Study of hydrogen-molecule guests in type II clathrate hydrates using a force-matched potential model parameterised from ab initio molecular dynamics

Christian J. Burnham, Zdenek Futera, and Niall J. English

Study of hydrogen-molecule guests in type II clathrate hydrates using a force-matched potential model parameterised from *ab initio* molecular dynamics

Christian J. Burnham,[a)] Zdenek Futera,[b)] and Niall J. English[a)]
School of Chemical and Bioprocess Engineering, University College Dublin, Belfield, Dublin 4, Ireland

The force-matching method has been applied to parameterise an empirical potential model for water-water and water-hydrogen intermolecular interactions for use in clathrate-hydrate simulations containing hydrogen guest molecules. The underlying reference simulations constituted *ab initio* molecular dynamics (AIMD) of clathrate hydrates with various occupations of hydrogen-molecule guests. It is shown that the resultant model is able to reproduce AIMD-derived free-energy curves for the movement of a tagged hydrogen molecule between the water cages that make up the clathrate, thus giving us confidence in the model. Furthermore, with the aid of an umbrella-sampling algorithm, we calculate barrier heights for the force-matched model, yielding the free-energy barrier for a tagged molecule to move between cages. The barrier heights are reasonably large, being on the order of 30 kJ/mol, and are consistent with our previous studies with empirical models [C. J. Burnham and N. J. English, J. Phys. Chem. C **120**, 16561 (2016) and C. J. Burnham *et al.*, Phys. Chem. Chem. Phys. **19**, 717 (2017)]. Our results are in opposition to the literature, which claims that this system may have very low barrier heights. We also compare results to that using the more *ad hoc* empirical model of Alavi *et al.* [J. Chem. Phys. **123**, 024507 (2005)] and find that this model does very well when judged against the force-matched and *ab initio* simulation data. *Published by AIP Publishing.*
https://doi.org/10.1063/1.4999909

INTRODUCTION

A type II clathrate hydrate is a water-ice crystal with a structure comprising "large" and "small" polyhedral cavities, with hydrogen-bonded water molecules forming the vertex of each polyhedron.[1] They possess an interesting structure: the lattice can be thought of as formed from tessellating polyhedral "cages," wherein the cages can enclathrate guest molecular species. Type II (or "sII") clathrate consists of "large" cages of type $5^{12}6^4$ (i.e., having twelve pentagonal faces and six hexagonal ones) and small-cage 5^{12} dodecahedra, each composed of five water pentamers, with each cage holding various numbers of guest H_2 molecules.[1] A typical large-cage structure, featuring H_2-molecule guests, is shown in Fig. 1.

Although on short time scales the hydrogen-molecule guests appear trapped in the cages, on longer time scales, it is possible for them to hop between the cages such that they will eventually diffuse throughout the lattice. In order for the guest molecule to pass from one cage to another without breaking hydrogen bonds, it is required to thread either one of the water pentamers or hexamers making up the faces of the polyhedral cages. In practice, the energy barrier for a guest hydrogen molecule to pass through a water pentamer is much greater than for it to thread through a hexamer, and so in this paper, as with our previous studies,[2,3] we will exclusively focus on the movement of hydrogen molecules through water hexamers. This occurs exclusively between adjoining large $5^{12}6^4$ cages, the smaller dodecahedral cages having only pentagonal faces.

Trinh *et al.* used Density Functional Theory (DFT)-based Born-Oppenheimer molecular dynamics (MD), in conjunction with umbrella sampling, to calculate the free-energy profile and barriers for a guest hydrogen molecule to pass between two large cages in the clathrate structure.[4] These authors found quite low-energy barriers in the range of 10 kJ/mol (at 100 K). Furthermore, they argue that increasing temperature will act to further lower the barriers. In our two previous papers,[2,3] we used a somewhat similar approach to Trinh *et al.*, albeit with an empirical force-field, in order to explore both the classical behavior and nuclear-quantised behavior of H_2 guest molecules in the sII clathrate hydrate structure.

Our first study used an umbrella sampling technique in conjunction with classical and centroid path-integral (PI) MD to ascertain the free-energy barrier heights for guest H_2 molecules to pass between adjoining large cages, through the hexagonal faces.[2] Here, we used the empirical model of Alavi *et al.*[5] for H_2-H_2 and H_2-water interactions, whilst PI-MD accounts for the effect of nuclear quantisation on equilibrium averages, such as free-energy barriers. We found that, contrary to the study of Trinh *et al.*, our empirical model did not predict low-energy barriers for guest hydrogen molecule transport between cages; in contrast, we found

---

[a)]Authors to whom correspondence should be addressed: christian.burnham@ucd.ie and niall.english@ucd.ie
[b)]Present address: Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, United Kingdom.
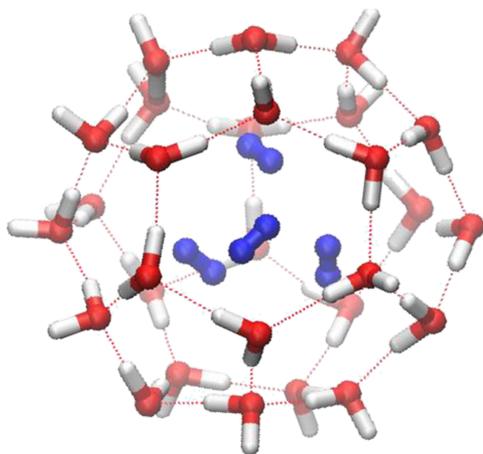
FIG. 1. Clathrate-hydrate $5^{12}6^4$ large cage with quadruple (n = 4) occupation of H_2-molecule guests. This structure favours a tetrahedral configuration for the guest molecules, with each H_2 molecule "sitting" behind one of the four hexagonal faces.

barriers around three times larger.[2] Of course, the sizable discrepancy between our results could be primarily due to differences in force-fields/functionals, but the differences were very large; indeed, it would be somewhat surprising if reasonably parameterised empirical force fields would perform so rather poorly. Also, in Ref. 2, it was concluded that the effect of including nuclear quantum effects, perhaps counterintuitively, served to *raise* the barriers, presumably because quantisation acts to increase the effective size of the hydrogen molecules, rendering it more difficult for them to thread the bridging hexamers.

In our "follow-up" study, we focused on the dependency of barrier heights vis-à-vis changes in temperature and found that, again contrary to expectations, the barrier height appears to *decrease* with respect to increasing temperature.[3] Although it is somewhat difficult to explain exactly why it should be borne in mind that raising the temperature acts to increase the kinetic energy at both the minima and the transition state, it is not the case that raising the temperature *must* reduce the barrier.

