Energy benchmarks for methane-water systems from quantum Monte Carlo and second-order Møller-Plesset calculations

M. J. Gillan, D. Alfè, and F. R. Manby

Citation: The Journal of Chemical Physics **143**, 102812 (2015); doi: 10.1063/1.4926444
View online: http://dx.doi.org/10.1063/1.4926444
View Table of Contents: http://scitation.aip.org/content/aip/journal/jcp/143/10?ver=pdfcov
Published by the AIP Publishing

Energy benchmarks for methane-water systems from quantum Monte Carlo and second-order Møller-Plesset calculations

M. J. Gillan,[1,2,3,a)] D. Alfè,[1,2,3,4] and F. R. Manby[5]
[1]*London Centre for Nanotechnology, University College London, Gordon St., London WC1H 0AH, United Kingdom*
[2]*Department of Physics and Astronomy, University College London, Gower St., London WC1E 6BT, United Kingdom*
[3]*Thomas Young Centre, University College London, Gordon St., London WC1H 0AH, United Kingdom*
[4]*Department of Earth Sciences, University College London, Gower St., London WC1E 6BT, United Kingdom*
[5]*Centre for Computational Chemistry, School of Chemistry, University of Bristol, Bristol BS8 1TS, United Kingdom*

(Received 26 March 2015; accepted 26 June 2015; published online 9 July 2015)

The quantum Monte Carlo (QMC) technique is used to generate accurate energy benchmarks for methane-water clusters containing a single methane monomer and up to 20 water monomers. The benchmarks for each type of cluster are computed for a set of geometries drawn from molecular dynamics simulations. The accuracy of QMC is expected to be comparable with that of coupled-cluster calculations, and this is confirmed by comparisons for the CH_4-H_2O dimer. The benchmarks are used to assess the accuracy of the second-order Møller-Plesset (MP2) approximation close to the complete basis-set limit. A recently developed embedded many-body technique is shown to give an efficient procedure for computing basis-set converged MP2 energies for the large clusters. It is found that MP2 values for the methane binding energies and the cohesive energies of the water clusters without methane are in close agreement with the QMC benchmarks, but the agreement is aided by partial cancelation between 2-body and beyond-2-body errors of MP2. The embedding approach allows MP2 to be applied without loss of accuracy to the methane hydrate crystal, and it is shown that the resulting methane binding energy and the cohesive energy of the water lattice agree almost exactly with recently reported QMC values. © *2015 AIP Publishing LLC.* [http://dx.doi.org/10.1063/1.4926444]

I. INTRODUCTION

Solid methane hydrate has been intensively studied because of its importance in the oil industry,[1] its potential as a future energy source,[2,3] and its relevance to climate change,[4] while methane-water solutions have been studied for related reasons.[5,6] In addition, methane molecules in aqueous solution are pertinent to the study of the hydrophobic interactions that are so important throughout biology.[7–10] Computer simulations based on empirical force fields have been an important tool for investigating processes such as the nucleation and growth of methane hydrate and the dissolution of methane gas in water,[2,5,11–27] and simulations based on density functional theory (DFT)[6,28–32] have also been valuable for the study of methane-water systems. Energy benchmarks from accurate quantum-mechanical calculations are essential for testing both force fields and DFT methods.[33] Here, we report a substantial set of benchmarks from the quantum Monte Carlo (QMC) technique[34–36] for thermally disordered methane-water clusters containing up to 20 water monomers. Using these benchmarks together with other recent QMC results,[37,38] we assess the accuracy of the second-order Møller-Plesset (MP2) technique for both methane-water clusters and the methane hydrate crystal.

Quantum chemistry calculations have been an important source of information for parameterizing intermolecular force fields since the early work of Clementi and co-workers on water in the 1970s.[39,40] Today the CCSD(T) (coupled cluster with single and double excitations and perturbative triples)[41] technique is often regarded as the "gold standard" of quantum chemical accuracy for non-covalent binding in such systems, being capable of an accuracy for dimer binding energies of typically 0.2 mE_h (~0.13 kcal/mol ≃ 5 meV). For water and some other molecular systems, the accuracy of the more efficient MP2 technique[41] falls not far short of this.[42,43] The last 10 years have seen a surge in the use of these methods to create sophisticated and accurate force fields, particularly for water, and also for other systems. Work of this kind on methane-water systems includes Refs. 44 and 45. The use of MP2 and CCSD(T) benchmarks to test DFT methods has also become widespread. An early study of the CH_4-H_2O dimer[44] showed that standard DFT approximations can be very inaccurate, one reason being their neglect of van der Waals dispersion. Much subsequent work has confirmed this, and the recent extensive survey by Liu *et al.*[46] on methane-water clusters showed that even dispersion-inclusive DFT methods can suffer from large errors.

The computational demands of MP2 and CCSD(T) increase very rapidly with the number of atoms, so that energy benchmarks near the complete basis-set (CBS) limit have generally been limited to rather small clusters. Furthermore,

a)Author to whom correspondence should be addressed. Electronic mail: m.gillan@ucl.ac.uk

the implementation of these methods in periodic boundary conditions is technically challenging, so that accurate MP2 and CCSD(T) benchmarks for crystals are seldom reported (but see, e.g., Refs. 47–52). QMC techniques offer an alternative approach. There is ample evidence that the accuracy of current QMC techniques for weak non-covalent interactions is near that of CCSD(T).[53–59] However, the computational demands of QMC grow much more slowly with system size, so that basis-set converged calculations on systems containing many tens of atoms are feasible, and the techniques work equally well for clusters and for periodic systems. There has already been a considerable body of QMC work reported on water (including ice structures and liquid water)[55,60–65] and other molecular systems, which have given new ways of probing the deficiencies of DFT approximations. The very recent QMC benchmarks on the methane clathrate cage (a CH_4 monomer surrounded by 20 H_2O monomers)[37] and the methane hydrate crystal[38] have confirmed previous indications that widely used DFT approximations suffer from substantial errors, even if dispersion is explicitly included.

The QMC benchmarks presented here are designed to complement those reported for the methane clathrate cage[37] and the methane hydrate crystal.[38] The clusters we treat consist of a single CH_4 surrounded by a number n of H_2O monomers ranging from 1 to 20, with the configurations for each cluster being drawn from molecular dynamics (MD) simulations, so that they sample a range of intermolecular separations and molecular orientations.

To illustrate the usefulness of the QMC benchmarks, we use them to test the accuracy of MP2 for methane-water energetics. For the clusters studied here, direct MP2 calculations can be performed close to the CBS limit. However, it is also of interest to test the accuracy of MP2 when it is implemented using a recently developed embedded many-body technique,[66,67] since this technique is much more efficient and also allows us to perform MP2 calculations on periodic systems. We shall show that the embedded many-body implementation of MP2 gives results that are almost identical to those of standard MP2. Our comparisons of MP2 with the cluster QMC benchmarks show that its accuracy is considerably better than that of typical DFT approximations, and suggest that MP2 should also be quite accurate for the methane hydrate crystal. The recently published QMC benchmarks on the crystal[38] together with the embedding-MP2 calculations presented here will allow us to confirm this.

