Molecular dimers of methane clathrates: *ab initio* potential energy surfaces and variational vibrational states

Michael P. Metz,[1] Krzysztof Szalewicz,[1, *] János Sarka,[2]

Roland Tóbiás,[3, 4] Attila G. Császár,[3, 4] and Edit Mátyus[3, †]

[1]*Department of Physics and Astronomy,*

University of Delaware, Newark, Delaware 19716

[2]*Department of Chemistry and Biochemistry,*

Texas Tech University, Lubbock, Texas, 79409-1061, USA

[3]*Institute of Chemistry, ELTE Eötvös Loránd University,*

Pázmány Péter sétány 1/A, Budapest, H-1117, Hungary

[4]*MTA-ELTE Complex Chemical Systems Research Group,*

P.O. Box 32, H-1518 Budapest 112, Hungary

[*] szalewic@udel.edu

[†] matyus@chem.elte.hu

S1. FITTING METHODOLOGY

The PESs of this study were generated using the automated methodology described in Ref. [1], implemented in the autoPES software package [2]. In this subsection of the SI, we briefly describe the procedure as applied to the present work. Starting from the monomer geometries, the process includes the following major steps: (a) generation of the asymptotic component of the PES, (b) generation of a grid of close-range dimer configurations and of the corresponding *ab initio* computed interaction energies (referred to as grid points), (c) fitting the data with a global analytic function, and (d) iterative quality assessment and improvement of the PES.

A grid of about $10\,000$ long-range dimer configurations is selected and the corresponding interaction energies are computed using the COM-COM multipole expansion described in Sec. II.A of the main paper. These grid points are chosen by autoPES to have as short dimer separations as possible, while being sufficiently separated that charge overlap effects can be neglected. In the region where charge overlap effects are significant, which extends to about 1.5 times the radial van der Waals minimum separation, the asymptotic expansion is not a good approximation of interaction energies and supermolecular (or SAPT) calculations are used instead.

Dimer-grid configurations for the close-range CCSD(T) computations arei selected using the weighted random procedure described in Sec. II of Ref. [1]. In the first iteration, we used the OPLS-AA force field of Jorgensen and co-workers [3] to determine the weights and to approximately locate the minima on the surface in order to place a higher density of grid points there. In subsequent iterations of the grid generation algorithm, the OPLS-AA force field was replaced with the fitted PES from the previous iteration.

After the close-range and asymptotic grids are generated and the corresponding interaction energies are computed, the data is fitted with an analytic function V of the form

$$V = \sum_{a \in A} \sum_{b \in B} u_{ab}(r_{ab}) + V_{\text{ind}}(A, B), \tag{S1}$$

where a (b) go over the sets A (B) of fitting sites in monomers A (B) (the polarizable

2

site is not included, see below). The site-site function u_{ab} is of the form

$$u_{ab}(r_{ab}) = \left[1 + \sum_{i=1}^{3} a_i^{ab}(r_{ab})^i\right] e^{\alpha^{ab} - \beta^{ab} r_{ab}} + \frac{q_a q_b}{r_{ab}} - \sum_{n=6,8} f_n(\delta_n^{ab}, r_{ab}) \frac{C_n^{ab}}{(r_{ab})^n}, \quad (S2)$$

where f_n are the Tang–Toennies damping functions [4],

$$f_n(\delta, r) = 1 - e^{-\delta r} \sum_{m=0}^{n} \frac{(\delta r)^m}{m!}. \quad (S3)$$

The sets A and B include all atoms in the respective monomers, and in addition include off-atomic fitting sites. As in earlier work [5–7], not all pairs of sites carry all the interactions. In particular, the polarization site carries only the polarization interaction, the exponential terms are present for all site-site pairs, except those involving the polarization site, the partial charges are placed only on atoms and on the two off-plane sites on the side of hydrogens (D1 sites) in the case of water but on all sites in the case of methane, and the dispersion interactions are only between atoms.