In any event, as stated in part already for Ref. 2, in both Refs. 2 and 3, we used the Alavi/TIP4P (2005) empirical model, this being a combination of the H_2 intermolecular force-field of Alavi *et al.*[5] together with the four-site TIP4P-2005 model of Abascal and Vega.[6] The intermolecular hydrogen molecule-water interactions are given by standard electrostatics and a simple combining rule for the 12-6 Lennard-Jones parameters.[5] This is a relatively simple model, being a combination of point-charge electrostatics and intermolecular Lennard-Jones interactions between each pair of molecules, so it is unclear whether it should be expected to give very accurate predictions for free-energy barrier heights or any other properties. In addition, we were, and continue to be, interested in nuclear quantum effects; given this, it seems that this model should be considered to be an "effective" model, already parameterised to include (albeit in a very rough way) such quantal effects. Strictly, this would make it less suitable for use with a path-integral method; although, in practice, as long as the clathrate hydrate lattice is

reasonably harmonic, it is unlikely to make much difference to the results.

Alavi *et al.*[7] have also used a model in which the H_2–H_2 intermolecular interactions were taken from a parameterisation by Wang,[8] who fitted the model parameters to reproduce the observed second virial coefficient in the range 160–423 K with an approach which took into account of second-order quantum corrections. Alavi *et al.* used this model to study the stability and structure of type II clathrate hydrates.[7] However, this work does seem to suffer from the drawback that it used classical simulation approaches only.

Recently, Cendagorta *et al.* published a calculation of free-energy profiles for enclathrated guest H_2 molecules using an approach not too dissimilar to ours.[9] These authors used the 4-site q-TIP4P/F model of water[10] in conjunction with the Alavi model for H_2 intermolecular interactions. Intermolecular hydrogen molecule-water interactions were, as usual, handled by use of a simple combining rule. Their study employed MD simulation using both classical- and ring-polymer PI-MD algorithms in conjunction with the blue-moon ensemble algorithm of Carter *et al.*,[11] in order to sample the reaction path for H_2 molecules to travel between large cages. Their results were similar to ours for both the barrier-height sizes and the effect of nuclear-quantisation on those heights. Impressively, Cendagorta *et al.* were also able to predict actual rates for hydrogen molecule hopping over those barriers;[9] using ring-polymer rate theory, they found that at temperatures above 25-50 K, the quantum rate is lower than the classical rate due to the aforementioned size-effects, in which nuclear quantisation increases the *de facto* size of the hydrogen molecule, making it harder for it to thread through the bridging hexamer connecting large cages. However, at temperatures smaller than 25 K, tunneling effects dominate, and the situation is reversed, with the quantal rates being the orders of magnitude higher than their classical counterparts. Now, it is worth noting that the q-TIP4P/F model used by Cendagorta *et al.* was parameterised explicitly to reproduce experimental bulk liquid properties under simulations with nuclear quantum effects included (by way of a path-integral algorithm). Thus, it is not an effective water model of the sort we referred to above and, hence, is more suited (strictly or at least at first glance) for use with path-integral simulation.

In the present study, we shall use a yet-different parameterisation approach, creating a potential model which is fitted directly from *ab initio*, electronic-structure-based MD propagated classically with high-quality Density Functional Theory (DFT). Not only will such a model be expected to be more accurate, but it will be, strictly, a non-effective model, meaning that it will be suitable for use with PI-MD, allowing for a more rigorous treatment of nuclear quantum effects. In particular, the new model is parameterised against *ab initio* (AI)-MD trajectories using a force-matching approach, in which the mean square error between the *ab initio* forces and potential-model forces along the entire trajectory is minimised using numerical methods.

For increased accuracy, here, we will also allow for more complex non-electrostatic interactions than the simple Lennard-Jones 12-6 intermolecular interactions between each molecular pair, as used in the Alavi/TIP4P-2005 model.[5,6] Not

only will we allow each attractive-repulsive pair-interaction to assume a more general functional form, but we will also allow for multiple interaction sites per molecule (as opposed to just the single interaction site per molecule used in the Alavi/TIP4P-2005 case). In so doing, we hope that this will allow us to achieve the following main aims:

(i)   Create a more accurate molecular force-field for the simulation of hydrogen guest molecules in clathrate-hydrate systems;

(ii) determine whether the existing Alavi *et al.* models are doing a satisfactory job of reproducing the properties of interest for these systems;

(iii) determine whether our *ab initio* simulation can reproduce the low barrier heights of Trinh *et al.*[4]

This article is organised as follows: (i) We outline the functional form of the model; (ii) we describe in some detail the force-matching algorithm, (iii) Results and analysis, and (iv) Conclusions.

THE POTENTIAL MODEL

The model is given by simple electrostatics plus a polynomial attractive/repulsive interaction acting between nuclei,

$$U^m = \sum_i \sum_{j>i} \left[\frac{q_i q_j}{4\pi\epsilon_0} + C_{ij}(r_{ij}) \right] + intra, \tag{1}$$

where the i,j sum is taken over all *intermolecular* site-site pairs and $C_{ij}(r)$ is a polynomial in reciprocal distances,

$$C_{ij}(r) = \frac{A_{6,i,j}}{r^6} + \frac{A_{8,i,j}}{r^8} + \frac{A_{10,i,j}}{r^{10}} + \frac{A_{12,i,j}}{r^{12}}. \tag{2}$$

There are also separate intramolecular terms for both water and hydrogen molecules.

The polynomial coefficients $A_{6,i,j}$, $A_{8,i,j}$, etc. are assumed to be different for each different pair of atomic species, i.e., there is one A_6 value for all water-molecule hydrogen-nuclei–water-molecule oxygen-nuclei interactions and another for all water-molecule hydrogen-nuclei–hydrogen-molecule hydrogen-nuclei interactions, and so on.

The intramolecular interactions were modeled using the highly accurate Partridge-Schwenke intramolecular surface[12] for all water molecules, and a Morse oscillator was used for each H$_2$ molecule, which takes the form

$$U^{morse}(r) = D\left(1 - e^{-a(r-r_0)}\right)^2, \tag{3}$$

where r is the intramolecular H$_2$ separation, r_0 is the equilibrium separation, D is the Morse-oscillator well depth, and a is a dimensionless parameter given by

$$a = \sqrt{\frac{k}{2D}}, \tag{4}$$

where k is a harmonic spring constant.