II. TECHNIQUES

The QMC benchmarks reported here were computed by the diffusion Monte Carlo (DMC) technique[34–36] with the CASINO code.[68] The technical settings follow closely those used in our previous work on water systems.[56,61,64] Dirac-Fock pseudopotentials were used,[70,71] non-locality being treated by the usual locality approximation.[69] The trial wavefunctions were of single-determinant Slater-Jastrow form, the single-electron orbitals being obtained from DFT-LDA plane-wave calculations with the PWSCF package[72] using the plane-wave cutoff 4082 eV. These calculations of the orbitals are performed in periodic boundary conditions, with the length L of the cubic repeating cell taken large enough to ensure that residual interactions between periodic images of the clusters do not significantly affect the orbitals; we generally take $L = 40$ a.u. (21.2 Å). For the purpose of the CASINO calculations, the representation of the orbitals was transformed from plane-waves to a basis of B-splines.[73] The time step used in all the calculations was 0.005 a.u., as in our previous work. Tests with a time step of 0.002 a.u. on the $CH_4(H_2O)_{10}$ clusters showed differences of between 0.2 and 0.3 mE_h in the energies per monomer reported below.

Benchmarks will be presented for two important energies characterizing the clusters $CH_4(H_2O)_n$. The first, denoted by ΔE_{CH_4}, is the binding energy of the methane monomer, i.e., the negative of the energy needed to remove the methane from the cluster, leaving the positions of all water monomers unchanged. The second, ΔE_{bind}^{empty}, is the binding energy per monomer of the resulting "empty" water cluster relative to free water monomers. These two energies were studied in Ref. 38, where they were shown to be particularly important for the energetics of the methane hydrate crystal. We want to obtain benchmarks for these energies to somewhat better than chemical accuracy (notionally 1 kcal/mol), and we adopt the target of reducing all errors, including the statistical error on the DMC calculations themselves, to less than 1 mE_h (0.6 kcal/mol or 27 meV).

Both ΔE_{CH_4} and ΔE_{bind}^{empty} involve computing the difference of two very large energies (typically a few thousand eV), and this means paying close attention to cancelation of errors. We find that this is best achieved by requiring that the energies to be subtracted refer to systems containing exactly the same set of atoms in different configurations. For ΔE_{CH_4}, we ensure this by computing the DMC total energy of the CH_4-$(H_2O)_n$ cluster of interest and subtracting the DMC energy of exactly the same system but with the CH_4 monomer rigidly displaced to a distance large enough to render the interaction between the CH_4 and $(H_2O)_n$ fragments negligible. We use a similar approach to calculate ΔE_{bind}^{empty}. In this case, we compute the DMC total energy of the $(H_2O)_n$ cluster, and we subtract from it the DMC energy of the system obtained by rigidly displacing all the H_2O monomers radially outwards from the origin by a chosen scaling factor, the origin being chosen as the position of the C atom in the CH_4 monomer before it was removed. To correct for the remaining (mainly electrostatic) interactions between the H_2O monomers, we compute them using MP2. (Note that the value chosen for the scaling factor used in the outward displacement is limited by the periodic boundary conditions employed in calculating the single-electron orbitals.)

For all the systems studied, we compare our QMC benchmarks with MP2 values of ΔE_{CH_4} and ΔE_{bind}^{empty} and we also examine δCCSD(T) corrections, i.e., the energy changes on going from MP2 to coupled-cluster CCSD(T). All these quantum chemistry calculations were performed with the MOLPRO code.[74,75] For both MP2 and δCCSD(T), we employ explicitly correlated F12 methods,[76] which greatly accelerate basis-set convergence. The calculations all employ correlation consistent aug-cc-pVXZ basis sets,[77,78] referred to here as AVXZ, where X is the cardinality. For all the MP2 calculations, we employ AVQZ basis sets, unless otherwise noted. These methods are expected to yield MP2 values of ΔE_{CH_4} and ΔE_{bind}^{empty} within ~0.1 mE_h of the CBS limit. The δCCSD(T)

corrections are straightforward to compute for the CH_4-H_2O dimer, where an accuracy of ~20 μE_h is obtained for these corrections using AVTZ basis sets with the counterpoise method. However, for CH_4-$(H_2O)_5$ and the larger clusters, we do not attempt direct CCSD(T) calculations on the whole cluster, and instead we content ourselves with the computation of 2- and 3-body δCCSD(T) corrections. The 2-body corrections are treated with AVTZ and counterpoise, and the 3-body corrections with AVDZ and full counterpoise (i.e., the basis set of the entire trimer is used for the computation of the monomer and dimer energies).

The embedding procedure that we use to compute the MP2 energies of the larger $CH_4(H_2O)_n$ clusters and the methane hydrate crystal is based on the embedded many-body expansion (EMBE) described by Bygrave, Allan, and Manby (BAM),[66] and we performed the calculations using a development version of the MOLPRO code. We follow here the procedure used in our recent MP2 calculations on water clusters and ice structures.[67] In this procedure, we separate the total energy into its Hartree-Fock (HF) and electron-correlation components, and EMBE is applied only to the correlation energy, the HF component being computed by standard methods. By "standard methods," we mean that for the clusters, the HF components of ΔE_{CH_4} and ΔE_{bind}^{empty} are computed directly with MOLPRO using density fitting[79] with AVQZ basis sets, which ensure basis-set convergence to be within ~0.1 mE_h. The corresponding quantities for the methane hydrate crystal are computed with the VASP code,[80] following the methods reported earlier for ice structures.[67]

Returning to the correlation components of ΔE_{CH_4} and ΔE_{bind}^{empty}, we recall that in the BAM form of EMBE, the embedding fields are the sum of Coulomb fields due to the charge distributions of the monomers and exchange-repulsion fields which are also obtained from these distributions via the Wheatley-Price relation between exchange-overlap interaction and density overlap.[81] The charge distributions from which the fields are derived are computed self-consistently by HF calculations on the monomers. In our work on water clusters and ice structures, we tested the effect of truncating the EMBE at the 2-body level and we showed that for the clusters, this yields a very accurate approximation to the MP2 correlation energy; we show that the same is true in the present work. In applying the 2-body-truncated EMBE to periodic systems, a spatial cutoff is applied, so that electron correlation is included only for pairs of monomers separated by a specified cutoff radius R_c. Convergence with respect to R_c must, of course, be tested.

III. QMC AND MP2 BENCHMARKS FOR CLUSTERS

A. The CH_4-H_2O dimer

We start by confirming the accuracy of QMC for the CH_4-H_2O dimer, which is clearly necessary for confidence in its ability to deliver useful benchmarks. Our criterion is that it must agree closely with "gold-standard" CCSD(T) calculations. While checking this, we also take the opportunity to assess the errors of the MP2 approximation. To avoid any restriction to special geometries, we make our comparisons for a set of configurations spanning a wide range of intermolecular separations and molecular orientations.

Our set of dimer configurations is drawn from a DFT-based MD simulation of a single methane molecule in liquid water performed with the BLYP exchange-correlation functional.[82–84] The simulation was performed using Langevin dynamics at the very high temperature of 1000 K to ensure that the system sampled configurations having short C-O distances. The usual periodic boundary conditions were employed, with 31 H_2O molecules and a single CH_4 molecule in the repeating cell, the mass density being 0.995 g/cm^3. The monomers in the simulated system undergo internal vibrations, so that the energy of each dimer configuration relative to free CH_4 and H_2O monomers is the sum of (a) the 1-body distortion energies of the individual monomers and (b) the 2-body interaction energy. We are interested here only in the interaction energy, and we therefore modify the configurations drawn from the simulation by restoring each monomer to its equilibrium gas-phase geometry.[85] With the configurations adjusted in this way, the 2-body interaction energy of each is simply its total energy minus the sum of the energies of the isolated equilibrium monomers.

As indicated in Sec. II, the benchmark CCSD(T) values of the interaction energies are computed as the sum of the MP2 values and the δCCSD(T) corrections. With AVQZ basis sets for MP2 and AVTZ for the δCCSD(T) difference, F12 and counterpoise being used for both, we expect values of the interaction energies to be within ~0.1 mE_h of the CBS limit. The DMC energies of the 100 dimer configurations were computed as explained in Sec. II, each run being continued until the rms statistical error was reduced to ~60 μE_h.