The term V_{ind} is a polarization model consisting of a *single* polarizable site with isotropic polarizability $\alpha_{\bar{a}}$ located near the COM of monomer A (and similarly for monomer B). We use the bar to indicate that site $\bar{a}$ is in general different from the sites in Eq. (S2). The polarization site responds to the external field by creating an induced dipole moment,

$$\boldsymbol{\mu}_{\bar{a}}^{\text{ind}} = \alpha_{\bar{a}} \left[\boldsymbol{\mathcal{E}}_{\bar{a}} + \boldsymbol{T} \boldsymbol{\mu}_{\bar{b}}^{\text{ind}}\right], \quad (S4)$$

where $\boldsymbol{\mathcal{E}}_{\bar{a}}$ is the damped electric field at the polarizable site of monomer A due to all permanent point charges q_b of monomer B,

$$\boldsymbol{\mathcal{E}}_{\bar{a}} = \sum_{b \in B} f_1(\delta_1^{\bar{a}b}, r_{\bar{a}b}) \frac{q_b \boldsymbol{r}_{\bar{a}b}}{r_{\bar{a}b}^3}, \quad (S5)$$

and $\boldsymbol{T}$ is the damped dipole-dipole 3×3 interaction tensor

$$\boldsymbol{T} = 3 f_3(\delta_{\text{p}}^{\bar{a}\bar{b}}, r_{\bar{a}\bar{b}}) \frac{\boldsymbol{r}_{\bar{a}\bar{b}} \otimes \boldsymbol{r}_{\bar{a}\bar{b}}}{r_{\bar{a}\bar{b}}^5} - f_3(\delta_{\text{p}}^{\bar{a}\bar{b}}, r_{\bar{a}\bar{b}}) \frac{1}{r_{\bar{a}\bar{b}}^3}, \quad (S6)$$

with $\boldsymbol{r}_{\bar{a}\bar{b}}$ denoting the displacement vector from the polarizable center of monomer

3

A to that of monomer B, $\boldsymbol{r}_{\bar{a}b}$ denoting the displacement from the polarizable center of monomer A to the site b of monomer B, and $\otimes$ denoting the Kronecker product. Equation (S4) is iterated to convergence. The polarization energy is then

$$V_{\text{ind}}(A, B) = -\frac{1}{2} \left(\boldsymbol{\mu}_{\bar{a}}^{\text{ind}} \cdot \boldsymbol{\mathcal{E}}_{\bar{a}} + \boldsymbol{\mu}_{\bar{b}}^{\text{ind}} \cdot \boldsymbol{\mathcal{E}}_{\bar{b}} \right) . \tag{S7}$$

The formulas given above can be easily extended to clusters larger than the dimer and the fit subroutines produced by autoPES will do it automatically if applied to such systems. The polarizable center for methane is located directly at the COM, while for water the polarizable site is slightly displaced from the COM, in the same location as used in Ref. [6].

For methane, four off-atomic sites were used, placed near the faces of the tetrahedron formed by the four hydrogen atoms, in such a way that the line through the site and the carbon atom passes through the center of a given face. To minimize the fitting error, the distance between the off-atomic sites and the center of the molecule was optimized simultaneously with the other fit parameters. For water, which requires more off-atomic sites than methane to achieve the same fit accuracy due to its lower symmetry, the 22 off-atomic sites from Ref. [7] were used. This, together with the atomic sites, makes a total of 25 sites plus the polarization site. Several of these sites are symmetry related, so that the total number of symmetry-distinct sites not counting the polarizable site is 8 (thus, '8s' in CCpol-8s). The positions of all the off-atomic sites are given in the files *_fit_report.txt included in the SI.

We will denote the sum over a and b of the first term in Eq. (S2) as V_{exp}, the sum over the second term as V_{elst}, and the sum over the final term plus V_{ind} as $V_{\text{asym}}^{(2)}$. The inclusion of the induction model V_{ind} has a minor influence on the two-body PESs, and primarily serves to enable the PES to reproduce many-body polarization effects [8–10]. While we do not make use of this feature in the present work, we nonetheless included the model to extend the future applicability of the PESs, in particular to simulations of methane clathrates.