A three-site charge model is used for the hydrogen molecules, with a positive charge of $q_{H2,H}$ on the H nuclei and a negative charge of $-2\,q_{H2,H}$ on a massless M-site, positioned at the midway point, $(r_1 + r_2)/2$, where r_1 and r_2 are the nuclear coordinates. The water model uses a standard TIP4P-like four-site arrangement, with charges of $q_{W,H}$ on the H nuclei

and a charge of $-2\,q_{W,H}$ on an M-site positioned at $\mathbf{r}_m = \gamma(\mathbf{r}_1 + \mathbf{r}_2)/2$, where γ is a model parameter. The water-molecule O sites are assigned zero charge.

As something of a numerical experiment in order to gauge the effects of distributed site-site interactions on the water-hydrogen molecule interaction, we will also attempt a parameterisation of the force-matched model in which the intermolecular water–hydrogen-molecule $C_{ij}(r)$ interactions are limited to a single $C_{ij}(r)$ term acting between each water O and hydrogen molecule M-site. All of the other parameters are kept the same as the original force-matched model. We will call this re-parameterisation FMM-exp (experimental force-matched model), to underline that it is more a numerical experiment than a necessarily suggested model.

The Alavi/TIP4P-2005 model shares essentially the same functional form with the new model we develop here, except that it limits the attractive/repulsive polynomial to the "classic" r^{-6} and r^{-12} terms of a Lennard-Jones potential for all site-site interactions, and it limits those interactions to one Van der Waals interaction site per molecule. Both models have three electrostatic interaction sites. In our previous work with this model,[2,3] the Partridge-Schwenke intramolecular surface was used for the water molecules, with a simple harmonic oscillator for the hydrogen molecules, and we adopt the same approach here.

FORCE MATCHING

Model parameters are fit using a force-matching approach, in which the model is parameterised to best reproduce the DFT-based forces along configurations generated classically propagated *ab initio* MD, with the underlying detail thereof outlined below. In general, the force-matching algorithm was first developed by Ercolessi and Adams,[13] as a general method for fitting potential energy surfaces by minimising the square difference between potential-model and reference forces. There have been several variants on this basic procedure, but one approach worth noting is that of Izvekov and Voth,[14] who employed third-order cubic splines to model inter-particle short-range interactions; the crucial advantage here is that their method renders the interaction linear in the spline parameters, thus reducing conveniently parameter optimisation into a linear-algebra problem.

In the present work, we will use a similar approach, except that, instead of cubic splines, our $C_{ij}(r)$ functions are polynomials in powers of inverse distance; this affords a convenient form of representation and potential-model numerical evaluation. However, the general strategy is the same given that the $C_{ij}(r)$ functions [cf. Eq. (2)], like Izvekov-Voth cubic splines, are linear in their coefficients.

Let **P** be the set of all model parameters **P** $= \{P_1, P_2, \ldots, P_N\}$, where each P_i is one of the A polynomial coefficients outlined in the section titled "The potential model." Also, let $\epsilon(\mathbf{P})$ be the sum of the square differences between the model and DFT-based forces, where the sum is taken over all relevant particles and all steps,

$$\epsilon(\mathbf{P}) = \frac{1}{3N} \sum_{i,n} \left(\mathbf{F}^m_{i,n}(\mathbf{P}) - \mathbf{F}^{DFT}_{i,n} \right)^2, \tag{5}$$

where the i index runs over the number of steps in the trajectory, the n index is over the number of particles, and N is the total number of steps multiplied by the total number of particles such that $\epsilon(\mathbf{P})$ is the average square difference per particle per step per Cartesian direction.

It will prove useful to rewrite the above in terms of the force gradients (with respect to the parameters). We define $\mathbf{d}_{i,n,k}$ to be

$$\mathbf{d}_{i,n,k} = \frac{\partial \mathbf{F}^m_{i,n}(\mathbf{P})}{\partial P_k}, \tag{6}$$

which are independent of $\mathbf{P}$ because $\mathbf{F}$ is linear in $\mathbf{P}$ (for the particular functional form of the potential energy we are using).

The model forces can then be recast as

$$\mathbf{F}^m_{i,n}(\mathbf{P}) = \mathbf{F}^{m(0)}_{i,n} + \sum_j \mathbf{d}_{i,n,j} P_j, \tag{7}$$

where $\mathbf{F}^{m(0)}_{i,n} = \mathbf{F}^m_{i,n}(\mathbf{P} = 0)$, which again holds because $\mathbf{F}$ is linear in $\mathbf{P}$. Note, this term contains all the intramolecular and Coulombic forces.

The error term can then be written as

$$\epsilon(\mathbf{P}) = \frac{1}{3N} \sum_{i,n} \left(\sum_j \mathbf{d}_{i,n,j} P_j \right)^2. \tag{8}$$

We wish to minimise $\epsilon(\mathbf{P})$ with respect to the parameters, $\mathbf{P}$, which will result in parameters having the lowest square error between model and DFT forces. The derivatives are given by

$$\frac{\partial \epsilon(\mathbf{P})}{\partial P_k} = \frac{2}{3N} \sum_{i,n} \mathbf{d}_{i,n,k} \cdot \left(\sum_j \mathbf{d}_{i,n,j} P_j \right). \tag{9}$$

At the minimum, $\partial \epsilon(\mathbf{P})/\partial P_k = 0$, and we have

$$\mathbf{F}^{DFT}_{i,n} - \mathbf{F}^{m(0)}_{i,n} = \sum_j \mathbf{d}_{i,n,j} P_j, \tag{10}$$

which can be inverted via finding the left inverse matrix of $\mathbf{d}$ to give P_j. In practice, the resulting matrices are quite large, and so a conjugate-gradient approach was used to solve for $\mathbf{P}$.

To simplify the process, we decided on force-matching the water-water, H_2-water, and H_2-H_2 interactions separately. The water-water parameters were obtained first from force-matching AI-MD of a clathrate structure containing no guest H_2 molecules, and then H_2-water parameters were determined from fitting to AI-MD of a structure in which each cage (both large and small) has a single guest H_2 molecule. Finally, with these parameterisations in place, H_2-H_2 interactions were obtained from force-matching AI-MD of a structure in which each large cage has four guest H_2 molecules and each small cage is singly occupied.