The differences between DMC and CCSD(T) values of the dimer interaction energy are plotted against C-O separation of the monomers in Fig. 1, which also presents the corresponding

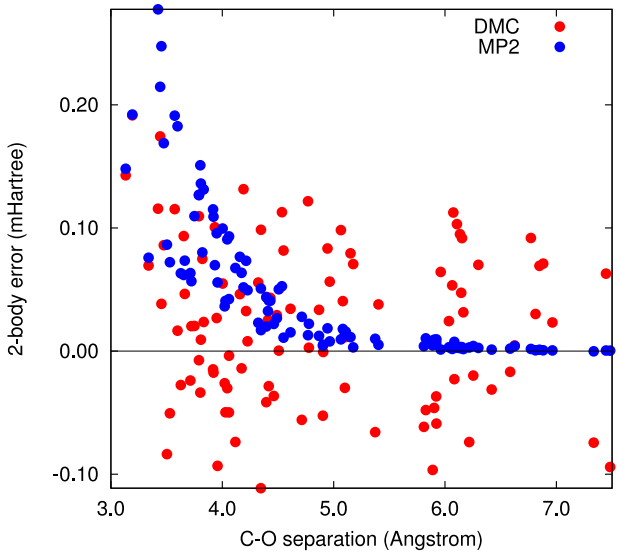

FIG. 1. Interaction energy of the CH_4-H_2O dimer as function of C-O separation for a thermal sample of 100 configurations (see text). Red and blue symbols show deviations of DMC and MP2 values of interaction energy from benchmark CCSD(T) values. Note that deviations in the case of DMC are mainly due to statistical sampling error. Units of energy and distance: mE_h and Å.

differences between MP2 and CCSD(T). The DMC differences arise almost entirely from the statistical sampling error of DMC itself, as is clear from the fact that the mean value of the differences shown in Fig. 1 is only 20 μE_h (0.5 meV), which is probably beyond the accuracy of CCSD(T) itself. The rms fluctuation of the difference between DMC and CCSD(T) is 64 μE_h, which is only slightly greater than the DMC sampling error of 60 μE_h. By contrast, the figure shows that MP2 is systematically underbound, with an error that increases quite rapidly as the monomers approach each other. At the C-O separation of 3.8 Å typical of the methane hydrate crystal at ambient pressure, the MP2 error may be over ~0.1 mE_h. As we point out later, such an error is not negligible for the overall energetics of the hydrate.

The present comparisons suggest that any systematic errors of DMC are considerably smaller than those of MP2.

B. The $CH_4(H_2O)_5$ cluster

As well as yielding accurate benchmarks, our study of the $CH_4(H_2O)_5$ clusters serves two other purposes: it allows us to make further tests of DMC against CCSD(T), and it provides tests of the embedding approximation for computing MP2 correlation energy (see Sec. II). As with the dimer, we created a set of configurations of the cluster by drawing them from a MD simulation of methane dissolved in water.

The MD simulation used in this case was performed with a classical force field on a large periodic system of 2944 H_2O molecules and 512 CH_4 molecules, both kinds of molecules being treated as rigid. The force field used to generate the MD was TIP4P/ICE[86] for H_2O-H_2O interactions, and the Tse-Klein-McDonald force field[87] for CH_4-CH_4 interactions, with CH_4-H_2O interactions constructed by the usual Lorentz-Berthelot combining rules. The original purpose of this simulation was to study the nucleation of methane hydrate and for that reason, it extended over a period of nearly 1 μs. Cluster configurations were created from the simulation as follows. A set of snapshots was first taken from the simulation. In each snapshot, a methane monomer was then chosen at random, with the condition that the C-C distance to every other methane monomer is at least 7.5 Å, so that the first methane is entirely surrounded by water molecules. Finally, the five nearest water monomers (as measured by the C-O distance) were taken to form the CH_4-$(H_2O)_5$ cluster. This procedure, which we also use to create the larger CH_4-$(H_2O)_n$ clusters, is designed to produce random configurations having some relevance to methane hydration in solution.

We focus here on the energies ΔE_{CH_4} and ΔE_{bind}^{empty} for each cluster configuration (see Sec. II): the binding energy of the methane monomer and the binding energy per monomer of the empty water cluster resulting from removal of the methane. The DMC value of ΔE_{CH_4} for each configuration was computed as the difference of energies in two runs: one for the given CH_4-$(H_2O)_5$ configuration and the other for the configuration obtained by displacing the CH_4 monomer (see Sec. II). The displacement used for this purpose was 10 Å. Likewise, ΔE_{bind}^{empty} was computed from the difference of two DMC energies: that of the given $(H_2O)_5$ configuration and that of the configuration obtained by moving all H_2O monomers radially outwards by a scaling factor, the value of this factor being chosen as 1.5. The remaining interactions between the monomers in the scaled configuration were computed by MP2 and subtracted. The DMC runs were always continued long enough to reduce the statistical sampling error on ΔE_{CH_4} to ~0.4 mE_h (the sampling errors on ΔE_{bind}^{empty} are considerably smaller). We have used the resulting DMC benchmarks for ΔE_{CH_4} and ΔE_{bind}^{empty} to test direct MP2 calculations on the given configurations. The MP2 values of ΔE_{CH_4} and ΔE_{bind}^{empty} were obtained by simple subtraction of the appropriate numbers of monomer energies from the directly calculated MP2 energies of the CH_4-$(H_2O)_5$ and $(H_2O)_5$ clusters. With AVQZ basis sets and the F12 techniques employed, the MP2 values of ΔE_{CH_4} and ΔE_{bind}^{empty} are expected to be within ~0.1 mE_h of the CBS limit.

Parity plots of the MP2 values of ΔE_{CH_4} and ΔE_{bind}^{empty} *vs* the DMC values are shown in Fig. 2. It appears that MP2 makes ΔE_{CH_4} somewhat underbound compared with DMC, the mean difference between the two values being 0.52 mE_h and the rms fluctuation of the difference being 0.31 mE_h. (Note that a considerable part of this rms fluctuation is attributable to DMC statistical error.) For ΔE_{bind}^{empty}, MP2 agrees much more closely with DMC, the mean difference being 15 μE_h and the rms fluctuation of the difference being 53 μE_h.

The underbinding of ΔE_{CH_4} values given by MP2 is expected from our results for the CH_4-H_2O dimer, but it is interesting to ask whether the differences between the MP2 and DMC values of ΔE_{CH_4} are attributable entirely to the 2-body errors of MP2. We test this by computing the 2-body δCCSD(T) correction, i.e., the difference between CCSD(T) and MP2 values of the sum of 2-body contributions to ΔE_{CH_4}, which we then add to the total MP2 value of ΔE_{CH_4}. The resulting 2-body-corrected MP2 values are included in the parity plots of Fig. 2. As expected, the 2-body δCCSD(T) corrections make ΔE_{CH_4} more negative, and the agreement with DMC is thereby improved. In a similar way, we have also computed the 3-body δCCSD(T) correction, and the resulting 2- and 3-body-corrected MP2 values of ΔE_{CH_4} (see figure) reveal that the 3-body corrections are small but are always positive. The mean difference between 2- and 3-body corrected MP2 values and DMC values of ΔE_{CH_4} is less than 10 μE_h and the rms fluctuation of this difference is 0.31 mE_h. We have also computed the 2- and 3-body δCCSD(T) corrections to MP2 for ΔE_{bind}^{empty}, but they turn out to be extremely small (see lower panel of Fig. 2).

