate precisely the PES for a methane molecule surrounded by an arbitrary number of water monomers ($CH_4(H_2O)_n$). The approximation is at the three-body level of the many-body representation. Besides accuracy, computational costs of potential calls made with the fitted three-body CH_4−H_2O−H_2O potentials (see Table 1) need to be estimated when building such a complex PES.

The methane surface we have employed in our calculations is by Warmbier *et al.* However, the "plug-and-play" feature of the methane-hydrate surface allows us to associate our high-level three-body potential with different fitted surfaces for the monomers, to improve the accuracy for given applications. For example, other accurate PESs exist for methane, e.g., a

recent one due to Tennyson and co-workers.[68]

For the $CH_4-H_2O-H_2O$ trimer, three-body is the highest order term in the many-body representation. If all the terms in the analytical PES for $CH_4(H_2O)_n$ (n=2) were with no error, then the PES would be able to reproduce the exact dissociation energy (D_e). In a previous work,[59] we have presented two analytical intrinsic two-body CH_4-H_2O potentials and applied them to the dimer system. One fitted potential (called in our previous work PES_{2b}-PI) was obtained by means of the primary and secondary invariant technique, while the second one (PES_{2b}-CSM) is a pruned purified potential. By employing the same optimized geometry for the $CH_4-H_2O-H_2O$ trimer as before, high-level CCSD(T)-F12b/haTZ *ab initio* calculations were performed to evaluate each term of one-, two- and three-body interactions. For comparison, the same calculations were performed with the analytical PES using P-422111*/3, the potential that better approximates the intrinsic three-body *ab initio* value for the trimer geometry. The zero of energy was set for the three isolated monomers in their equilibrium configurations. The two-body energy from PES_{2b}-PI is almost exact if compared to the *ab initio* value, and the estimate of PES_{2b}-CSM is also very accurate. Our most reliable estimate of D_e for the trimer is 2371.3 cm^{-1}. It is obtained by using PES_{2b}-PI for the intrinsic two-body CH_4-H_2O interactions. Our value agrees well with the *ab initio* result of 2403.3 cm^{-1}. The difference is mainly due to the intrinsic two-body H_2O-H_2O term, the error of which is partially compensated by the overestimation of the three-body contribution. The main reason for these discrepancies lies in the different levels of electronic theory adopted. In fact, the two-body H_2O-H_2O PES is fitted to CCSD(T)/aVTZ energies, P-422111*/3 to MP2-F12/haTZ energies, while *ab initio* calculations were performed with CCSD(T)-F12b/haTZ. The three-body interaction is shown to have a non-negligible impact, accounting for about 5% of the dissociation energy. The energy of each term in the many-body representation is listed in Table 2.

To further point out the accuracy of our three-body potentials, we present in Figure 3 three 1-d cuts for the three potentials with lowest fitting rmse. The cuts are not included

Table 2: Energy of each term in the many-body representation of the CH_4−H_2O−H_2O trimer. Energies are in cm^{-1}. *Ab initio* calculations employ CCSD(T)-F12b/haTZ.

	ab initio	PES
CH_4 1-body	16.4	12.3
$H_2O(1)$ 1-body	16.8	18.4
$H_2O(2)$ 1-body	5.1	5.8
CH_4−$H_2O(1)$ 2-body	-232.9	-233.9[a],-252.6[b]
CH_4−$H_2O(2)$ 2-body	-330.0	-330.5[a],-331.7[b]
H_2O−H_2O 2-body	-1756.3	-1712.5
CH_4−H_2O−H_2O 3-body	-122.4	-130.9[c]
CH_4−H_2O−H_2O D_e	-2403.3	-2371.3[a,c]

[a] PES_{2b}-PI, [b] PES_{2b}-CSM, [c] P-422111*/3.

in the database of fitted energies. Starting from the optimized equilibrium geometry of the trimer, the cuts describe the change in three-body interaction energy against variation of the distance between the methane carbon atom and the oxygen atom of one of the two water monomers. In the cuts reported, all monomer internal geometries are frozen along the cut and the C-O distance is modified in one case moving the methane (upper panel) and in the other case shifting the water monomer (lower panel). The second water monomer is held at its initial position. The three potentials show excellent accuracy down to C-O distances of about 3 Angstroms, a short distance where two-body repulsive interactions start to be increasingly predominant.[59] The upper panel clearly points out that purified potentials rigorously ensure the zero-interaction asymptotic limit.