Force-matching, in and of itself, can only hope to fit the parameter values over the range of the input data. Naturally, if extrapolated to unphysically short intermolecular separation, the polynomials "blow up," and this can create unphysical "black-holes" in the model potential-energy surface, as the potential energy diverges to either plus or minus infinity at distances smaller than those sampled. To prevent any MD simulation (whether by PI or classical propagation) based on

these fitted potential models from falling into one of these black-holes, it was found sometimes necessary to "repair" the model potential energy such that each polynomial pair inter-action becomes strongly repulsive at short range. We accomplished this by first tabulating the polynomial function $C_{ij}(r)$ over N points in the interval $r_{min} \ldots r_{max}$ in steps of 0.1 Å, where r_{min} and r_{max} are the minimum and maximum distances sampled in the course of force-matching the pair-interaction. An extra "trial" point at short-range, $r < r_{min}$, and with a strongly repulsive energy, is then added to the table, with a distance and energy chosen to render the function repulsive at that distance. The $N + 1$ data points are then refit with another 6-8-10-12 polynomial, $C'_{ij}(r)$. Usually, the refit can be made to give a very good approximation to both the initial function over the sampling range, $r_{min} \ldots r_{max}$, and to pass smoothly through the trial repulsive point even at short-ranges. Of course, the new function could still diverge to minus infinity at an even shorter range, but the trial point can be chosen so as to create a *de facto* insurmountable barrier to prevent the system from reaching any possible further remaining black hole.

In any event, the force-matching procedure fixes the polynomial coefficients for non-Coulombic intermolecular inter-actions [cf. Eq. (2)] but still leaves the charges and Morse-oscillator parameters to be determined. These were fitted systematically to minimise the value of ϵ, the square error between the AI-MD and potential-model forces.

SIMULATION AND FITTING DETAILS

Born-Oppenheimer (BO)-MD simulations were done within the framework of DFT under periodic boundary conditions (PBC), using a vdW-DF dispersion functional based on the generalised-gradient approximation (GGA). We employed DRSLL (Dion, Rydberg, Schroder, Langreth, and Lundqvist) non-local correlation correction,[15,16] which includes van der Waals interactions. The functional was recently shown to perform exceptionally well for liquid water[17–22] and has also yielded realistic results for H_2-hopping free-energy barriers in H_2 hydrates[3] and Raman-vibrational properties vis-à-vis experiment.[23] The simulations were performed with a time step of 0.2 fs in canonical (NVT) ensemble. System temperature was imposed at 258 K by a Nosé-Hoover thermostat[24,25] with fictitious mass 120 Ry fs^2. Charge density, described in double-zeta polarized (DZP) basis set, was converged to an accuracy of 10^{-4} a.u. at each MD step. The calculations were performed by the electronic-structure software package SIESTA.[26]

The AIMD trajectories were run for 5 ps at 258 K on a single unit cell in the NVT ensemble, for large-cage occupation numbers n = 1 . . . 4, single-occupation of small cages, as well as for an entirely empty lattice. The cubic unit cell contains 136 water molecules in a clathrate hydrate structure, with each unit cell holding eight large $5^{12}6^4$ cages and sixteen small 5^{12} cages. In addition, in order to fit for H_2-water interactions specifically, we ran an BO-MD simulation with singly filled large and small cages, in which the water lattice was fixed, and H_2 were moving, set at 600 K; this affords more intimate contact of H_2 and water and a closer approach toward the cage edges and faces; the fitting results were consistent with the

respective singly occupied 258 K simulation with a vibrating lattice, but the higher-temperature simulation allows for greater confidence and scope to construct more robust extrapolations of the potential-model fit to more intimate H_2-water contact more redolent of (toward) hopping, hexamer-dilation events, where more simplistic expressions and functional forms of interaction energies and forces between H_2 and the hexamer face tend to become more questionable.

Each structure was first run in the NPT ensemble at 258 K and 2 kbar beforehand to equilibrate to experimentally relevant conditions,[23] producing equilibrium cell lengths of 16.5–16.6 Å. This is some 3% shorter than the cell length we used previously with the Alavi/TIP4P-2005 empirical model of 17.1 Å.[2,3] In the present study, all calculations, including those with empirical/fitted models, were performed with a 16.6 Å cell length, in a manner consistent with the DFT results. Properties of interest include gauging the barrier height; indeed, in principle, this is expected to have a dependence on both system density and pressure, as well as temperature—however, assessing these dependencies lies outside the scope of the present work.

The hydrogen-molecule–water-molecule intermolecular interactions were fit to the n = 1 BO-MD trajectories with a mean square error $\epsilon(\mathbf{P}) = 1.1 \times 10^{-3}$ (Ry/Å)^2 for each H nucleus [and $\epsilon(\mathbf{P}) = 1.1 \times 10^{-3}$ (Ry/Å)^2 for the FFM-exp model]. The hydrogen molecule–hydrogen molecule intermolecular interactions were fit to the n = 4 BO-MD trajectories, with a mean square error $\epsilon(\mathbf{P}) = 2.1 \times 10^{-3}$ (Ry/Å)^2 for each H nucleus. Only hydrogen molecules residing in the large cages were considered for fitting.

Classical and PI-MD simulations with the empirical (Alavi/TIP4P-2005) and force-fitted models were each run for 1.7 ns in the NVT ensemble at 130 and 260 K with a Nosé-Hoover thermostat[24,25] under periodic boundary conditions, using the Ewald-summation method for long-range Coulombic interactions;[27] the time step was 1 fs, which was found to be satisfactory from Hamiltonian-conservation metrics. Nuclear quantum effects were incorporated by four- and eight-replica PI-MD for the 260 and 130 K simulations, respectively. This may seem like quite a small number of replicas by the standards of PI simulations for aqueous systems in the literature, but, as we have previously shown,[2] this number of beads is sufficient to converge the free-energy profiles, which are our properties of interest.

Hopping of hydrogen molecules between large cages requires them traversing quite a significant barrier, which is often unlikely to be sampled on the nanosecond time scale in an unbiased molecular-dynamics simulation. Bearing this in mind, the umbrella-sampling (US) algorithm[28] is used here in order to sample the relatively high free-energy barriers involved in the guest-molecule hopping. We apply US in conjunction with standard MD algorithms (with both classical and PI propagation) to restrain the system to the vicinity of a point along the reaction coordinate (the distance between large-cage centres along a hopping path).[2,3] Multiple runs are then performed at different points along the reaction coordinate, before the results are combined via the weighted histogram analysis method (WHAM)[29] to produce statistics over the entire region spanned.[2,3]

In the present work, as in our previous studies,[2,3] the restraint is effected via a harmonic spring term attached to the centre of mass (COM) of a single "tagged" H_2 molecule, with the other end positioned at a given fraction of the vector joining the centres of two adjoining large cages. In this way, US serves to obtain the free energy for a tagged hydrogen molecule to move from one large cage to an adjoining one, over a barrier where the guest molecule has to thread a bridging hexamer shared by the two $5^{12}6^4$ polyhedra.

RESULTS AND ANALYSIS

The final force-matched parameters are given in Table I, with the other parameters in Table II. We begin by noting that

TABLE I. A_n parameters, in units kJ/mol/Å^n. Note that "M" refers to "dummy" sites on both H_2 and TIP4P-style water[5,6] (with both species flexible).