The embedding approach to the calculation of the MP2 correlation energy will be important later in the paper, and it is convenient to test its accuracy on the CH_4-$(H_2O)_5$ cluster. As explained in Sec. II, we compute an approximation to the true MP2 correlation energy by truncating the embedded MBE at the 2-body level, while the HF energy is computed directly without embedding. We show in Fig. 3 the parity plots of the MP2 values of ΔE_{CH_4} and ΔE_{bind}^{empty} computed by this embedding approximation and the standard method. It is clear that the errors of the embedding approximation are rather small. In fact, the mean and the rms fluctuations of the deviations of the embedding values from the standard values are −0.13 mE_h and 62 μE_h for ΔE_{CH_4} and 20 μE_h and 10 μE_h for ΔE_{bind}^{empty}. The conclusion is that the embedding approximation can safely be used to compute the MP2 values of the energies. In the

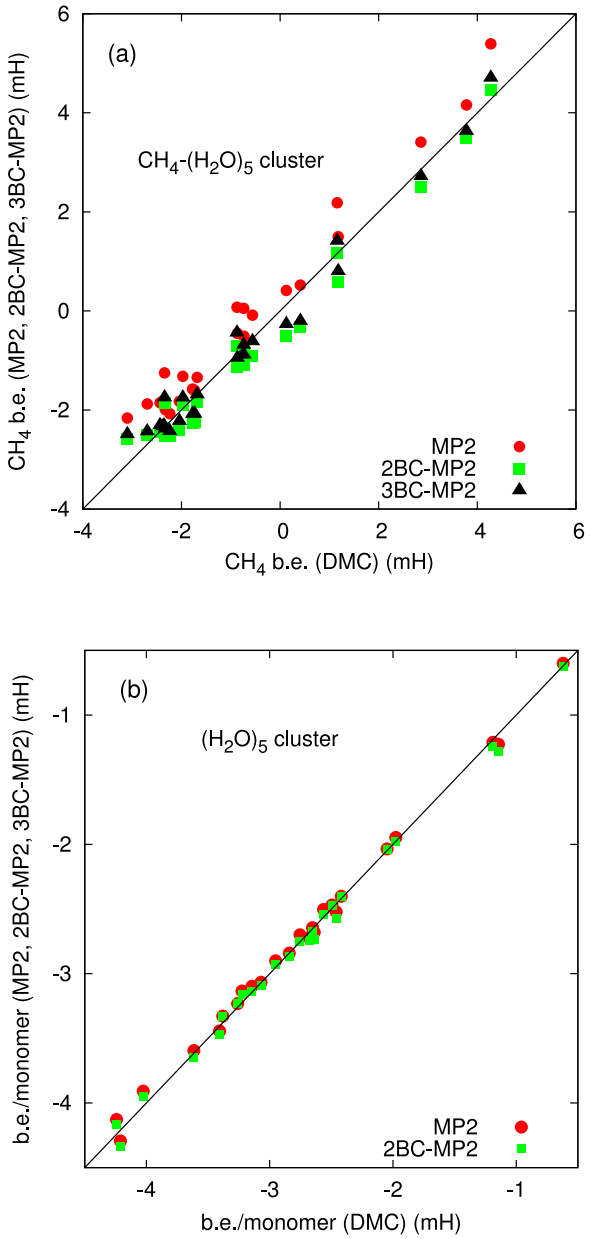

FIG. 2. Binding energy ΔE_{CH_4} of the methane monomer (upper panel) and binding energy per monomer ΔE_{bind}^{empty} of the empty water cluster for 25 configurations of the CH_4-$(H_2O)_5$ cluster. Panels show parity plots of uncorrected MP2 (red circles), MP2 with 2-body δCCSD(T) corrections (green squares), and MP2 with 2- and 3-body δCCSD(T) corrections (black triangles) *vs* DMC values.

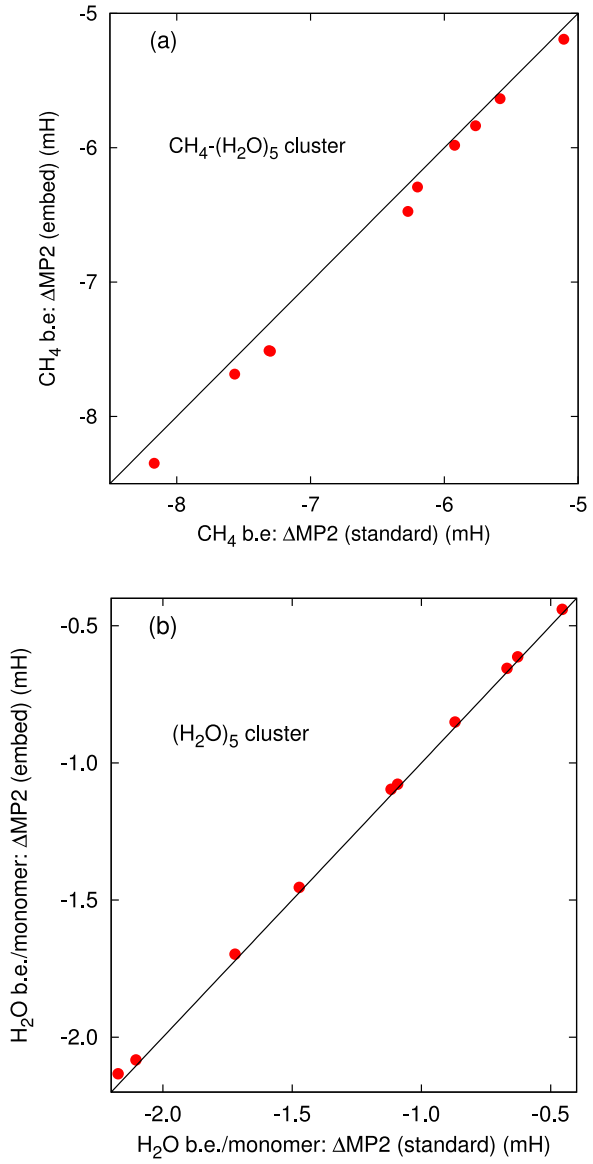

FIG. 3. Parity plots of ΔMP2 correlation components of ΔE_{CH_4} and ΔE_{bind}^{empty} computed by the 2-body embedding approximation and by the standard method.

following, if it is necessary to distinguish MP2 energies computed by embedding from those computed in the standard way, we refer to the embedding approximation of MP2 as E-MP2.

C. Clusters from CH_4-$(H_2O)_{10}$ to CH_4-$(H_2O)_{20}$

Using the procedure described in Sec. III B, we created configuration sets for the clusters CH_4-$(H_2O)_n$, for n = 10, 15, and 20 by drawing them from the same MD simulation of the bulk water-methane solution. In each case, DMC benchmark values of ΔE_{CH_4} and ΔE_{bind}^{empty} were computed by continuing the runs long enough to reduce the statistical sampling errors on ΔE_{CH_4} to ~0.4 mE_h for n = 10 and 15 and ~0.5 mE_h for n = 20 (the sampling errors on ΔE_{bind}^{empty} are considerably smaller). We report DMC values of ΔE_{bind}^{empty} only for the CH_4-$(H_2O)_{10}$ cluster. For clusters of these sizes, direct MP2 calculations with F12 and AVQZ basis sets are still feasible, but in practice, it is more convenient to use the 2-body-truncated EMBE approximation to compute the MP2 correlation energy and then add this to the directly calculated HF energy to obtain the embedded-MP2 approximation referred to here is E-MP2.

