AIP Publishing

This manuscript was accepted by J. Chem. Phys. Click here to see the version of record.

Explicit treatment of hydrogen bonds in the Universal Force Field: validation and application for Metal-Organic Frameworks, hydrates and host-guest complexes

Damien E. Coupry
Wilhelm-Ostwald-Institut für Physikalische und Theoretische Chemie, Fakultät für Chemie und Mineralogie, Universität Leipzig, Linnéstr. 2, 04103 Leipzig, Germany

Matthew A. Addicoat*
School of Science and Technology, Nottingham Trent University, Clifton Lane, NG11 8NS Nottingham, UK

Thomas Heine
Wilhelm-Ostwald-Institut für Physikalische und Theoretische Chemie, Fakultät für Chemie und Mineralogie, Universität Leipzig, Linnéstr. 2, 04103 Leipzig, Germany

A straightforward means to include explicit hydrogen bonds within the Universal Force Field is presented. Instead of treating hydrogen bonds as non-bonded interaction subjected to electrostatic and Lennard-Jones potentials, we introduce an explicit bond with negligible bond order, thus maintaining the structural integrity of the H-bonded complexes and avoiding the necessity to assign arbitrary charges to the system. The explicit hydrogen bond changes the coordination number of the acceptor site and the approach is thus most suitable for systems with under-coordinated atoms, such as many metal-organic frameworks, however, it also shows excellent performance for other systems involving a hydrogen-bonded framework. In particular, it is an excellent means for creating starting structures for molecular dynamics and for investigations employing more sophisticated methods.

The approach is validated for the hydrogen bonded complexes in the S22 dataset and then employed for a set of metal-organic frameworks from the Computation-Ready Experimental (CoRE) database and several hydrogen bonded crystals including water ice and clathrates. We show that direct inclusion of hydrogen bonds reduces the maximum error in predicted cell parameters from 66% to only 14% and the mean unsigned error is similarly reduced from 14% to only 4%. We posit that with the inclusion of hydrogen bonding, the solvent-mediated breathing of frameworks such as MIL-53 is now accessible to rapid UFF calculations, which will further the aim of rapid computational scanning of metal-organic frameworks while providing better starting points for electronic structure calculations.

ACCEPTED MANUSCRIPT

* matthew.addicoat@ntu.ac.uk

This manuscript was accepted by J. Chem. Phys. Click here to see the version of record.

I. INTRODUCTION

Metal Organic Frameworks (MOFs) are now a well-known class of crystalline, porous materials, where inorganic connectors are joined by organic linkers, forming a 3-dimensional network. Since the publication of MOF-5 in 1999[1], research into MOFs has grown almost exponentially and there are now several thousand synthesised MOFs.[2] Several reviews have been written on various aspects of MOF chemistry,[3, 4] including their interactions with water.[5, 6]

Water is an important solvent for MOFs, it serves as a synthetic medium,[7–9] a structural component,[10] and an adsorbate.[11] Water plays a major role in mediating defects and dissolution of many MOFs.[12–14] Some MOFs, such as the UiO series, are highly water stable[15], while others break down or transform[16–19] when exposed to water. Proposed applications of MOFs include separation of organic contaminants[20] and arsenic[21] from aqueous mixtures and capture of water from air[11].

The requirement to computationally scan or screen the nearly infinite number of possible MOFs led to the development of UFF4MOF.[22, 23] UFF4MOF extends the Universal Force Field of Rappé[24] to include several atom types present in MOFs but not accounted for in the original list of atom types. UFF4MOF thus permits the rapid calculation of structures for a wide range of actual and hypothetical MOFs. The structures predicted using UFF4MOF are typically within 5% of experimental cell parameters, and it is therefore of interest whether similar accuracy can be maintained when the structural parameters are significantly determined by hydrogen bonding within the framework. MIL-53[25] represents a notable case of water not only stabilizing, but determining the framework structure. The transition from (evacuated) large pore to narrow pore to hydrated and superhydrated large pore has been extensively studied by both experimental and computational methods.[26–29] The empty, large pore (LP) structures of MIL-53(M) (M = Cr, Al, Fe) have all been replicated using UFF4MOF, and in this work we seek to replicate the hydrated narrow pore (NP) structures, whose cell parameters are dictated by the hydrogen bonds that occur between the guest water molecules and the framework.[30, 31]