	A_6	A_8	A_{10}	A_{12}
	Force-matched model			
H_2 M–H_2 M	$-1.216\,06 \times 10^3$	$-1.169\,6 \times 10^5$	$3.821\,72 \times 10^5$	$1.506\,13 \times 10^5$
H_2 M–H_2 H	$1.725\,58 \times 10^4$	$-4.435\,75 \times 10^4$	$-1.196\,56 \times 10^5$	$3.995\,05 \times 10^5$
H_2 H–H_2 H	$-3.290\,43 \times 10^4$	$2.715\,27 \times 10^5$	$-8.429\,8 \times 10^5$	$9.299\,67 \times 10^5$
H_2O O–H_2 M	$1.126\,15 \times 10^5$	$-1.151\,12 \times 10^6$	$4.475\,57 \times 10^6$	$-6.070\,71 \times 10^6$
H_2O O–H_2 H	$-6.163\,06 \times 10^4$	$6.464\,04 \times 10^5$	$-2.540\,23 \times 10^6$	$3.609\,5 \times 10^6$
H_2O H–H_2 H	$-1.437\,84 \times 10^4$	$1.887\,67 \times 10^5$	$-8.442\,66 \times 10^5$	$1.274\,95 \times 10^6$
H_2O O–H_2O O	$2.581\,25 \times 10^4$	$-5.354\,58 \times 10^5$	$3.897\,22 \times 10^6$	$-7.536\,87 \times 10^6$
	FMM-exp			
H_2O O–H_2 M	$-9.043\,13 \times 10^4$	$1.361\,07 \times 10^6$	$-6.900\,7 \times 10^6$	$1.183\,95 \times 10^7$
	Alavi/TIP4P-2005 model			
H_2 M–H_2 M	$-8.968\,84 \times 10^2$	0	0	$7.051\,2 \times 10^5$
H_2O O–H_2 M	$-1.661\,99 \times 10^3$	0	0	$7.051\,2 \times 10^5$
H_2O O–H_2O O	$-3.079\,79 \times 10^3$	0	0	$3.060\,1 \times 10^6$

TABLE II. Other model parameters (partial charges and intramolecular terms).

	Force-matched	Alavi/TIP4P-2005
H_2O q_H (lel)	0.57	0.5564
H_2 q_H (lel)	0.55	0.4932
H_2O γ (dimensionless)	0.2639	0.2639
H_2 D (kJ/mol)	261.6	Lim D → ∞[a]
H_2 r_0 (Å)	0.7662	0.7417
k (kJ/mol Å^2)	3405.5	3475.0

[a]Note: The Alavi/TIP4P-2005 model uses a harmonic oscillator for the intramolecular H_2 molecules, which is equivalent to a Morse oscillator with an infinite well-depth.

the force-matched model parameters exhibit reasonably similar charges to those used in the original Alavi/TIP4P model. It is interesting that the Alavi model's H_2 charges, which were found by parameterising to the gas-phase quadrupole moment,[5] are not too dissimilar to the force-matched model's effective charges produced by fitting against the DRSLL data for the bulk; indeed, this is not unexpected given that the polarisability of the two-electron H_2 molecules are relatively low.

Figures 2 and 3 show the free-energy profile curves for large-cage occupations n = 1...4 for a guest-H_2 molecule to move between two large cages in the clathrate structure. The *x* coordinate is the perpendicular distance of the guest molecule from the centre of the bridging hexamer such that *x* = 0 is the transition state corresponding to the guest lying at the centre of the bridging hexamer. Results are shown for both the Alavi/TIP4P-2005 and force-matched models; results are also shown for both classical and path-integral simulation, in which the nuclear quantum effects are accounted for.

Examining results from the DFT simulations *per se*, Figs. 2 and 3 show that the BO-MD simulations do not sample the transition state for inter-cage hopping, even in the case of the 600 K simulation with fixed lattice (results not shown). This is unfortunate but entirely expected—the barriers are just too high to be traversed on these 5 ps time scales affordable in DFT. Indeed, even molecular-dynamics simulations with empirical

potentials struggle to overcome these barriers over typical amenable simulation (nanosecond-to-microsecond) time scales—hence the need for enhanced sampling.

We will now compare US-biased simulations using the force-matched model. It can be seen that this model gives a quite good reproduction of the DFT free-energy curves across all large-cavity occupations n = 1...4. To an extent, this is to be expected *ipso facto*, as the model was fit to a subsection of the data (i.e., n = 1 for H_2-water interactions and n = 4 for H_2-H_2 BO-MD simulations and, of course, water-water interactions prior to this from empty-lattice BO-MD). However, the force-matched model was fit to best-match the instantaneous forces, and it is still significant that it is then able to reproduce the free-energy curves. It is also important to note that the force-matched model appears to give good agreement over all occupations, showing that the model is reasonably transferable.

Before discussing the path-integral-sampled curves, we first turn to Fig. 3 to gauge how the Alavi/TIP4P-2005 model compares. It turns out that this model also does a rather creditable job of reproducing the DFT-based, BOMD-sampled free-energy curves. Indeed, this may be regarded as somewhat surprising, in view of the less complex model, not fitted against any BO-MD trajectory data. Its agreement with the DFT-based data is not quite as good as that seen with the force-matched model, but it is still mostly satisfactory.