Parity plots of the E-MP2 values of ΔE_{CH_4} and ΔE_{bind}^{empty} *vs* the DMC benchmarks for the CH_4-$(H_2O)_{10}$ cluster are presented in Fig. 4, and parity plots of ΔE_{CH_4} for the CH_4-$(H_2O)_{15}$ and CH_4-$(H_2O)_{20}$ clusters in Fig. 5. The figures also report the E-MP2 values with 2- and 3-body δCCSD(T) corrections (the corrected values were obtained for only a subset of the configurations in some cases, since they are laborious to compute). We see that E-MP2 values of ΔE_{CH_4} al-

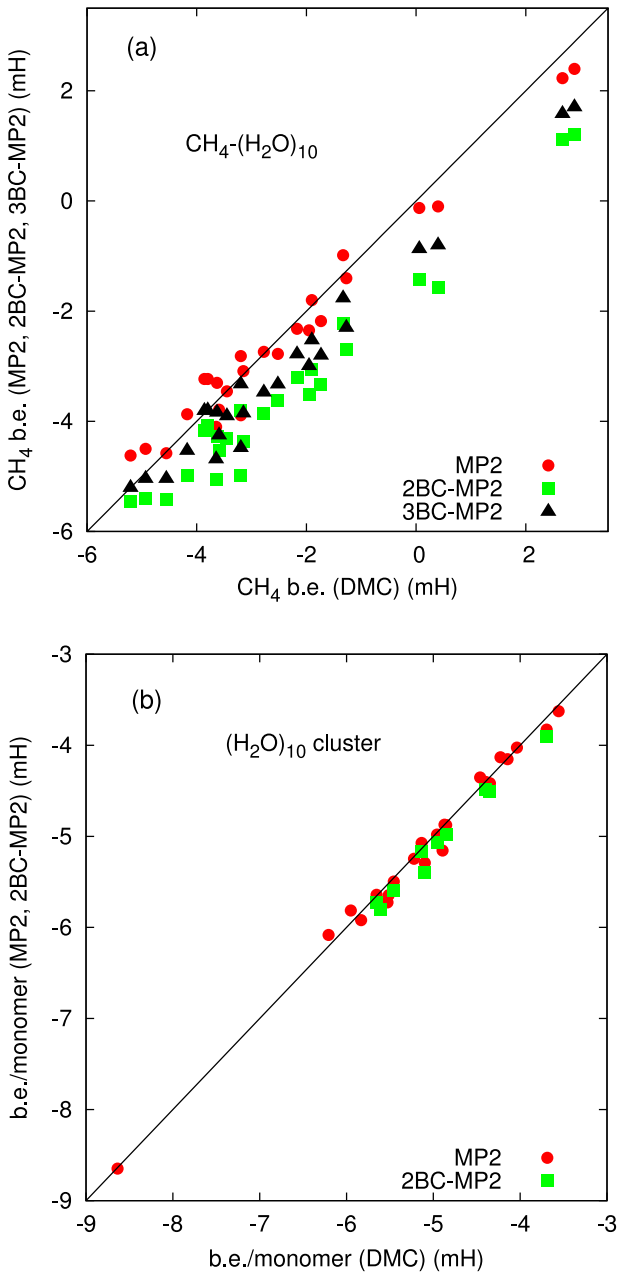

FIG. 4. Binding energy ΔE_{CH_4} of the methane monomer (upper panel) and binding energy per monomer ΔE_{bind}^{empty} of the empty water cluster (lower panel) for 25 configurations of the CH_4-$(H_2O)_{10}$ cluster. Panels show parity plots of uncorrected MP2 (red circles), MP2 with 2-body δCCSD(T) corrections (green squares), and MP2 with 2- and 3-body δCCSD(T) corrections (black triangles) *vs* DMC values.

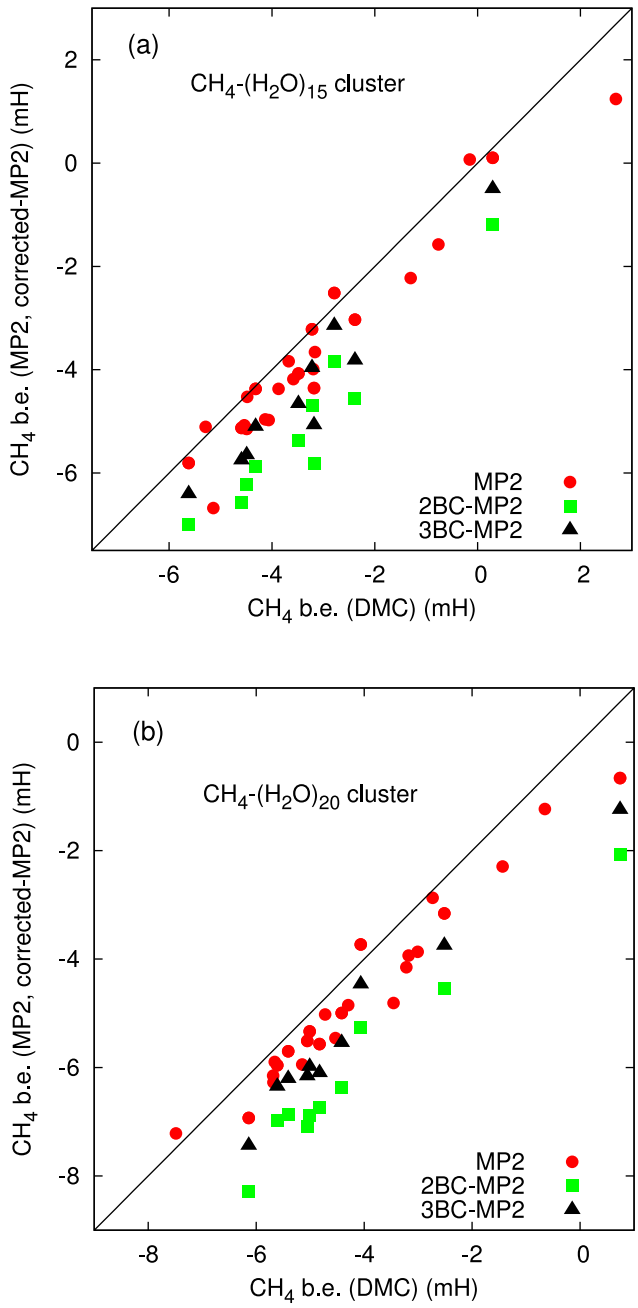

FIG. 5. Methane binding energy ΔE_{CH_4} in clusters CH_4-$(H_2O)_{15}$ (upper panel) and CH_4-$(H_2O)_{20}$ (lower panel). Parity plots compare values from MP2 (red circles), 2-body-corrected MP2 (green squares), and 2- and 3-body-corrected MP2 (black triangles) with values from DMC.

ways agree fairly closely with the QMC values, the mean differences for $n = 10$, 15, and 20 being -0.02, -0.52, and -0.57 mE_h, and the rms fluctuations of these differences being 0.38, 0.48, and 0.40 mE_h. However, the corrected values indicate that this agreement benefits from cancelation of errors. If we make only 2-body δCCSD(T) corrections, then the mean differences between the 2-body corrected E-MP2 values and the QMC values are -1.1, -1.7, and -1.9 mE_h for the three clusters. On the other hand with both 2- and 3-body corrections, the mean differences are -0.7, -1.0, and -1.1 mE_h. Just as for the CH_4-$(H_2O)_5$ cluster, the 2-body corrections are always negative and the 3-body corrections always positive. Since the sum of 2- and 3-body corrections does not completely restore agreement with DMC, it seems that even beyond-3-body corrections to MP2 may also be significant. As in the case of CH_4-$(H_2O)_5$, the agreement of MP2 values with DMC for ΔE_{bind}^{empty} is very close for the CH_4-$(H_2O)_{10}$ cluster, the mean difference and rms fluctuation being only -0.04 and 0.10 mE_h. The introduction of 2-body δCCSD(T) corrections slightly worsens the agreement, the mean difference now becoming -0.14 mE_h, and the rms fluctuation remaining essentially the same. However, these differences are negligible for most practical purposes.