Treatment of hydrogen bonds is challenging for classical force fields. On one hand, the bonds are readily formed and broken in solution, so that an explicit, permanent bond would be an inappropriate description if investigating dynamical processes or, for example, the liquid state. Many force fields treat hydrogen bonds as non-bonded interactions: the bonding is achieved by attractive components arising from the London dispersion and the Coulomb interaction. In UFF, the former is defined by a Lennard-Jones potential, while the latter relies on the definition of charges. As the Coulomb interaction dominates, charge definition is crucial for the results. However, this approach works well for water and is particularly useful in QM/MM calculations, as was previously shown in DFTB/UFF calculations with electrostatic embedding.[32, 33].

Several force fields attempt to explicitly treat hydrogen bonds: MM2[34], MM3[35] and AMBER force fields all employ (optionally in the case of AMBER) an angle-independent 10,12 Lennard-Jones potential to describe non-bonded interactions. MM3 was later updated to include directional hydrogen bonding[36]. Paton and Goodman[37] reviewed these force fields along with OPLS*, OPLSAA and MMFF, which do not include explicit hydrogen bonding and concluded that for the S22 database, the OPLSAA and MMFF force fields performed best, yielding accurate geometries and interaction energies. The same two force fields also ranked highest against the JSCH2005 database, but some optimized geometries differed from the benchmark geometry.

Despite these advances, there generally remains a technical problem. For host-guest systems that are treated entirely using a force field, it is often not practical to define charges to treat the electrostatic interactions. Indeed, although a charge equilibration scheme was designed for use with UFF [38], the original parameterization was done without an electrostatic model: the implementation is thus open to interpretation. Some codes forego explicit charges completely [39, 40]. UFF also explicitly includes electrostatic interactions for 1st and 2nd neighbours, so that most electrostatics is screened for a bonded system. This, of course, does not hold for non-bonded interactions, where significant electrostatic contributions arise due to the proximity of atoms formally carrying a charge. There is a further, merely technical aspect for the choice of explicit hydrogen bond that is crucial when studying static hydrogen-bonded frameworks including MOFs, ice, clathrates, or inclusion compounds: upon generation their starting structure, the non-bonded character in conjunction with substantial forces at initial geometry optimization, or, worse, kinetic energy gain upon molecular dynamics startup, makes it hard to converge to or to maintain the anticipated local minimum representing the desired framework topology.

ACCEPTED MANUSCRIPT

This manuscript was accepted by J. Chem. Phys. Click here to see the version of record.

II. DEFINITION OF EXPLICIT HYDROGEN BONDS WITHIN THE UNIVERSAL FORCE FIELD

To provide a fast and topology-preserving structure generation method, and to avoid the definition of charges in order to account for the Coulomb interaction in the hydrogen bonds, we take advantage of the implicit definition of the electronegativity correction r_{EN}, which is contributing to the UFF natural bond length r_{ij}:

$$r_{ij} = r_i + r_j + r_{\mathrm{BO}} - r_{\mathrm{EN}} \tag{1}$$

where r_i and r_j are the bond radii of atoms i and j respectively, r_{BO} is a bond order correction to the bond distance. The bond order correction, in turn, is defined as:

$$r_{\mathrm{BO}} = -\lambda(r_i + r_j)\ln(n) \tag{2}$$

where λ is a proportionality constant derived using propane, propene and propyne with the C_3, C_2 and C_1 radii and is equal to 0.1332. There is precedent for assigning unique bond orders for special types of bonds, with the amide bond order set to 1.41 in order to reproduce the C-N bond distance in *N*-methylformamide.[24] Assigning a similar bond order for hydrogen bonds requires selecting an appropriate reference. At first glance, this is less than straightforward, given the variety of hydrogen bonds that may be encountered in framework structures. However, considering the prime importance of water to the synthesis, structure and behaviour of MOFs, a reasonable reference is the the prototypical C_s global minimum of the water dimer. We further require that our treatment of hydrogen bonds can reproduce the hydrogen bond mediated breathing that occurs in hydrated MIL-53.