In any event, we now turn to the path-integral curves in Figs. 2 and 3. As we have shown before in Refs. 2 and 3, inclusion of nuclear quantisation acts to *increase* the free-energy barrier height for inter-cage H_2-hopping. At first glance, as remarked previously, this may seem somewhat counterintuitive, as one might expect tunneling effects to reduce the barrier, but the H_2 molecules are not in a simple double-well potential, and, in any case, tunneling effects are probably quite small in this system (at these temperatures).[7] In our previous work,[2,3] we rationalised the barrier increase as follows. First, one must note that in order to pass through the barrier, the guest molecules have to squeeze through the bridging water-hexamer, and consequently, the effective size of the guest
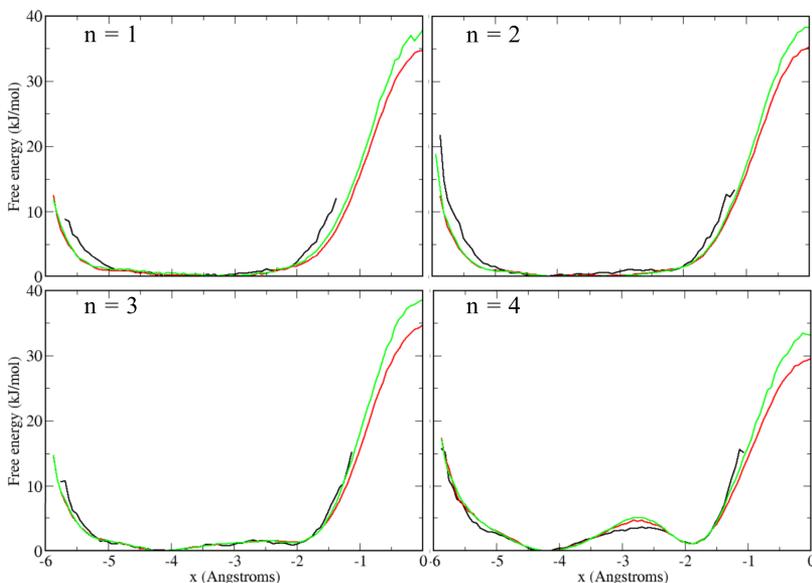
FIG. 2. DFT (black line) and force-matched (red line) free-energy scans, for large-cage occupancies 1...4 at 260 K. The green curve shows results for the force-matched model with nuclear quantum effects included by way of a four-bead path-integral simulation.
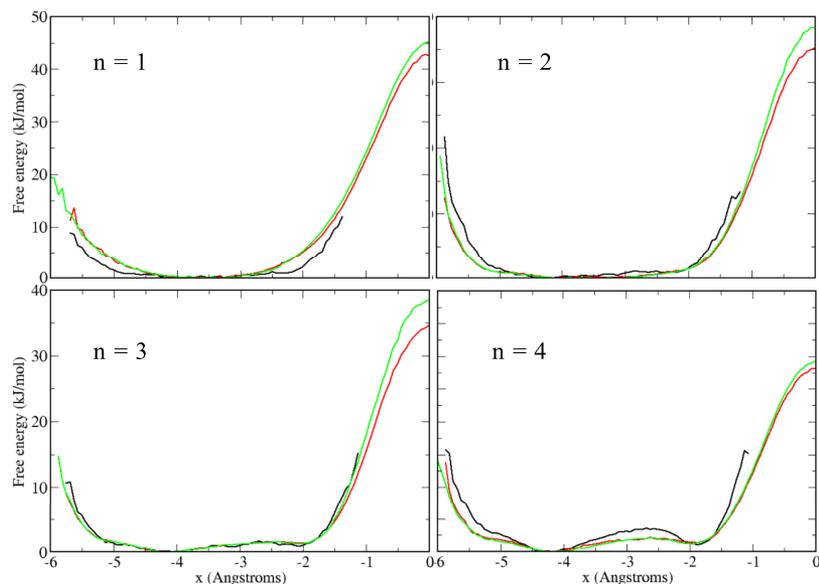
FIG. 3. Same as Fig. 2 but for the Alavi/TIP4P-2005 model.

matters a great deal to the barrier height, with larger molecules finding it more difficult to thread the hexamer. It is certainly plausible that nuclear quantisation is acting to increase the guest H_2 molecule's effective size, and it is this which is raising the barrier; in fact, in Ref. 3, we showed unambiguously that nuclear quantum effects do indeed serve to increase the size of the H_2 molecules—not along the bond direction *per se* so much as in the directions transverse thereto, with the nuclei delocalising over a large solid angle due to quantisation of each molecule's two rotational degrees of freedom.

The calculated inter-cage barrier heights for the empirical models are plotted in Fig. 4. Both classical and path-integral results are shown, and results are shown for both 130 and 260 K. Examining the classical-propagation results first, it can be seen that there is a modest difference between the two models, with the force-matched model tending to predict lower barrier heights than the Alavi/TIP4P-2005 case; although, at 260 K, the Alavi/TIP4P-2005 model has fractionally lower barrier heights for large-cage occupations of 3 and 4 guest molecules. The two models also give somewhat different predictions for the change in the barrier heights over different occupations, with the force-matched model showing fairly constant barrier heights in the range n = 1...3, with the barrier only reducing for large-cage

quadruple occupation, whereas the Alavi/TIP4P-2005 model evinces a significant downward trend in the barrier height over the entire n = 1...4 occupation range. Despite all of these noted differences, it is once again striking just how well the relatively simple Alavi/TIP4P-2005 model performs. If one assumes that the force-matched model has fairly realistic barriers, owing in large part to the high quality of vdW-DF treatment, then the Alavi/TIP4P-2005 model gives reasonably good agreement for most large-cavity occupancies. The main exception seems to be for the n = 1 occupation case, in which the Alavi/TIP4P-2005 model predicts a substantially higher barrier (at both temperatures) than found with the force-matched potential; in any event, real-world large-cage occupancies are, on average, higher than 1,[2,3] meaning that the Alavi/TIP4P-2005 model still performs rather well for systems of more practical interest.

Turning to the path-integral-sampled results, we see a very noticeable difference in the effects of nuclear quantisation between the two models, namely, that the force-matched model exhibits a much larger increase in the barrier height at each occupation when nuclear-quantum effects are included, compared to that yielded by the Alavi/TIP4P-2005 model. It would be interesting to determine the main reason(s) for the sizeable difference between the two models in the effects
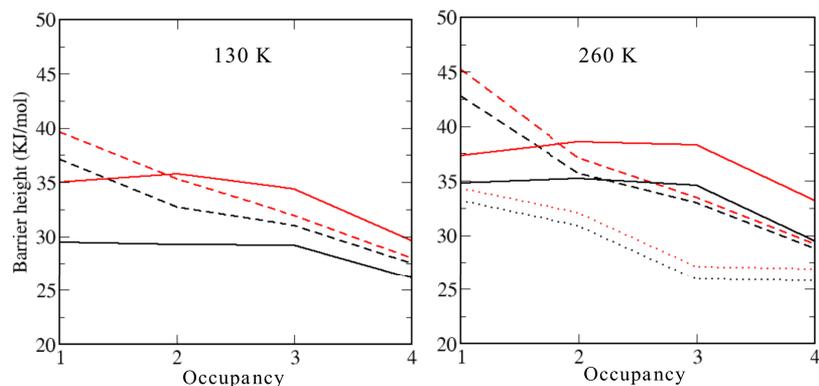
FIG. 4. Calculated inter-cage hopping free-energy barrier heights for large-cage occupancies 1...4 at 130 and 260 K for force-matched model (solid line), Alavi/TIP4P-2005 model (dashed line), and FMM-exp model (dotted line). Classical results are in black and path-integral results are in red.

of quantisation. The first point to note is that this difference is apparent even in the large-cage single-occupation occupation case (n = 1), so it is unlikely to be related to H_2–H_2 intermolecular interactions. Neither is it likely to be due to the fact that the force-matched potential uses a Morse-oscillator intramolecular H_2 potential, as opposed to the harmonic oscillators we used in our implementation of a flexible version of the Alavi/TIP4P (2005) model; the difference between the two models in intramolecular terms is relatively minor, with both having similar force-constants and equilibrium distances, and it would be surprising if such anharmonicity makes that much of a difference. Neither are the electrostatics the likely source of the difference—with both models using similar partial charges (cf. Table II).