IV. QMC AND MP2 FOR METHANE HYDRATE CRYSTAL

Having learnt something about the accuracy of MP2 for methane-water clusters, we now want to see how it performs for the methane hydrate crystal in the sI structure. The MP2 values of the characteristic energies ΔE_{CH_4} and ΔE_{bind}^{empty} are calculated by the embedding approximation, whose accurate reproduction of standard MP2 was demonstrated above, and these will be compared with the recently reported QMC values. The sI methane hydrate crystal has cubic symmetry, and the unit cell contains 46 H_2O monomers and 8 CH_4 monomers in the unit cell. All the present calculations are performed at the experimental lattice parameter of 11.77 Å with the atomic positions of Ref. 38. In order to obtain ΔE_{CH_4} and ΔE_{bind}^{empty}, we performed all calculations on both the "filled" crystal with all methane monomers in place and the "empty" water network obtained by removing all methane monomers with the positions of all water monomers remaining unchanged.

The HF binding energies of the filled and empty crystals relative to isolated monomers were calculated using the VASP code,[80] as explained in Sec. II. Technical settings such as plane-wave cutoff, density of k-point sampling, etc., were systematically improved until the values of HF energy per monomer were converged to better than 20 μE_h per monomer. The resulting values of ΔE_{CH_4} and ΔE_{bind}^{empty} are reported below.

The embedded many-body procedure for computing the MP2 correlation energy in periodic boundary conditions is summarized in Sec. II, where we note that embedded 2-body terms are included only for pairs of monomers whose separation is less than a specified cutoff distance R_c. We must therefore ensure that the correlation energy is adequately converged with respect to both basis set and R_c. Throughout this work, we compute MP2 energies using F12 and AVQZ basis sets, which give convergence to be within ~100 μE_h/monomer. However, for the small contributions to correlation energy from well separated monomers, AVTZ basis sets should suffice. Our strategy is therefore as follows. We first perform all the embedding calculations with F12 and AVTZ, increasing R_c until acceptable convergence is achieved. We then repeat the calculations with F12 and AVQZ, but now the R_c value need only ensure that the difference between AVQZ and AVTZ values is acceptably converged. Our final MP2 correlation energy is then the converged AVTZ value plus the converged difference AVQZ—AVTZ.

TABLE I. Convergence of MP2 correlation components of ΔE_{CH_4} and ΔE_{bind}^{empty} (mE_h units) with respect to cutoff radius R_c and basis set. Values of correlation components are computed using 2-body truncated EMBE with AVTZ basis sets and F12. Basis-set corrections (changes of values on replacing AVTZ with AVQZ) are given in parentheses. Extrapolated values in final row are obtained by adding basis-set corrections and assuming that error due to finite R_c decays as $1/R_c^3$.

R_c (Å)	ΔE_{CH_4} (mE_h)	ΔE_{bind}^{empty} (mE_h)
5.0	−12.29 (−0.04)	−10.82 (−0.10)
7.5	−12.98 (−0.04)	−11.07 (−0.11)
9.0	−13.20	−11.13
Extrap	−13.3	−11.3

TABLE II. Binding energy ΔE_{CH_4} of the methane molecule (mE_h/CH_4 monomer) in the methane hydrate sI crystal and cohesive energy ΔE_{bind}^{empty} of the empty water lattice (mE_h/H_2O monomer) computed with diffusion Monte Carlo (DMC) and second-order Møller-Plesset (MP2). The MP2 values of ΔE_{CH_4} and ΔE_{bind}^{empty} are the sum of the Hartree-Fock (HF) and MP2 correlation (ΔMP2) contributions. DMC values are from Ref. 38.

	HF	ΔMP2	MP2	DMC
ΔE_{CH_4}	4.4	−13.3	−8.9	−8.9
ΔE_{bind}^{empty}	−10.4	−11.3	−21.7	−21.7

The results reported in Table I show that convergence with respect to basis set and R_c is readily achieved. We extrapolate the AVTZ values of the correlation parts of ΔE_{CH_4} and ΔE_{bind}^{empty} to their $R_c \to \infty$ limit by assuming that the long-distance interaction falls off with the usual dispersion dependence of $1/R^6$, which implies that the residual error due to the R_c cutoff falls off as $1/R_c^3$.[88] The table suggests that the uncertainty in this extrapolation is no more than ~0.1 mE_h. We take the basis-set correction to be the change on going from AVTZ to AVQZ, and the table indicates that this correction is almost independent of R_c. The table reports the extrapolated values of the correlation parts of ΔE_{CH_4} and ΔE_{bind}^{empty}, which we then add to the HF parts to obtain the final embedded MP2 values, reported in Table II, where we also include the DMC values from Ref. 38. The agreement between the E-MP2 and DMC values of both ΔE_{CH_4} and ΔE_{bind}^{empty} is remarkably, and perhaps to some extent fortuitously, close, as we discuss further in Sec. V.

V. DISCUSSION AND CONCLUSIONS

Methane-water systems are challenging for commonly used electronic-structure methods. In order to make useful predictions, these methods need to achieve better than chemical accuracy (~1 kcal/mol) for the two energies focused on here: the methane binding energy ΔE_{CH_4} and the cohesive energy per monomer ΔE_{bind}^{empty} of the water subsystem. In the methane hydrate crystal, each CH_4 is surrounded by 20 H_2O neighbours, so that chemical accuracy for ΔE_{CH_4} requires an accuracy of 0.05 kcal/mol ≃ 0.1 mE_h ≃ 2 meV for the interaction with each H_2O neighbor. The same is true of methane in aqueous solution, where the hydration number is also *ca.* 20.[10] Our main aims in this work have been to show that QMC can deliver benchmarks of this accuracy or better for a range of methane-water clusters and to use the benchmarks to assess the accuracy of MP2. In the course of doing this, we have also shown the effectiveness of a recently developed embedding scheme for MP2 calculations,[66,67] which allows us to compare MP2 and QMC for the hydrate crystal.

Our strategy has been to start with the CH_4-H_2O dimer and then to work through a number of CH_4-$(H_2O)_n$ clusters up to $n = 20$, in each case taking 25 configurations (100 in the case of the dimer) from MD simulations to avoid any limitation to equilibrium configurations. The statistical accuracy of the QMC benchmarks is always much greater than that demanded by chemical accuracy for ΔE_{CH_4} and ΔE_{bind}^{empty}. For the dimer, QMC is in essentially perfect accord with CCSD(T) energies,

and we showed that MP2 is underbound by a small but not completely negligible amount. Nevertheless, the MP2 values of ΔE_{CH_4} and ΔE_{bind}^{empty} agree with the QMC values to rather better than 1 kcal/mol for all the clusters, and the agreement is even closer for the methane hydrate crystal. This is an important outcome, because it indicates that for this type of system, MP2 is almost as good as QMC as a source of benchmarks.

For the CH_4-$(H_2O)_n$ clusters ($n \geq 5$), we have not attempted to perform direct CCSD(T) calculations near the basis-set limit, but we have presented results for the 2- and 3-body δCCSD(T) corrections to MP2. As expected from other recent work,[37] these corrections are significant, and if only 2-body corrections are made, the agreement with QMC deteriorates appreciably. However, as also expected, the 3-body corrections partially cancel the 2-body corrections, so that the agreement is partially restored. This means that the good agreement between MP2 and QMC relies significantly on cancelation of 2- and 3-body errors.