Treating hydrogen bonds as explicit bonds without bond order does have negative consequences: The dynamical breaking and formation of hydrogen bonded networks, essential for the description of the liquid state, becomes impossible (though it may work if the Reactive Force Field, ReaxFF[41], was used instead of UFF). We note that defining an explicit hydrogen bond increases the coordination number of the central atom by one, however, this typically does not pose a problem for otherwise under-coordinated sites, and fully coordinated atoms typically are not hosting extra solvent. Secondly, because the defined bonding network is changed in the hydrogen-bonded system, it is not possible to calculate binding energies or relative stabilities using this approach. However, for the problem of rapid structural pre-optimization, neither of these disadvantages apply and furthermore, they are offset by the ability to rapidly produce high quality structures, without requiring the relatively expensive calculation of electrostatic terms.

As we will show below, for "frozen" configurations of hydrogen bonded dimers our approach performs significantly better than the traditional UFF treatment using nonbonded interactions with Coulomb interactions defined via HF/3-21G charges, and should provide a much better starting point for subsequent electronic structure calculations.

The water dimer has been extensively studied over decades, by both experimental[42, 43] and computational[44–46] means. Early calculations yielded a H$\cdots$O distance of 1.72 Å[47], a subsequent study including some of the same authors yielded a value of 1.80 Å[48]. By the 1990's, calculations using MP2 and Coupled Electron Pair Theory (CEPA-1)[49] yielded geometries and energies in good agreement with experiment[50, 51] and the "gold standard" CCSD(T)/QZ geometry found in the S22 database[52] also agrees very well, with $R_{H\cdots O} = 1.952\mathring{A}$ (corresponding to $R_{OO} = 2.91\mathring{A}$). If we therefore consider the UFF bond length and bond order correction in Equations 1 and 2, we note that λ, r_i and r_j and r_{EN} are all fixed within the UFF framework. Therefore, for water, where the oxygen atom has the O_3 atom type and hydrogen has the H_ atom type, $r_i + r_j = 0.354 + 0.658 = 1.012\mathring{A}$, the electronegativity correction for a H-O_3 bond is 0.0021 and thus the required bond order correction is equal to 0.9380 Å. A bond order of 0.001 yields correction of 0.9312 Å and noting the diminishing returns of further optimization of such a "bond order", combined with the large range of acceptable bond lengths for a hydrogen bond, we propose to this value as a reasonable approximation to a bond order for describing hydrogen bonds in UFF calculations on metal-organic frameworks. Employing this bond order yields a bond length of 2.025 Å for a H$\cdots$N_R bond and 1.899 Å where an O_2 atom is the proton acceptor, both of which are reasonable lengths for a hydrogen bond.

III. RESULTS AND DISCUSSION

To check the basic sensibility of using a bond order to correct hydrogen bond distances, we undertook geometry optimizations of the hydrogen bonded complexes in the S22 database. The C_{2h} ammonia dimer was excluded from analysis as the angle terms resulting from the addition of the two hydrogen bonds considerably change the geometry of the dimer. The hydrogen bond distances of the other six complexes are shown in Table I and this is the only metric we employ for these non-bonded clusters.[53] Calculations were undertaken in the General Utility Lattice Program (GULP)[54, 55] except those employing atomic charges, where deMonNano[56] was employed. Hartree-Fock

ACCEPTED MANUSCRIPT

This manuscript was accepted by J. Chem. Phys. Click here to see the version of record.

TABLE I. Hydrogen bond distances calculated using UFF for hydrogen bonded complexes in the S22 database[52]. Percent errors are calculated as $(X_{\mathrm{UFF}}$ - $X_{\mathrm{ref}})/X_{\mathrm{ref}} \times 100$, where X_{UFF} denotes UFF-predicted and X_{ref} denotes the original *ab initio* value.