The partial charges q_x and the distributed induction plus dispersion coefficients C_n^{ab} are fit to the interaction energies on the asymptotic grid obtained from the COM-COM multipole expansion. In the latter case, the values of V_{ind} are subtracted

from the sum of the COM-COM induction and dispersion energies before the fit. Because the terms which depend on all other free parameters in Eq. (S2) decay exponentially with distance, this ensures the accuracy of the PES in the asymptotic regime. The parameters δ_n^{ab}, $\delta_1^{\bar{a}b}$, $\delta_p^{\bar{a}\bar{b}}$, α^{ab}, β^{ab}, and a_i^{ab} are then fit to the total interaction energy using the short-range grid, see Sec. VI of Ref. [1] for details of the fitting procedure. In the case of PESs based on SAPT, the energy decomposition given by that theory can be leveraged to fit the damping parameters in V_{elst} and $V_{\text{asym}}^{(2)}$ terms to individual components of the interaction energy [1]. Because the present work uses supermolecular calculations, we instead fit all parameters simultaneously to the total interaction energy. For the isotropic polarizabilities, we have used the experimental values of 9.922 a.u. for water [11] and 17.68 a.u. for methane [12]. The autoPES code has also the capability of computing *ab initio* distributed asymptotic expansions [13, 14], but it results in anisotropic atom-atom functions which cannot be used in most of molecular simulation software.

For all cases where multiple atomic or off-atomic sites are equivalent due to the geometric symmetry of the molecules, the parameters associated with all sites of a given symmetry type are constrained to have identical values. There are 8 such symmetry-distinct site types in the case of water and 3 in the case of methane (plus the polarization site in each case).

We constrain the C_n^{ab} parameters to be positive and use the geometric mean combination rule $C_n^{ab} = \sqrt{C_n^a C_n^b}$, where each of the symmetry-distinct fitting sites is associated with a single parameter C_n^a or C_n^b, rather than using a separate free parameter C_n^{ab} for each pair of sites. Similarly, the α^{ab}, β^{ab}, and all δ parameters are constrained to satisfy the arithmetic sum combination rule. Such constraints decrease the number of free parameters in the fit without significantly affecting accuracy.

S2. LEVEL OF SAPT

We specify here the level of symmetry-adapted perturbation theory (SAPT) used in Table IV of the main paper. This level includes terms approximately equivalent to the supermolecular MBPT through fourth order. The SAPT corrections are

partitioned into four physically interpretable energy components. For a complete description of the terms, see Ref. [15].

$$E_{\text{int}}^{\text{SAPT}} = E_{\text{elst}}^{(1)} + E_{\text{exch}}^{(1)} + E_{\text{indx}} + E_{\text{dispx}}^{(2)} \tag{S8}$$

$$E_{\text{elst}}^{(1)} = E_{\text{elst}}^{(10)} + E_{\text{elst,resp}}^{(12)} + E_{\text{elst,resp}}^{(13)} \tag{S9}$$

$$E_{\text{exch}}^{(1)} = E_{\text{exch}}^{(10)} + \epsilon_{\text{exch}}^{(1)}(\text{CCSD}) \tag{S10}$$

$$E_{\text{indx}}^{(2)} = E_{\text{ind,resp}}^{(20)} + E_{\text{exch}-\text{ind,resp}}^{(20)} + {}^{t}E_{\text{ind}}^{(22)} + {}^{t}E_{\text{exch}-\text{ind}}^{(22)} \tag{S11}$$

$$E_{\text{dispx}}^{(2)} = E_{\text{disp}}^{(20)} + E_{\text{disp}}^{(21)} + E_{\text{disp}}^{(22)} + E_{\text{exch}-\text{disp}}^{(20)} \tag{S12}$$

For polar systems, the term $\delta_{\text{int,resp}}^{\text{HF}}$ is usually included in the indx component (then this component becomes of infinite order in the interaction operator V since $\delta_{\text{int,resp}}^{\text{HF}}$ is of infinite order).

S3. INTERNAL COORDINATES

During this work, two different sets of internal coordinates were utilized. In the PES construction with the autoPES code, space-fixed Euler angles were used, while during the variational computations with the GENIUSH code, body-fixed Euler angles were used (in addition to the intermolecular COM-COM distance R).

The R and the space-fixed Euler angles used in autoPES and reported in Table S1 are defined in Ref. [1]. The Cartesian coordinates corresponding to these sets of R and Euler angles are displayed in Table S2. The reference orientations of monomers are given in files *_fit_report.txt. The water monomer is in xz plane with O at a positive value on z axis and H's at negative z. The methane monomer has C and two H's in yz plane, one of these H's at a positive value on z axis and the other one has positive y and negative x coordinates. The positions of the other two H's follow from symmetry. The COM of monomer A is always at (0,0,0), and that of monomer B is at (0,0,R). In the WM19 surface, methane is monomer A, while water is monomer B.