In the course of this work, we have shown that the embedded many-body scheme of Bygrave *et al.*[66] provides an efficient way of performing MP2 calculations on large methane-water clusters and the methane-hydrate crystal, which appears to give results that are almost indistinguishable from those of standard MP2. As in our recent embedded-MP2 work on water clusters and ice structures,[67] we use the embedded many-body expansion truncated at 2-body level to compute only the correlation energy, with the Hartree-Fock energy computed by standard methods. We pointed out in our work on water systems[67] that this way of applying MBE is expected to be accurate for MP2 since by definition, MP2 includes electron correlation only up to the 2-body level, so that truncation at this level loses almost nothing. It is worth noting that very recent developments have made it possible to perform direct MP2 calculations on periodic systems,[50,51] and it appears to us that such direct calculations on the methane hydrate crystal should be feasible. Indeed, MP2-based Monte Carlo and even MD simulations on solid and liquid methane-water systems could probably be performed now with appropriate computer resources.

We plan to use the benchmarks reported here to test and calibrate both force fields and DFT approximations. Up to now, the force fields used to model methane-water systems have mostly been constructed by fitting to experimental data for quantities such as the solubility of methane gas in water.[5] It has sometimes been thought[6] that DFT methods must be more accurate than such empirical force fields, even though these methods have difficulty in describing dispersion interactions, which are known to be vital. The dangers of relying too heavily on presently available DFT methods are highlighted by recent work on methane hydrate and the isolated clathrate cage,[37,38] which reveal errors in ΔE_{CH_4} of over 5 kcal/mol with both standard and dispersion-inclusive approximations. It has been suggested that the many-body exchange-overlap effects known to be important in water systems may be implicated in some of these errors.[89] QMC benchmarks have already proved their worth for analysing the errors of DFT approximations in water systems,[56,61,64,65] and we expect the present benchmarks to be equally useful for methane-water systems.

ACKNOWLEDGMENTS

The project employed the facilities of the Oak Ridge Leadership Computing Facility at ORNL, which is supported by the Office of Science of the DOE under Contract No. DE-AC05-00OR22725. We thank G. Csányi and A. Nichol for permission to use configurations from the MD simulations described in Sec. III A and S. J. Cox for permission to use configurations from the MD simulations described in Sec. III B. We also thank C. Taylor for technical assistance with the embedding calculations and K. D. Jordan for many discussions.

[1]C. A. Koh, E. D. Sloan, A. K. Sum, and D. T. Wu, "Fundamentals and applications of gas hydrates," Annu. Rev. Chem. Biomol. Eng. **2**, 237–257 (2011).
[2]A. K. Sum, D. T. Wu, and K. Yasuoka, MRS Bull. **36**, 205 (2011).
[3]C. A. Koh, A. K. Sum, and E. D. Sloan, J. Nat. Gas Sci. Eng. **8**, 132 (2012).
[4]K. A. Kvenvolden, Rev. Geophys. **31**, 173, doi:10.1029/93RG00268 (1993).
[5]H. Docherty, A. Galindo, C. Vega, and E. Sanz, J. Chem. Phys. **125**, 074510 (2006).
[6]J. L. Li, R. Car, C. Tang, and N. S. Wingreen, Proc. Natl. Acad. Sci. U. S. A. **104**, 2626 (2007).
[7]W. L. Jorgensen, J. Gao, and C. Ravimohan, J. Phys. Chem. **89**, 3470 (1985).
[8]P. H. K. De Jong, J. E. Wilson, G. W. Neilson, and A. D. Buckingham, Mol. Phys. **91**, 99 (1997).
[9]H. A. Scheraga, J. Biomol. Struct. Dyn. **16**, 447 (1998).
[10]S. F. Dec, K. E. Bowler, L. L. Stadterman, C. A. Koh, and D. Sloan, J. Am. Chem. Soc. **128**, 414 (2006).
[11]B. J. Anderson, J. W. Tester, G. P. Borghi, and B. L. Trout, J. Am. Chem. Soc. **127**, 17852 (2005).
[12]C. Moon, R. W. Hawtin, and P. M. Rodger, Faraday Discuss. **136**, 367 (2007).
[13]H. Jiang, K. D. Jordan, and C. E. Taylor, J. Phys. Chem. B **111**, 6486 (2007).
[14]H. Jiang, E. M. Myshakin, K. D. Jordan, and R. P. Warzinski, J. Phys. Chem. B **112**, 10207 (2008).
[15]R. W. Hawtin, D. Quigley, and P. M. Rodger, Phys. Chem. Chem. Phys. **10**, 4853 (2008).
[16]J. F. Zhang, Y. Yang, E. Nakagawa, M. Rivero, S. K. Choi, and P. M. Rodger, J. Phys. Chem. B **112**, 10608 (2008).
[17]E. M. Myshakin, H. Jiang, R. P. Warzinski, and K. D. Jordan, J. Phys. Chem. A **113**, 1913 (2009).
[18]M. R. Walsh, C. A. Koh, E. D. Sloan, A. K. Sum, and D. T. Wu, Science **326**, 1095 (2009).
[19]L. C. Jacobson and V. Molinero, J. Phys. Chem. B **114**, 7302 (2010).
[20]L. C. Jacobson, W. Hujo, and V. Molinero, J. Am. Chem. Soc. **132**, 11806 (2010).
[21]S. Liang and P. G. Kusalik, Chem. Phys. Lett. **494**, 123 (2010).
[22]L. Jensen, K. Thomsen, N. von Solms, S. Wierzchowski, M. R. Walsh, C. A. Koh, E. D. Sloan, D. T. Wu, and A. K. Sum, J. Phys. Chem. B **114**, 5775 (2010).
[23]M. R. Walsh, G. T. Beckham, C. A. Koh, E. D. Sloan, D. T. Wu, and A. K. Sum, J. Phys. Chem. C **115**, 21241 (2011).
[24]S. Liang, D. Rozmanov, and P. G. Kusalik, Phys. Chem. Chem. Phys. **13**, 19856 (2011).
[25]R. Sakamaki, A. K. Sum, T. Narumi, R. Ohmura, and K. Yasuoka, J. Chem. Phys. **134**, 144702 (2011).
[26]G. J. Guo and P. M. Rodger, J. Phys. Chem. B **117**, 6498 (2013).
[27]P. Pirzadeh and P. G. Kusalik, J. Am. Chem. Soc. **135**, 7278 (2013).
[28]T. Ikeda and K. Terakura, J. Chem. Phys. **119**, 6784 (2003).
[29]A. Lenz and L. Ojamae, J. Phys. Chem. A **115**, 6169 (2011).
[30]M. Hiratsuka, R. Ohmura, A. K. Sum, and K. Yasuoka, J. Chem. Phys. **136**, 044508 (2012).
[31]M. Hiratsuka, R. Ohmura, A. K. Sum, and K. Yasuoka, J. Chem. Phys. **137**, 144306 (2012).
[32]L. Rossato, F. Rossetto, and P. L. Silvestrelli, J. Phys. Chem. B **116**, 4552 (2012).
[33]Y. Zhao and D. G. Truhlar, J. Chem. Theory Comput. **1**, 415 (2005).
[34]B. M. Austin, D. Y. Zubarev, and W. A. Lester, Chem. Rev. **112**, 263 (2012).
[35]W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, Rev. Mod. Phys. **73**, 33 (2001).
[36]R. J. Needs, M. D. Towler, N. D. Drummond, and P. L. Ríos, J. Phys.: Condens. Matter **22**, 023201 (2010).