Complex (symmetry)	Hydrogen bond Atom types	UFF with HF/3-21G charges	UFF with explicit H-bonds	Reference	% error
$(H_2O)_2$ (C_s)	O_3-H$\cdots$O_3	2.666	1.920	1.952	-1.7
ammonia dimer (C_{2h})	N_3-H$\cdots$N_3	2.854	-	2.504	-
Formic acid dimer (C_{2h})	O_R-H$\cdots$O_1	2.488	1.886	1.670	12.9
Formamide dimer (C_{2h})	N_R-H$\cdots$O_1	2.520	1.884	1.840	2.4
Uracil dimer (C_{2h})	N_R-H$\cdots$O_2	2.481	1.882	1.774	6.1
2-pyridoxine - 2-aminopyridine (C_1)	N_R-H$\cdots$N_R	2.609	2.030	1.860	9.1
	N_R-H$\cdots$O_1	2.585	1.858	1.874	-0.8
Adenine - thymine WC (C_1)	N_R-H$\cdots$N_R	2.556	2.030	1.819	11.6
	N_R-H$\cdots$O_1	2.477	1.857	1.929	-3.7

TABLE II. Error on bond angles calculated using UFF for hydrogen bonded complexes in the S22 database[52]. Percent errors are calculated as $100/N_{\mathrm{ang}} \times \sum_{i=1}^{N_{\mathrm{ang}}} (X_{\mathrm{UFF}} - X_{\mathrm{ref}})/X_{\mathrm{ref}}$ and rmse as $(1/N_{\mathrm{ang}} \times \sum_{i=1}^{N_{\mathrm{ang}}} (X_{\mathrm{UFF}} - X_{\mathrm{ref}})^2)^{\frac{1}{2}}$, where X_{UFF} denotes UFF-predicted and X_{ref} denotes the original *ab initio* value

	UFF, LJ only		UFF with explicit H-bonds	
Complex (symmetry)	RMSE	Mean error (%)	RMSE	Mean error (%)
$(H_2O)_2$ (C_s)	2.97341	2.6813	3.60928	2.85337
ammonia dimer (C_{2h})	6.89258	7.00757	22.4794	21.6555
Formic acid dimer (C_{2h})	6.97037	22.1015	7.08619	81.4685
Formamide dimer (C_{2h})	9.29674	39.9108	2.85878	14.4514
Uracil dimer (C_{2h})	3.60883	9.89631	3.02618	9.90769
2-pyridoxine - 2-aminopyridine (C_1)	3.57988	16.3517	2.85318	10.1884
Adenine - thymine WC (C_1)	3.7847	15.6545	4.11144	11.9441

calculations were undertaken in Gaussian09[57].

For these simple complexes, the results show reasonable agreement with the *ab initio* reference and are far superior to UFF + Lennard-Jones + electrostatics. The two N_R-H$\cdots$N_R bonds are overestimated by approximately 10%, as is the unusually short hydrogen bond in the formic acid dimer. Other bonds are within 6% of their reference values. Undertaking the calculations without specification of the hydrogen bond results in bond distances increasing by approximately 1Å which is clearly poor.

We note in Table I, the only failure of the explicit hydrogen-bond approach is the C_{2h} ammonia dimer, which upon inclusion of two explicit hydrogen bonds, optimizes to a singly hydrogen-bonded complex of C_s symmetry. In this case the two hydrogen bonds in the reference structure form H-N$\cdots$H angles of 58°, a significant deviation from the 106.7° angle of the N_3 parameter, and the optimizer prefers to allow a single hydrogen bond with a close-to-ideal angle (i.e. essentially tetrahedral geometry around the acceptor nitrogen atom), rather than two hydrogen bonds with large errors on the angle term. This is a general limitation of this approach, in that by making the hydrogen bond explicit, the coordination number of both the hydrogen atom and the accepting atom are increased by one and the hydrogen bonded atom consequently figures in the angle terms around the acceptor atom. In most cases, this effect is either desired or benign, such as in the case of hydrogen bonding to an under-coordinated metal atom in a paddlewheel.