We are left to conclude that the difference almost certainly lies in the details of the H_2-water $C_{ij}(r)$ pairwise intermolecular interactions and how they differ between the two models. This could be due to either the different parameterisation of the force-matched model, or it could be due to the force-matched model employing more site-site interactions, vis-à-vis the Alavi/TIP4P-2005 model's use of a single Lennard-Jones intermolecular interaction for each molecular pair. To test this, we turn to the FMM-exp model, in which the non-Coulombic site-site interactions are akin to the Alavi/TIP4P-2005 model, i.e., we limit ourselves to using a single non-Coulombic $C_{ij}(r)$ intermolecular interaction between each hydrogen-water molecular pair, acting between the O nucleus on each water and the M-site on each H_2-molecule. The resulting barrier heights at 260 K are shown in Fig. 4. Not only does the FMM-exp model predict significantly lower barrier heights than evidenced in the original force-matched model, but it also predicts a much smaller increase in the barrier when nuclear quantum effects are included, similar to the differences predicted by the Alavi/TIP4P-2005 model. This suggests that it is the use of distributed site–site intermolecular interactions in the force-matched model which is responsible for the large difference in the barrier height changes when nuclear quantum effects are included.

The underlying physical mechanisms at play in highlighting this interesting disparity are interesting, in terms of why the inclusion of distributed intermolecular interactions makes such a difference. Although it is difficult to be definitive, we suspect that the way distributed site-site interactions can better define the shape of the molecules is important to determine the barrier heights more rigorously for this system. Indeed, as noted previously, not only is the functional form for the exquisitely sensitive H_2-water interactions probably of great importance at, and near, the hexamer face (transition state) during hopping, but also the very assumption of the validity of pairwise interactions in an empirical model (whether fitted from high-quality DFT/BO-MD or from attempts to match relevant macroscopic observable properties) needs to be questioned. Obviously, DFT, especially with high-quality vdW-DF is inherently non-local in its determination of atomistic and intermolecular interactions, so it stands to reason that the greater the number of interaction sites when using a pairwise (or possibly even three-body) empirical potential will serve to "smear out" the underlying approximation of pairwise or
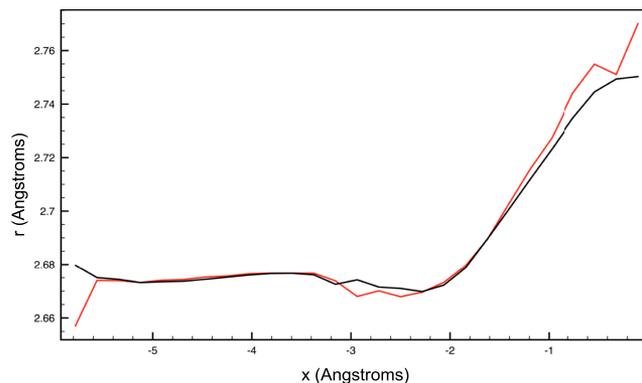
FIG. 5. Average hexamer radius as a function of the x coordinate for the cases of classical (black) and path-integral (red) sampling for n = 4 occupation, at 130 K.

few-body interactions at these "extreme" transition states, where high-quality DFT is arguably of better quality.

In Fig. 5, we show the calculated radius of the bridging hexamer versus the *x*-coordinate, where the radius is taken to be the instantaneous average distance from the geometric centre of the hexamer to its six oxygen atoms. Results are shown for the n = 4 occupation at 130 K, for both classical and path-integral cases. Unsurprisingly, there is a modest increase (from 2.68 to 2.75 Å) in the average hexamer radius when the tagged molecule is at the centre of a cage pair, as opposed to when it is inside a cage, i.e., the bridging hexamer dilates slightly when a guest hydrogen molecule passes from one cage to another. Although the statistics for the path-integral case are not quite as good as for the classical case, the path-integral results appear to show that the bridging hexamer is forced to dilate slightly more than in the classical case, consistent with our explanation that nuclear quantisation acts to increase the effective size of the guest H_2 molecules, rendering it more difficult for them to pass between cages.

It is also worth comparing briefly the calculated energies for the force-matched and Alavi/TIP4P (2005) models.

TABLE III. Relative energies (kJ/mol) of the various large-cage occupancies predicted by both the Alavi/TIP4P (2005) and force-matched models at 130 K, where mSnL indicates a clathrate hydrate structure with m hydrogen molecules per small cage and n hydrogen molecules per large cage. Each energy is reported relative to the gas phase at the same temperature.

	Alavi/TIP4P (2005)	Force-matched
	Classical	
1S0L	−63.8	−69.5
1S1L	−64.1	−78.7
1S2L	−64.4	−89.0
1S3L	−64.5	−100.2
1S4L	−64.5	−112.5
	Path integral	
1S0L	−53.1	−58.7
1S1L	−53.5	−67.9
1S2L	−53.7	−78.0
1S3L	−53.6	−89.0
1S4L	−53.4	−101.1

Relative energies (with respect to the gas phase) for different large-cage occupations are given in Table III. It can be seen that both models yield similar energies for the case of zero large-cage occupation but that the force-matched model predicts a quite large enhancement of energetic stability upon filling the large cages with hydrogen molecules, whereas the Alavi/TIP4P (2005) case shows little change in these relative energies with increasing occupation. It is somewhat difficult to assess which of the two models provides the more realistic predictions for the relative energetic stabilities for different occupations, but it is significant that they give such different results; indeed, future simulation studies will be needed to gauge which is in better accord with high-quality *ab initio* data.

CONCLUSIONS

Our implementation of a force-matching algorithm seems to be reasonably successful in reproducing the results of *ab initio* MD simulation for clathrate-hydrates with hydrogen-molecule guests. In particular, the resulting potential gives good (but not perfect) agreement with BOMD-sampled free-energy profiles for the hydrogen-molecule guests for a range of occupations, n = 1...4 hydrogen molecules per large cage. *Ab initio* data were also compared to results using a previous empirical model, Alavi/TIP4P-2005, a model which has a somewhat simpler form and which was created using a somewhat more *ad hoc* parameterisation methodology.[5] Rather surprisingly, the Alavi/TIP4P-2005 model gives very creditable results. Although the Alavi/TIP4P-2005 model is not quite as good as the force-matched model in reproducing the results from *ab initio* data, it is possibly all the more impressive that the model does so well, given that it was not parameterised to reproduce these BO-MD data. Here, we point out again that we have used a flexible version of the Alavi model; we have not tested as to whether the rigid version is similarly accurate.