A final and more challenging application of our intrinsic three-body CH_4−H_2O−H_2O potentials has been performed. It concerns the calculation of the dissociation energy of a methane molecule in a dodecahedral water cage. Three different conformers of the empty $(H_2O)_{20}$ dodecahedral cage have been optimized by employing the WHBB water potential, and the corresponding energies were calculated. Geometries of CH_4 in these three $(H_2O)_{20}$ cages have been optimized by means of the analytical PES for $CH_4(H_2O)_n$ (n=20), but with methane-water interactions limited to the CH_4−H_2O two-body level. The energies of these $CH_4@(H_2O)_{20}$ systems were also calculated. **D_e excluding the three-body**

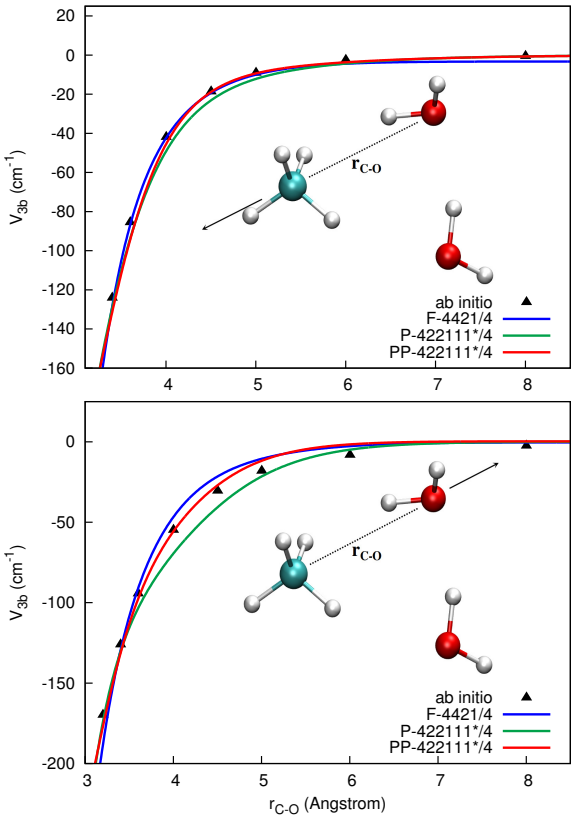


Figure 3: 1-d cuts for the intrinsic $CH_4{-}H_2O{-}H_2O$ three-body potential (V_{3b}^{MWW}). On the top panel, the C-O distance is varied by moving the methane monomer. In the lower panel, motion concerns the water monomer. The second water monomer is held at its initial position and geometry.

CH_4−H_2O−H_2O interaction was obtained as the difference between the calculated energies of the empty water cages and the corresponding $CH_4@(H_2O)_{20}$. To assess the impact of the intrinsic MWW interaction on D_e, geometries were not further optimized. The 190 three-body CH_4−H_2O−H_2O terms were directly evaluated using the analytical three-body potentials and then added together. For comparison, *ab initio* calculations of the three-body interactions for the three conformers were performed at the MP2-F12/haTZ level of theory, the same of our database. **Finally, D_e with three-body interaction was estimated as the difference between D_e without three-body CH_4−H_2O−H_2O interaction and the repulsive three-body MWW interaction**.