In calculations of spectra, a different convention for the Euler angles (body-fixed Euler angles) was used, as defined in Ref. [16]. The reference orientations of monomers are similar as in autoPES except that the roles of x and y axes are inverted, i.e., the water monomer is in yz plane instead in xz as in autoPES, whereas the methane monomer has two H's in xz plane. Moreover, in the water–methane dimer, water is monomer A, while methane is monomer B. The dimer minima geometries and energies are, of course, identical in both conventions. The coordinates of minima in this convention are listed in Table XV of the main text. Note that in addition to the different convention of the coordinates (and also due to the swap of the monomers in the case of WM19 surface), the permutation of the hydrogen atoms in the minima provided in Table XV is different from those of Table S1 and S2.

TABLE S1. Interaction energies, in cm^{-1}, and coordinates, R in Å and the Euler angles in degrees (°), at the minima of the WW19, WM19, and MM19 PESs developed in this work using space-fixed Euler angles.

Label	E_{int}	R	β_A	γ_A	α_B	β_B	γ_B
WW19	-1790.86	2.91688	56.3774	90.0000	180.000	60.3956	0.00000
WM19(GM)	-367.82	3.45748	66.8866	277.135	62.5973	63.7919	180.000
WM19(SM)	-241.49	3.79488	0.72790	261.498	312.875	178.116	146.048
MM19	-190.23	3.64358	70.5288	30.0000	23.2888	0.00000	66.7112

TABLE S2. Cartesian coordinates (in angstroms) corresponding to the sets of R and Euler angles from Table S1. Column 2 gives the atomic number.

Label	Z	x	y	z
WW19	8	0.055703710843378	0.000000000000000	0.037041197452746
	1	-0.442109123938075	-0.771034745128542	-0.293988598533781
	1	-0.442109198907332	0.771034745128542	-0.293988485792667
	8	-0.058162409540960	0.000000008190134	2.949927847037623
	1	0.842521154003280	-0.000000084037312	3.324977231600114
	1	0.080725564003651	-0.000000045969591	1.984214093742098
WM19(GM)	6	0.000000000000000	0.000000000000000	0.000000000000000
	1	1.020918059173883	0.000000000000000	0.435741571172055
	1	0.067333442735276	0.129994675924625	-1.100323585185535
	1	-0.499932658933534	-0.964305760271743	0.228749071327614
	1	-0.588318842975625	0.834311084347118	0.435832942685867
	8	0.027620611764974	0.053279393787077	3.487020493730036
	1	-0.062491607610028	-0.120547432078482	2.531218689584005
	1	-0.375946884925503	-0.725188255964299	3.914822387350833
WM19(SM)	6	0.000000000000000	0.000000000000000	0.000000000000000
	1	0.014101680805128	0.000000000000000	1.109930422591132
	1	1.030251323337329	-0.154729335939914	-0.383125884998858
	1	-0.656165224089125	-0.819002290765656	-0.361699940297949
	1	-0.388187780053333	0.973731626705571	-0.365104597294326
	8	0.001496011662352	-0.001611327848430	3.728028288393649
	1	-0.762420094154190	0.188256249055357	4.304461764863195
	1	0.738673006159245	-0.162678678259404	4.346510667848477
MM19	6	0.000000000000000	0.000000000000000	0.000000000000000
	1	1.046536898071173	0.000000000000000	0.370006650440808
	1	-0.523268403134842	0.906327599358239	0.370006569289151
	1	-0.000000017209305	-0.000000129197469	-1.110019999997679
	1	-0.523268477727026	-0.906327470160769	0.370006780267719
	6	0.000000000000000	0.000000000000000	3.643584408799999
	1	0.000000124508219	0.000000053592837	4.753604408797676
	1	-1.046536933837382	0.00000441970052	3.273577859521560
	1	0.523268802892789	0.906327286977970	3.273577639681969
	1	0.523268006436374	-0.906327782540859	3.273577727198794

TABLE S3. Character table of the molecular symmetry group of the methane dimer. The table was generated using GAP according to instructions of Ref. [17]. Zero characters are labeled with "." to enhance readability.