A. Metal-Organic Frameworks

Having thus established the validity of the approach, a set of framework materials where hydrogen bonding is important for maintaining structural integrity was selected from the Computation Ready Experimental (CoRE) database.[2] The original crystal structures were re-sourced from the Cambridge Structural Database[58] in order to

ACCEPTED MANUSCRIPT

This manuscript was accepted by J. Chem. Phys. Click here to see the version of record.

recover the solvent molecules. To this test set we add the particular case of MIL-53(Al) NP[25]. After assigning atom types to each structure, we detect hydrogen bonds by looking for hydrogen atoms. For every H atom that is directly connected to (O, S, N) we assign a larger covalent radius (of 2Å), then with this, we re-build the connectivity list and we add to the original connectivity list every new bond between these super-big hydrogen atoms and (O, N, S, F, Cl, Br, I) provided that the bond angle is between 140 and 220 degrees. For all structures we calculate the structure both with and without specification of hydrogen bonds and note the resultant cell parameters, which are listed in Table III.

TABLE III: Comparison of UFF calculated and experimental cell parameters of selected MOFs. The first 10 structures possess primarily intermolecular hydrogen bonds, the second 10 possess more intramolecular hydrogen bonds. Percent errors are calculated as $(X_{\mathrm{UFF}} - X_{\mathrm{exp}})/X_{\mathrm{exp}} \times 100$, where X_{UFF} denotes UFF-predicted and X_{exp} denotes the original value.

CSD Refcode		Experimental	UFF without H-bonds	UFF with H-bonds	% error without H-bonds	% error with H-bonds
MIL-53(Al) NP[25]	a =	19.504	18.593	19.423	-4.7	-0.4
	b =	15.201	21.839	15.427	43.7	1.5
	c =	6.569	6.366	6.470	-3.1	-1.5
CDLGLU01[59]	a =	11.575	10.099	11.217	-12.8	-3.1
	b =	10.764	15.304	9.920	42.2	-7.8
	c =	7.256	7.435	7.775	2.5	7.2
CUGLTM[60]	a =	11.084	13.404	10.989	20.9	-0.9
	b =	10.350	11.262	9.412	8.8	-9.1
	c =	7.238	6.330	7.246	-12.6	0.1
FUFREE[61]	a =	7.785	8.312	7.808	6.8	0.3
	b =	10.238	12.172	9.780	18.9	-4.5
	c =	15.851	15.524	16.074	-2.1	1.4
HURNOX[62]	a =	7.149	6.891	6.158	-3.6	-13.9
	b =	10.468	14.307	9.573	36.7	-8.5
	c =	11.295	10.156	11.518	-10.1	2.0
ICOWON[63]	a =	8.754	8.240	8.232	-5.9	-6.0
	b =	10.003	10.693	9.775	6.9	-2.3
	c =	11.790	11.446	11.479	-2.9	-2.6
JUCXEK[64]	a =	11.382	13.971	12.306	22.7	8.1
	b =	11.382	14.020	12.509	23.2	9.9
	c =	11.734	9.592	10.169	-18.3	-13.3
MAZTIR[65]	a =	19.014	19.839	19.215	4.3	1.1
	b =	6.843	9.715	6.782	42.0	-0.9
	c =	9.420	9.217	9.753	-2.2	3.5
MEHPAQ[66]	a =	7.221	6.350	7.205	-12.1	-0.2
	b =	10.295	11.218	9.948	9.0	-3.4
	c =	11.041	13.425	10.966	21.6	-0.7
MUTVUT[67]	a =	9.335	9.255	9.399	-0.9	0.7
	b =	22.203	25.664	23.168	15.6	4.3
	c =	27.514	26.756	27.710	-2.8	0.7
SALLAT[68]	a =	15.686	12.126	15.023	-22.7	-4.2
	b =	8.165	13.522	8.865	65.6	8.6
	c =	13.119	12.225	12.767	-6.8	-2.7
DUQSEO[69]	a =	8.884	7.930	7.606	-10.7	-14.4
	b =	13.093	11.963	12.266	-8.6	-6.3
	c =	13.135	16.660	14.021	26.8	6.7
FAPTUN[70]	a =	14.304	11.704	14.757	-18.2	3.2
	b =	16.970	17.091	16.907	0.7	-0.4
	c =	11.098	11.342	11.373	2.2	2.5
KOJCUI[71]	a =	18.326	21.776	18.356	18.8	0.2
	b =	25.300	21.273	24.369	-15.9	-3.7
	c =	7.524	6.804	6.900	-9.6	-8.3
RATVEP[72]	a =	7.764	9.142	8.044	17.8	3.6
	b =	10.177	11.938	10.240	17.3	0.6
	c =	15.949	15.840	16.222	-0.7	1.7
SARBOE[73]	a =	14.389	14.925	14.477	3.7	0.6
	b =	15.475	14.511	15.006	-6.2	-3.0
	c =	8.424	10.610	8.788	26.0	4.3
SIVKAK[74]	a =	7.583	9.243	8.340	21.9	10.0
	b =	31.086	27.033	29.315	-13.0	-5.7
	c =	13.799	14.199	13.523	2.9	-2.0
VEFLUP[75]	a =	8.002	7.283	7.460	-9.0	-6.8