Table 3 reports the results. *Ab initio* calculations demonstrate that the three conformers have similar three-body CH_4−H_2O−H_2O interactions, with conformer 2 and 3 slightly higher in energy. This feature is best reproduced by P-422111*/4 and PP-422111*/4. Looking at the single conformer energies, P-422111*/4 and PP-422111*/4 give the best estimate for conformer 2 and conformer 3, while conformer 1 is best approximated by F-4421/4. The best potentials provide energies that are within 0.1-0.15 kcal/mol (or less) of the *ab initio* value. This is a small error (about 40-50 cm^{-1}) if one considers that there is a total of 190 terms that sum up to yield the three-body energy. A combination of previous results for the trimer and results for cluster systems shows that potentials with maximum polynomial order M=4 are in general more accurate than those with M=3. However, the latter are much faster and have good accuracy, so in some applications they could be the preferred choice. Cluster calculations point out once more that three-body contributions to the total energy are not negligible.

As a final comparison, we note that recently, Deible et al.[69] performed *ab initio* and quantum Monte Carlo calculations to determine the D_e of $CH_4@(H_2O)_{20}$. The structure they employed is labeled as conformer 3 in this article, and the authors reported a D_e of 5.3 kcal/mol. However, they froze the geometry of $(H_2O)_{20}$ when methane is enclathrated, thus underestimating D_e. They have also calculated the 20 terms in the two-body CH_4−H_2O inter-

action and the 190 terms of the three-body $CH_4-H_2O-H_2O$ interaction for their geometry, employing the high-level CCSD(T)-F12b method with VTZ-F12 basis set and counterpoise correction. The sum of two-body energies and the sum of three-body energies are -5.85 kcal/mol and 1.01 kcal/mol. The difference with our estimates is mainly due to the different geometries (which impact to the two-body energy for about 0.2 kcal/mol), level of electronic theory and counterpoise correction employed.

Table 3: **Dissociation** energies and three-body energies for three conformers of $CH_4@(H_2O)_{20}$. *Ab initio* calculations are at the MP2-F12/haTZ level. Energies are in kcal/mol.

	Conformer 1[a]	Conformer 2[b]	Conformer 3[c]
D_e without 3-body	6.77	6.81	6.79
F-4421/4	0.72	0.54	0.54
P-422111*/4	0.59	0.63	0.66
PP-422111*/4	0.61	0.63	0.65
F-422111*/3	0.66	0.53	0.53
P-422111*/3	0.67	0.59	0.57
PP-422111*/3	0.63	0.60	0.56
ab initio	0.75	0.77	0.77
D_e(PP-422111*/4)	6.16	6.18	6.14

[a] Structure extracted from crystal structure in Appendix of Ref. 70
[b] Structure from WHBB[66]
[c] Structure from the supplementary material of Ref. 69

Summary and Conclusions

We have presented two approaches to fit many-body, non-covalent interactions via monomial symmetrization. In one, the fitting polynomial basis, obtained with the standard monomial symmetrization procedure, is purified from polynomials not able to reproduce the long-range zero-interaction limit. In the second approach, pruned purified fitting bases were obtained by employing partially permutationally invariant symmetry and by further reducing the previously purified bases to the subset of polynomials that depend exclusively on inter-monomer distances. The two procedures led to potentials that reproduce the correct asymptotic limit,

and that are at the same time accurate and faster to evaluate. This last aspect is important for applications to complex systems, where the number of many-body interactions can be huge and their evaluation by means of full-basis potentials too demanding.

The fitted potentials have been employed for calculations involving the $CH_4(H_2O)_2$ trimer and the $CH_4@(H_2O)_{20}$ cluster. Results are excellent from the point of view of both accuracy and reduction of computational overheads. For instance, PP-422111*/4 can evaluate all 190 three-body $CH_4-H_2O-H_2O$ interactions in the $CH_4@(H_2O)_{20}$ cluster with a computational effort which is less than 6 times that needed by the evaluation of just 20 two-body CH_4-H_2O interactions with a potential obtained with the very efficient invariant polynomial technique. This relative factor of 6 even drops to 1 if PP-422111*/3 is employed. As for accuracy, the three-body contribution to the binding energy of the trimer is reproduced with errors of just a handful of wavenumbers, and the sum of 190 three-body terms for the $CH_4@(H_2O)_{20}$ cluster is underestimated by no more than 40-50 cm^{-1}. Three-body interactions have been demonstrated to be not negligible if compared to two-body ones. The possibility to construct efficient purified potentials based on monomial symmetrization was crucial for application to a 11-atom system. In fact, the alternative technique based on primary invariant polynomials is currently available only for systems up to 10 atoms.