	$\hat{E}$	(14)(23)	(14)(23)(58)(67)	(243)	(243)(58)(67)	(243)(687)	(234)(687)	(34)(78)(9,10)	(1423)(78)(9,10)	(1423)(5867)(9,10)	(15)(26)(37)(48)	(1548)(2637)	(15)(264837)	(15)(26)(38)(47)(9,10)	(1547)(2638)(9,10)	(15)(264738)(9,10)
$X_1(1)$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
$X_2(1)$	1	1	1	1	1	1	1	−1	−1	−1	−1	−1	−1	1	1	1
$X_3(1)$	1	1	1	1	1	1	1	−1	−1	−1	1	1	1	−1	−1	−1
$X_4(1)$	1	1	1	1	1	1	1	1	1	1	−1	−1	−1	−1	−1	−1
$X_5(2)$	2	2	2	−1	−1	2	−1	.	.	.	.	.	.	2	2	−1
$X_6(2)$	2	2	2	−1	−1	2	−1	.	.	.	.	.	.	−2	−2	1
$X_7(2)$	2	2	2	−1	−1	−1	2	.	.	.	−2	−2	1	.	.	.
$X_8(2)$	2	2	2	−1	−1	−1	2	.	.	.	2	2	−1	.	.	.
$X_9(2)$	4	4	4	1	1	−2	−2	.	.	.	.	.	.	.	.	.
$X_{10}(6)$	6	2	−2	3	−1	.	.	−2	.	2	.	.	.	.	.	.
$X_{11}(6)$	6	2	−2	3	−1	.	.	2	.	−2	.	.	.	.	.	.
$X_{12}(9)$	9	−3	1	.	.	.	.	−1	1	−1	−3	1	.	3	−1	.
$X_{13}(9)$	9	−3	1	.	.	.	.	−1	1	−1	3	−1	.	−3	1	.
$X_{14}(9)$	9	−3	1	.	.	.	.	1	−1	1	−3	1	.	−3	1	.
$X_{15}(9)$	9	−3	1	.	.	.	.	1	−1	1	3	−1	.	3	−1	.
$X_{16}(12)$	12	4	−4	−3	1	.	.	.	.	.	.	.	.	.	.	.

[1] M. P. Metz, K. Piszczatowski, and K. Szalewicz, J. Chem. Theory Comput. **12**, 5895 (2016).

[2] M. P. Metz, K. Piszczatowski, and K. Szalewicz, "*autoPES: Automatic Intermolecular Potential Energy Surface Generation Software*," `http://www.physics.udel.edu/~szalewic/SAPT/index.html` (2016).

[3] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, J. Am. Chem. Soc. **118**, 11225 (1996).

[4] K. T. Tang and J. P. Toennies, J. Chem. Phys. **80**, 3726 (1984).

[5] E. M. Mas, R. Bukowski, K. Szalewicz, G. C. Groenenboom, P. E. S. Wormer, and A. van der Avoird, J. Chem. Phys. **113**, 6687 (2000).

[6] R. Bukowski, K. Szalewicz, G. C. Groenenboom, and A. van der Avoird, J. Chem. Phys. **128**, 094313 (2008).

[7] W. Cencek, K. Szalewicz, C. Leforestier, R. van Harrevelt, and A. van der Avoird, Phys. Chem. Chem. Phys. **10**, 4716 (2008).

[8] U. Gora, R. Podeszwa, W. Cencek, and K. Szalewicz, J. Chem. Phys. **135**, 224102 (2011).

[9] U. Góra, W. Cencek, R. Podeszwa, A. van der Avoird, and K. Szalewicz, J. Chem. Phys. **140**, 194101 (2014).

[10] O. Akin-Ojo and K. Szalewicz, J. Chem. Phys. **138**, 024316 (2013).

[11] W. F. Murphy, J. Chem. Phys. **67**, 5877 (1977).

[12] H. Watson and K. Ramaswamy, Proc. Royal Soc. London A: Math., Phys. and Eng. Sci. **156**, 144 (1936).

[13] F. Rob and K. Szalewicz, Chem. Phys. Lett. **572**, 146 (2013).

[14] F. Rob and K. Szalewicz, Mol. Phys. **111**, 1430 (2013).

[15] K. Szalewicz, Wiley Interdis. Rev.: Comp. Mol. Sci. **2**, 254 (2012).

[16] J. Sarka, A. G. Csaszar, S. C. Althorpe, D. J. Wales, and E. Matyus, Phys. Chem. Chem. Phys. **18**, 22816 (2016).

[17] R. Schmied and K. Lehmann, J. Mol. Spectrosc. **226**, 201 (2004).