ACCEPTED MANUSCRIPT

AIP Publishing
33.5%
This manuscript was accepted by J. Chem. Phys. Click here to see the version of record.

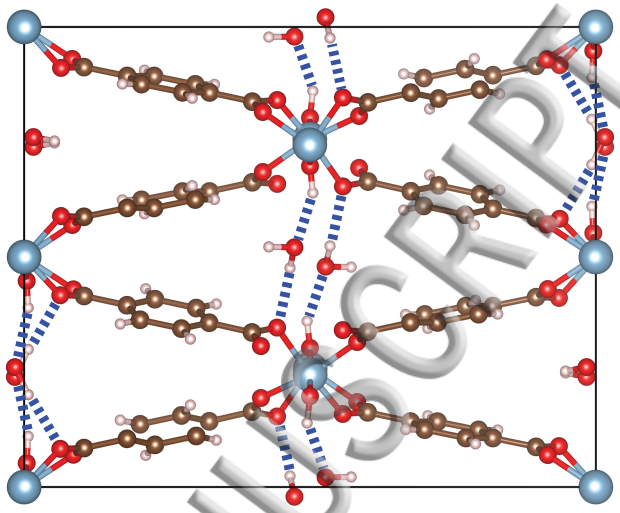

FIG. 1. MIL-53(Al) narrow pore structure (CCSD refcode: SABWAU [25]) with hydrogen bonds between included water and framework indicated by dashed blue lines.

	b =	9.091	10.877	9.031	19.6	-0.7
	c =	12.935	13.070	13.071	1.0	1.0
YORZAH[76]	a =	16.493	16.283	16.334	-1.3	-1.0
	b =	6.820	7.905	6.597	15.9	-3.3
	c =	22.220	23.331	21.906	5.0	-1.4
YUVSUE[77]	a =	15.435	17.188	15.518	11.4	0.5
	b =	15.435	17.188	15.793	11.4	2.3
	c =	22.775	18.122	21.005	-20.4	-7.8
ZNGLUD[78]	a =	11.190	13.840	12.126	23.7	8.4
	b =	10.463	11.273	10.112	7.7	-3.4
	c =	7.220	6.369	7.203	-11.8	-0.2
				Maximum Unsigned Error	65.6	14.4
				Mean Unsigned Error	13.9	4.0

In framework systems, the explicit inclusion of hydrogen bonds should yield even better agreement with experimental reference structure, as terms corresponding to regular covalent bonds far outnumber the contribution due to hydrogen bonds, and this is indeed the case. Without specifying hydrogen bonds, each structure has at least one cell parameter expand by 15% or greater, and the mean unsigned error on all cell parameters is 13.9%. Once hydrogen bonds are specified, the maximum unsigned error in cell parameters is only 14.4% and the mean unsigned error is only 4.0%.

The narrow pore structure of MIL-53(Al) is of particular note. Each pore contains two water molecules which are hydrogen-bonded to the framework oxygen atoms and the hydrogen of the framework hydroxy group. Optimising the structure with 16 hydrogen bonds specified (two hydrogen bonds per water molecule, see Figure 1), the cell parameters of MIL-53(Al) are all predicted within 1.5%, including the b dimension, which otherwise expands by over 43% to resemble the large pore structure.

ACCEPTED MANUSCRIPT

This manuscript was accepted by J. Chem. Phys. Click here to see the version of record.