The purified potentials here presented combine high accuracy and speed, thus opening up the possibility to undertake future high-level applications regarding spectroscopy and dynamics of methane hydrates. **All the methane-water-water interaction potentials reported are available upon request to the authors**.

Acknowledgements

We thank Professor Paul L. Houston for useful discussion. This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Award Number DE-FG02-97ER14782. C. Q. thanks NASA for financial

support through Grant No. 370NNX12AF42G from the NASA Astrophysics Research and Analysis program.

References

(1) Schatz, G. C. *Rev. Mod. Phys.* **1989**, *61*, 669–688.

(2) Bolton, K.; Hase, W. L.; Doubleday, C. *J. Phys. Chem. B* **1999**, *103*, 3691–3698.

(3) Sun, L.; Song, K.; Hase, W. L. *Science* **2002**, *296*, 875–878.

(4) Pu, J.; Truhlar, D. G. *J. Chem. Phys.* **2002**, *116*, 1468–1478.

(5) Troya, D.; Pascual, R. Z.; Schatz, G. C. *J. Phys. Chem. A* **2003**, *107*, 10497–10506.

(6) Sun, L.; Schatz, G. C. *J. Phys. Chem. B* **2005**, *109*, 8431–8438.

(7) Jasper, A. W.; Miller, J. A. *J. Phys. Chem. A* **2009**, *113*, 5612–5619.

(8) Jasper, A. W.; Miller, J. A. *J. Phys. Chem. A* **2011**, *115*, 6438–6455.

(9) Conte, R.; Aspuru-Guzik, A.; Ceotto, M. *J. Phys. Chem. Lett.* **2013**, *4*, 3407–3412.

(10) Hornik, K.; Stinchcombe, M.; White, H. *Neural Networks* **1989**, *2*, 359–366.

(11) Raff, L. M.; Malshe, M.; Hagan, M.; Doughan, D. I.; Rockley, M. G.; Komanduri, R. *J. Chem. Phys.* **2005**, *122*, 084104.

(12) Manzhos, S.; Wang, X.; Dawes, R.; Carrington, T. *J. Phys. Chem. A* **2006**, *110*, 5295–5304.

(13) Handley, C. M.; Popelier, P. L. A. *J. Phys. Chem. A* **2010**, *114*, 3371–3383.

(14) Behler, J.; Parrinello, M. *Phys. Rev. Lett.* **2007**, *98*, 146401.

(15) Darley, M. G.; Handley, C. M.; Popelier, P. L. A. *J. Chem. Theory and Comp.* **2008**, *4*, 1435–1448.

(16) Behler, J. *Phys. Chem. Chem. Phys.* **2011**, *13*, 17930–17955.

(17) Kondati Natarajan, S.; Morawietz, T.; Behler, J. *Phys. Chem. Chem. Phys.* **2015**, doi: 10.1039/C4CP04751F.

(18) Chen, J.; Xu, X.; Xu, X.; Zhang, D. H. *J. Chem. Phys.* **2013**, *138*, 221104.

(19) Carter, S.; Culik, S. J.; Bowman, J. M. *J. Chem. Phys.* **1997**, *107*, 10458–10469.

(20) Yagi, K.; Oyanagi, C.; Taketsugu, T.; Hirao, K. *J. Chem. Phys.* **2003**, *118*, 1653–1660.

(21) Rauhut, G.; Barone, V.; Schwerdtfeger, P. *J. Chem. Phys.* **2006**, *125*, 054308.

(22) Jäckle, A.; Meyer, H.-D. *J. Chem. Phys.* **1998**, *109*, 3772–3779.

(23) Thompson, K. C.; Jordan, M. J. T.; Collins, M. A. *J. Chem. Phys.* **1998**, *108*, 8302–8316.