[37]M. J. Deible, O. Tuguldur, and K. D. Jordan, J. Phys. Chem. B **118**, 8257 (2014).
[38]S. J. Cox, M. D. Towler, D. Alfè, and A. Michaelides, J. Chem. Phys. **140**, 174703 (2014).
[39]H. Popkie, H. Kistenmacher, and E. Clementi, J. Chem. Phys. **59**, 1325 (1973).
[40]O. Matsuoka, E. Clementi, and M. Yoshimine, J. Chem. Phys. **64**, 1351 (1976).
[41]T. Helgaker, P. Jorgensen, and J. Olsen, *Molecular Electronic Structure Theory* (Wiley, New York, 2000).
[42]F. F. Wang, G. Jenness, W. A. Al-Saidi, and K. D. Jordan, J. Chem. Phys. **132**, 134303 (2010).
[43]L. Goerigk and S. Grimme, Phys. Chem. Chem. Phys. **13**, 6670 (2011).
[44]Z. T. Cao, J. W. Tester, and B. L. Trout, J. Chem. Phys. **115**, 2550 (2001).
[45]Q. S. Du, P. J. Liu, and J. Deng, J. Chem. Theory Comput. **3**, 1665 (2007).
[46]Y. Liu, J. J. Zhao, F. Y. Li, and Z. F. Chen, J. Comput. Chem. **34**, 121 (2013).
[47]A. Grüneis, M. Marsman, and G. Kresse, J. Chem. Phys. **133**, 074107 (2010).
[48]L. Maschio, B. Civalleri, P. Ugliengo, and A. Gavezzotti, J. Phys. Chem. A **115**, 11179 (2011).
[49]F. Goeltl, A. Grüneis, T. Bučko, and J. Hafner, J. Chem. Phys. **137**, 114111 (2012).
[50]M. Del Ben, J. Hutter, and J. VandeVondele, J. Chem. Theory Comput. **8**, 4177 (2012).
[51]M. Del Ben, M. Schoenherr, J. Hutter, and J. VandeVondele, J. Phys. Chem. Lett. **4**, 3753 (2013).
[52]G. H. Booth, A. Grüneis, G. Kresse, and A. Alavi, Nature **493**, 365 (2013).
[53]I. G. Gurtubay and R. J. Needs, J. Chem. Phys. **127**, 124306 (2007).
[54]J. Ma, D. Alfè, A. Michaelides, and E. Wang, J. Chem. Phys. **130**, 154303 (2009).
[55]B. Santra *et al.*, Phys. Rev. Lett. **107**, 185701 (2011).
[56]M. J. Gillan, F. R. Manby, M. D. Towler, and D. Alfè, J. Chem. Phys. **136**, 244105 (2012).
[57]M. Dubecky, P. Jurečka, R. Derian, P. Hobza, M. Otyepka, and L. Mitas, J. Chem. Theory Comput. **9**, 4287 (2013).
[58]A. Ambrosetti, D. Alfè, R. A. DiStasio, and A. Tkatchenko, J. Phys. Chem. Lett. **5**, 849 (2014).
[59]A. Benali, L. Shulenburger, N. A. Romero, J. Kim, and O. A. von Lilienfeld, J. Chem. Theory Comput. **10**, 3417 (2014).
[60]B. Santra, A. Michaelides, M. Fuchs, A. Tkatchenko, C. Filippi, and M. Scheffler, J. Chem. Phys. **129**, 194111 (2008).
[61]D. Alfè, A. P. Bartók, G. Csányi, and M. J. Gillan, J. Chem. Phys. **138**, 221102 (2013).
[62]M. J. Gillan, D. Alfè, A. P. Bartók, and G. Csányi, J. Chem. Phys. **139**, 244504 (2013).
[63]F. Wang, M. Deible, and K. D. Jordan, J. Phys. Chem. A **117**, 7606 (2013).
[64]D. Alfè, A. P. Bartók, G. Csányi, and M. J. Gillan, J. Chem. Phys. **141**, 014104 (2014).
[65]M. A. Morales, J. R. Gergely, J. McMinis, J. M. McMahon, J. Kim, and D. M. Ceperley, J. Chem. Theory Comput. **10**, 2355 (2014).
[66]P. J. Bygrave, N. L. Allan, and F. R. Manby, J. Chem. Phys. **137**, 164102 (2012).
[67]M. J. Gillan, D. Alfè, P. J. Bygrave, C. R. Taylor, and F. R. Manby, J. Chem. Phys. **139**, 114101 (2013).
[68]R. J. Needs, M. D. Towler, N. D. Drummond, and P. López-Ríos, Casino 2.12 User Manual, 2013.
[69]L. Mitaš, E. L. Shirley, and D. M. Ceperley, J. Chem. Phys. **95**, 3467 (1991).
[70]J. R. Trail and R. J. Needs, J. Chem. Phys. **122**, 014112 (2005).
[71]J. R. Trail and R. J. Needs, J. Chem. Phys. **122**, 174109 (2005).
[72]P. Giannozzi *et al.*, J. Phys.: Condens. Matter **21**, 395502 (2009).
[73]D. Alfè and M. J. Gillan, Phys. Rev. B **70**, 161101 (2004).
[74]H.-J. Werner *et al.*, MOLPRO, version 2012.1, a package of *ab initio* programs, 2012, see http://www.molpro.net.
[75]H. J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, and M. Schutz, Wiley Interdiscip. Rev.: Comput. Mol. Sci. **2**, 242 (2012).
[76]W. Klopper, F. R. Manby, S. Ten-No, and E. F. Valeev, Int. Rev. Phys. Chem. **25**, 427 (2006).
[77]T. H. Dunning, J. Chem. Phys. **90**, 1007 (1989).
[78]R. A. Kendall, T. H. Dunning, and R. J. Harrison, J. Chem. Phys. **96**, 6796 (1992).
[79]R. Polly, H. J. Werner, F. R. Manby, and P. J. Knowles, Mol. Phys. **102**, 2311 (2004).
[80]G. Kresse and J. Furthmüller, Phys. Rev. B **54**, 11169 (1996).
[81]R. J. Wheatley and S. L. Price, Mol. Phys. **69**, 507 (1990).
[82]A. D. Becke, Phys. Rev. A **38**, 3098 (1988).
[83]C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).
[84]Since we use BLYP without correcting for non-local electron correlation (dispersion), an accurate approximation to the real-world system is not necessarily expected. However, our aim here is simply to create a set of CH_4-H_2O configurations spanning a wide range of C-O separations and molecular orientations.
[85]We take the equilibrium O–H bond length and H-O-H angle in the H_2O monomer to be 0.958 Å and 104.5°, and the C–H bond-length in the CH_4 monomer to be 1.088 Å. The equilibrium geometry of the H_2O monomer is restored in two steps: first, the O–H bonds are restored to their equilibrium length, keeping their direction fixed; second, the H-O-H angle is restored to its equilibrium value, keeping the direction of the H-O-H bisector fixed. The position of the O atom is held fixed throughout. For the CH_4 monomer, a sequence is adopted for the four H atoms, and the same procedure as for the H_2O monomer is applied to the first two H atoms. The lengths and directions of the other two C–H bonds are then adjusted to give the tetrahedral equilibrium geometry.
[86]J. L. F. Abascal, E. Sanz, R. G. Fernández, and C. Vega, J. Chem. Phys. **122**, 234511 (2005).
[87]J. S. Tse, M. L. Klein, and I. R. McDonald, J. Chem. Phys. **81**, 6146 (1984).
[88]For large R_c, the residual error is approximately proportional to the 3-dimensional integral of $1/r^6$ over the region lying beyond the radius R_c, this integral being $4\pi \int_{R_c}^{\infty} dr r^2 \cdot r^{-6} = 4\pi/3R_c^3$.
[89]M. J. Gillan, J. Chem. Phys. **141**, 224106 (2014).