B. Hydrogen-bonded crystal structures

The prediction of the structures of hydrogen-bonded MOFs is perhaps the primary use-case of this approach, however, the structure of any static hydrogen-bonded system is amenable to calculation this way. To illustrate the approach' broader utility, a variety of structures where hydrogen bonding is crucial to defining the structure were recalculated. Hydrogen bonds were inserted following the literature description of each structure. Table IV shows the results for these structures.

1. *Methanol*

The case of α-methanol[79] deserves special consideration as it illustrates how the choice of hydrogen bonds affects the optimized structure. There are two distinct types of hydrogen bond in α-methanol; a bond from the hydroxyl hydrogen to the neighbouring oxygen with a OH$\cdots$O distance of 1.75Å and a CH$\cdots$O distance of 2.51Å, both illustrated in Figure 2. Choosing only the strong OH$\cdots$O results in a correct estimate of the a parameter, but exchanges the errors on the b and c parameters. Including the CH$\cdots$O in addition results in an underestimation of both the b and c parameters and a moderate overestimation of a. While none of the three calculated structures is a clear winner, arguably including only the OH$\cdots$O results in the best quality structure. This choice is then consistent with the structure of β-methanol[80], the high temperature phase, which possesses only OH$\cdots$O hydrogen bonds. These bonds, arranged in sheets, are crucial to the accurate description of the structure, neglecting them results in a strong distortion of the crystal, with the a parameter being underestimated by 32% and the c parameter being overestimated by 48%.

2. *Water ice and hydrate structures*

The classic hydrogen-bonded structure is that of water. The approach of fixing hydrogen bonds as actual bonds, makes the description of fluxional systems, such as liquid water, impossible. However, the structure of non-fluxional ice is readily amenable to calculation, the structures of both cubic and hexagonal water ice were calculated using all hydrogen bonds, and making each oxygen atom formally tetrahedral with two single bonds and two hydrogen bonds. Employing this approach results in a nearly uniform overestimation of lattice parameters of 4.5-7.5% for both structures, a clear improvement over neglecting the hydrogen bonds which results in parameters being over/underestimated by 20-40%.

The success of the explicit approach to hydrogen bonds in water ice has further significance in describing clathrate structures. Methane hydrates, in particular are the subject of renewed research as, found on the ocean floor and in polar regions they are estimated to contain up to 12% of all the organic carbon on Earth[81], making them an important energy resource. In addition, hydrates are often formed within gas pipelines, where they are unwanted and cause significant damage[82]. The structures of three methane hydrates, MH-I, MH-II and MH-H (hexagonal)[83] were calculated, yielding results broadly similar to those of water ice, whereby specification of hydrogen bonds gives a structure with cell parameters uniformly overestimated by approximately 7% and neglecting those bonds results in very poor structures with parameters under- and overestimated by up to 40%. More modest improvement is seen for the structure of n-butanol hexahydrate (CCSD Refcode WUVZIW).

3. *Host-guest inclusion complexes and cocrystals*

To further illustrate the diversity of hydrogen-bonded systems to which this simple approach may be applied, structures described as being hydrogen-bonded were sourced from the CCSD[58]. Structures included two urea-based inclusion compounds (ABAZOS[84] and WARWOB[85]), a macrocyclic inclusion complex, ABUCIJ[86], a porous diamide matrix, ABEBUF[87], a Cu coordination compound forming a hydrogen-bonded helicate, SIYRAU[88] and three cocrystals of pyrogallol[4]arenes and the ionic liquid 1-ethyl-3-methylimidazolium ethylsulfate[89]. In these cases, neglecting hydrogen bonds typically leads to at least one cell parameter in error by greater than 10%. Overall, employing explicit hydrogen bonds reduces the maximum unsigned error for these complexes from 48.3% to 12.1% and the mean unsigned error from 12.6% to 4.2%.

ACCEPTED MANUSCRIPT

This manuscript was accepted by J. Chem. Phys. Click here to see the version of record.