(24) Collins, M. A. *J. Chem. Phys.* **2007**, *127*, 024104.

(25) Zou, S.; Bowman, J. M. *Chem. Phys. Lett.* **2003**, *368*, 421–424.

(26) Zhang, X.; Zou, S.; Harding, L. B.; Bowman, J. M. *J. Phys. Chem. A* **2004**, *108*, 8980–8986.

(27) Chen, C.; Braams, B.; Lee, D. Y.; Bowman, J. M.; Houston, P. L.; Stranges, D. *J. Phys. Chem. A* **2011**, *115*, 6797–6804.

(28) Conte, R.; Houston, P. L.; Bowman, J. M. *J. Phys. Chem. A* **2014**, *118*, 7742–7757.

(29) Bowman, J. M.; Czako, G.; Fu, B. *Phys. Chem. Chem. Phys.* **2011**, *13*, 8094–8111.

(30) Homayoon, Z.; Bowman, J. M. *J. Phys. Chem. A* **2013**, *117*, 11665–11672.

(31) Shepler, B. C.; Braams, B. J.; Bowman, J. M. *J. Phys. Chem. A* **2008**, *112*, 9344–9351.

(32) Fu, B.; Han, Y.-C.; Bowman, J. M.; Angelucci, L.; Balucani, N.; Leonori, F.; Casavecchia, P. *Proc. Nat. Acad. Sc.* **2012**, *109*, 9733–9738.

(33) Xie, C.; Li, J.; Xie, D.; Guo, H. *J. Chem. Phys.* **2012**, *137*, 024308.

(34) Li, J.; Carter, S.; Bowman, J. M.; Dawes, R.; Xie, D.; Guo, H. *J. Phys. Chem. Lett.* **2014**, *5*, 2364–2369.

(35) Li, J.; Guo, H. *Phys. Chem. Chem. Phys.* **2014**, *16*, 6753–6763.

(36) Fu, B.; Han, Y.; Bowman, J. M. *Faraday Discuss.* **2012**, *157*, 27–39.

(37) Braams, B. J.; Bowman, J. M. *Int. Rev. Phys. Chem.* **2009**, *28*, 577–606.

(38) Bosma, W.; Cannon, J.; Playoust, C. *J. Symb. Comp.* **1997**, *24*, 235 – 265.

(39) Huang, X.; Braams, B. J.; Bowman, J. M. *J. Chem. Phys.* **2005**, *122*, 044308.

(40) Xie, Z.; Bowman, J. M. *J. Chem. Theory Comput.* **2010**, *6*, 26–34.

(41) Software can be downloaded at *http://www.chemistry.emory.edu/faculty/bowman/msa/index.html* (accessed Jan. 18, 2011).

(42) Li, J.; Jiang, B.; Guo, H. *J. Chem. Phys.* **2013**, *139*, 204103.

(43) Jiang, B.; Guo, H. *J. Chem. Phys.* **2014**, *141*, 034109.

(44) Paukku, Y.; Yang, K. R.; Varga, Z.; Truhlar, D. G. *J. Chem. Phys.* **2013**, *139*, 044309.

(45) Conte, R.; Houston, P. L.; Bowman, J. M. *J. Phys. Chem. A* **2013**, *117*, 14028–14041.

(46) Varandas, A. J. C.; Rodrigues, S. P. J. *J. Chem. Phys.* **1997**, *106*, 9647–9658.

(47) Rodrigues, S. P. J.; Varandas, A. J. C. *J. Phys. Chem. A* **1998**, *102*, 6266–6273.

(48) Tang, K. T.; Toennies, J. P. *J. Chem. Phys.* **1984**, *80*, 3726–3741.

(49) Marcelli, G.; Sadus, R. J. *J. Chem. Phys.* **1999**, *111*, 1533–1540.

(50) Tainter, C. J.; Skinner, J. L. *J. Chem. Phys.* **2012**, *137*, 104304.

(51) Gora, U.; Cencek, W.; Podeszwa, R.; van der Avoird, A.; Szalewicz, K. *J. Chem. Phys.* **2014**, *140*, 194101.