We showed that the two models give surprisingly different predictions for the free-energy scans, when nuclear-quantisation effects are included, with the force-matched model predicting a much larger (~3×) increase in the barrier height, than observed using the Alavi/TIP4P-2005 model, when nuclear quantum effects are included. We were able to show that this appears due to the fact that the force-matched model uses a distributed attractive-repulsive interaction acting between various water-hydrogen molecule atomic site pairs, as opposed to the Alavi/TIP4P-2005 model which uses a single Lennard-Jones interaction acting between the hydrogen-molecule COMs (M-sites) and the water molecule. Furthermore, it was hypothesised that the use of distributed pair-interactions helps the force-matched model better capture the effective shape of the molecules.

No evidence was found for the very low-energy hopping barriers found by Trinh *et al.*[4] Using the force-matched model, as with the Alavi/TIP4P-2005 models used in the present study and our previous publications,[2,3] we find that decreasing the temperature acts to *raise* the barrier heights, not to lower them as reported by Trinh *et al.*[4]

It would seem that the above effects, i.e., (i) the presence of high-energy barriers and (ii) the increase in the barrier heights

with temperature, were not artefacts due to inaccurate empirical models and are present even in a force-matched model which is directly fitted to *ab initio* data. In terms of future outlook, high-quality DFT with non-local correlation is arguably of more promising fidelity for treating H_2/guest hopping at cage-face-crossing "extreme" transition states, but this is compounded by unbiased AI-MD's lack of ability to sample these very events. Still, the use of higher-temperature AI-MD with a restrained lattice (to prevent thermal decomposition, at least over picosecond time scales) does allow for a closer approach of H_2 to the hexamer faces and serves to offer a glimpse of the "breakdown" of pairwise energy/force determinations with an insufficient number of interaction sites to capture the sensitivities and subtleties of the molecular shape and non-local interactions with H_2 (or sufficiently small guest, in general) at such "close quarters" and in intimate contact with the hexamer. This motivates the quest for a larger number of interaction sites, as we have indeed found and argued in the present study. Further, the inclusion of biased-sampling methods, such as US (and WHAM) or Blue-Moon approaches, to classically propagated and PI-propagated AI-MD, using high-quality non-local correlation in the underlying DFT (or even PBC-based local-MP2 methods)[30] would certainly offer further progress in accurate and rigorous treatment of cage-hopping free-energy calculations.

SUPPLEMENTARY MATERIAL

See supplementary material in which FORTRAN code for implementing the force-matched potential is provided.

ACKNOWLEDGMENTS

[1] Y. F. Makogon, *Hydrates of Hydrocarbons* (Penn Well Books, Tulsa, OK, 1997).

[2] C. J. Burnham and N. J. English, J. Phys. Chem. C **120**, 16561 (2016).

[3] C. J. Burnham, Z. Futera, and N. J. English, Phys. Chem. Chem. Phys. **19**, 717 (2017).

[4] T. T. Trinh, M. H. Waage, T. S. van Erp, and S. Kjelstrup, Phys. Chem. Chem. Phys. **17**, 13808 (2015).

[5] S. Alavi, J. A. Ripmeester, and D. D. Klug, J. Chem. Phys. **123**, 024507 (2005).

[6] J. L. F. Abascal and C. Vega, J. Chem. Phys. **123**, 234505 (2005).

[7] S. Alavi, D. D. Klug, and J. A. Ripmeester, J. Chem. Phys. **128**, 064506 (2008).

[8] W. F. Wang, J. Quant. Spectros. Rad. Transfer **76**, 23 (2003).

[9] J. R. Cendagorta, A. Powers, T. J. H. Hele, O. Marsalek, Z. Bačić, and M. E. Tuckerman, Phys. Chem. Chem. Phys. **18**, 32169 (2016).

[10] S. Habershon, T. E. Markland, and D. E. Manolopoulos, J. Chem. Phys. **131**, 024501 (2009).

[11] E. A. Carter, G. Ciccotti, J. Hynes, and R. Kapral, Chem. Phys. Lett. **156**, 472 (1989).

[12] H. Partridge and D. W. Schwenke, J. Chem. Phys. **106**, 4618 (1997).

[13] F. Ercolessi and J. B. Adams, Europhys. Lett. **26**, 583 (1994).

[14] S. Izvekov and G. A. Voth, J. Chem. Phys. **123**, 134105 (2005).

[15] M. Dion, H. Rydberg, E. Schröder, D. C. Langreth, and B. I. Lundqvist, Phys. Rev. Lett. **92**, 246401 (2004).

[16] G. Román-Pérez and J. M. Soler, Phys. Rev. Lett. **103**, 096102 (2009).

[17] J. Wang, G. Román-Pérez, J. M. Soler, E. Artacho, and M.-V. Fernández-Serra, J. Chem. Phys. **134**, 024516 (2011).

[18] C. Zhang, J. Wu, G. Galli, and F. Gygi, J. Chem. Theory Comput. **7**, 3054 (2011).

[19] F. Corsetti, E. Artacho, J. M. Soler, S. S. Alexandre, and M.-V. Fernández-Serra, J. Chem. Phys. **139**, 194502 (2013).

[20] A. Bankura, A. Karmakar, V. Carnevale, A. Chandra, and M. L. Klein, J. Phys. Chem. C **118**, 29401 (2014).

[21] N. J. English, Energies **8**, 9383 (2015).

[22] M. J. Gillan, D. Alfè, and A. Michaelides, J. Chem. Phys. **144**, 130901 (2016).

[23] Z. Futera, M. Celli, L. del Rosso, C. J. Burnham, L. Ulivi, and N. J. English, J. Phys. Chem. C **121**, 3690 (2017).

[24] S. Nosé, J. Chem. Phys. **81**, 511 (1984).

[25] W. G. Hoover, Phys. Rev. A **31**, 1695 (1984).

[26] J. M. Soler, E. Artacho, J. D. Gale, A. García, J. Junquera, P. Ordejón, and D. Sánchez-Portal, J. Phys.: Condens. Matter **14**, 2745 (2002).

[27] W. Smith, *Elements of Molecular Dynamics* (2015).

[28] G. M. Torrie and J. P. Valleau, J. Comput. Phys. **23**, 187 (1977).

[29] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, J. Comput. Chem. **13**, 1011 (1992).

[30] N. J. English and J. M. D. MacElroy, Chem. Eng. Sci. **121**, 133 (2015).