(52) Yu, K.; Schmidt, J. R. *J. Chem. Phys.* **2012**, *136*, 034503.

(53) Shank, A.; Wang, Y.; Kaledin, A.; Braams, B. J.; Bowman, J. M. *J. Chem. Phys.* **2009**, *130*, 144314.

(54) Wang, Y.; Shepler, B. C.; Braams, B. J.; Bowman, J. M. *J. Chem. Phys.* **2009**, *131*, 054511.

(55) Mancini, J. S.; Bowman, J. M. *J. Chem. Phys.* **2013**, *139*, 164115.

(56) Mancini, J. S.; Bowman, J. M. *J. Phys. Chem. A* **2014**, *118*, 7367–7374.

(57) Mancini, J. S.; Bowman, J. M. *J. Phys. Chem. Lett.* **2014**, *5*, 2247–2253.

(58) Conte, R.; Houston, P. L.; Bowman, J. M. *J. Chem. Phys.* **2014**, *140*, 151101.

(59) Qu, C.; Conte, R.; Houston, P. L.; Bowman, J. M. *Phys. Chem. Chem. Phys.* **2015**, doi: 10.1039/C4CP05913A.

(60) Sloan, E. D.; Koh, C. A. *Clathrate hydrates of natural gases*, 3rd ed.; CRC Press, Taylor & Francis Group, 2008; pp 1–44.

(61) Werner, H.-J.; Adler, T. B.; Manby, F. R. *J. Chem. Phys.* **2007**, *126*, 164102.

(62) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007.

(63) Werner, H.-J. et al. MOLPRO, version 2010.1, a package of ab initio programs. 2010; see http://www.molpro.net.

(64) Adler, T. B.; Knizia, G.; Werner, H.-J. *J. Chem. Phys.* **2007**, *127*, 221106.

(65) Knizia, G.; Adler, T. B.; Werner, H.-J. *J. Chem. Phys.* **2009**, *130*, 054104.

(66) Wang, Y.; Huang, X.; Shepler, B. C.; Braams, B. J.; Bowman, J. M. *J. Chem. Phys.* **2011**, *134*, 094509.

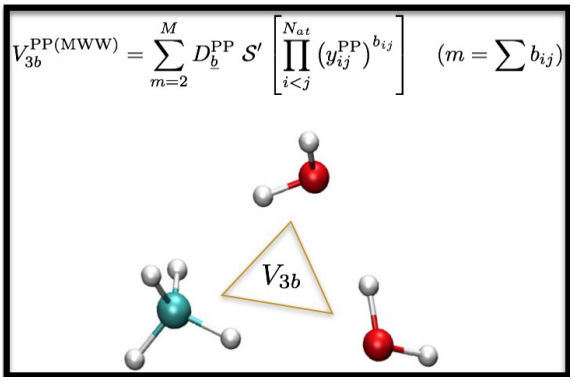
(67) Warmbier, R.; Schneider, R.; Sharma, A. R.; Braams, B. J.; Bowman, J. M.; Hauschildt, P. H. *Astron. Astrophys.* **2009**, *495*, 655–661.

(68) Yurchenko, S. N.; Tennyson, J.; Barber, R. J.; Thiel, W. *J. Mol. Spectrosc.* **2013**, *291*, 69–76.

(69) Deible, M. J.; Tuguldur, O.; Jordan, K. D. *J. Phys. Chem. B* **2014**, *118*, 8257–8263.

(70) Sparks, K. A. Configurational properties of water clathrates through molecular simulation. Ph.D. thesis, Massachusetts Institute of Technology, 1991.

For TOC only


$V_{3b}^{\mathrm{PP(MWW)}} = \sum_{m=2}^{M} D_{\underline{b}}^{\mathrm{PP}}\ \mathcal{S}' \left[\prod_{i<j}^{N_{at}} \left(y_{ij}^{\mathrm{PP}}\right)^{b_{ij}} \right] \quad (m = \sum b_{ij})$
V_{3b}