TABLE IV. Comparison of UFF calculated and experimental cell parameters of selected hydrogen bonded crystals. Percent errors are calculated as $(X_{\mathrm{UFF}} - X_{\mathrm{exp}})/X_{\mathrm{exp}} \times 100$, where X_{UFF} denotes UFF-predicted and X_{exp} denotes the original value.

CSD Refcode		Experimental	UFF without H-bonds	UFF with H-bonds	% error without H-bonds	% error with H-bonds
Ic ice[90]	*a*	6.358	8.805	6.839	38.5	7.6
	b	6.358	9.019	6.641	41.8	4.5
	c	6.358	9.019	6.641	41.8	4.5
Ih ice[90]	*a*	4.506	5.483	4.756	21.7	5.5
	b	4.506	6.284	4.717	39.5	4.7
	c	7.346	5.205	7.776	-29.2	5.9
alpha methanol[79]	*a*	4.873	4.645	4.955	-4.7	1.7
Figure 2(c)	*b*	4.641	5.000	4.487	7.7	-3.3
	c	8.867	9.835	7.798	10.9	-12.1
alpha methanol[79]				4.876		0.1
Figure 2(b)				5.178		11.6
				8.592		-3.1
beta methanol[80]	*a*	6.409	4.362	6.702	-31.9	4.6
	b	7.199	7.268	7.099	1.0	-1.4
	c	4.649	6.895	4.554	48.3	-2.0
ammonia[91]	a	5.138	5.313	4.919	3.4	-4.3
	b	5.138	5.313	4.919	3.4	-4.3
	c	5.138	5.313	4.919	3.4	-4.3
Methane Hydrate-I[83]	*a*	11.620	12.181	12.492	4.8	7.5
	b	11.620	16.164	12.490	39.1	7.5
	c	11.620	10.436	12.431	-10.2	7.0
Methane Hydrate-II[83]	*a*	11.890	12.262	12.745	3.1	7.2
	b	11.890	11.775	12.723	-1.0	7.0
	c	11.890	15.180	12.712	27.7	6.9
Methane Hydrate-H[83]	*a*	11.910	10.826	12.789	-9.1	7.4
	b	11.910	14.074	12.813	18.2	7.6
	c	9.894	11.302	10.456	14.2	5.7
WUVZIW[92]	*a*	7.400	7.393	7.355	-0.1	-0.6
	b	24.448	24.958	25.248	2.1	3.3
	c	14.265	15.399	14.707	8.0	3.1
PgC_2 Cocrystal 5 [89]	*a*	10.059	9.642	9.693	-4.1	-3.6
	b	14.556	15.927	13.886	9.4	-4.6
	c	15.558	17.731	14.945	14.0	-3.9
PgC_4 Cocrystal 6[89]	*a*	21.075	23.775	21.334	12.8	1.2
	b	12.796	13.363	12.452	4.4	-2.7
	c	18.229	18.353	17.438	0.7	-4.3
PgC_4 Cocrystal 7[89]	*a*	21.075	23.490	20.088	11.5	-4.7
	b	12.796	13.506	12.380	5.6	-3.2
	c	18.229	18.251	17.118	0.1	-6.1
ABAZOS[84]	*a*	16.338	17.022	16.306	4.2	-0.2
	b	21.933	22.805	20.839	4.0	-5.0
	c	16.338	16.870	16.310	3.3	-0.2
ABEBUF[87]	*a*	10.722	10.989	10.851	2.5	1.2
	b	10.900	11.831	11.230	8.5	3.0
	c	27.635	25.745	27.648	-6.8	0.0
ABUCIJ[86]	*a*	11.281	10.924	11.076	-3.2	-1.8
	b	17.888	18.370	17.973	2.7	0.5
	c	23.950	24.319	23.863	1.5	-0.4
SIYRAU[88]	*a*	13.456	13.861	13.257	3.0	-1.5
	b	14.394	13.727	14.647	-4.6	1.8
	c	15.935	18.019	16.945	13.1	6.3
WARWOB[85]	*a*	19.297	22.634	20.706	17.3	7.3
	b	4.616	5.599	4.589	21.3	-0.6
	c	8.705	6.954	7.972	-20.1	-8.4
				Maximum Unsigned Error	48.3	12.1
				Mean Unsigned Error	12.6	4.2

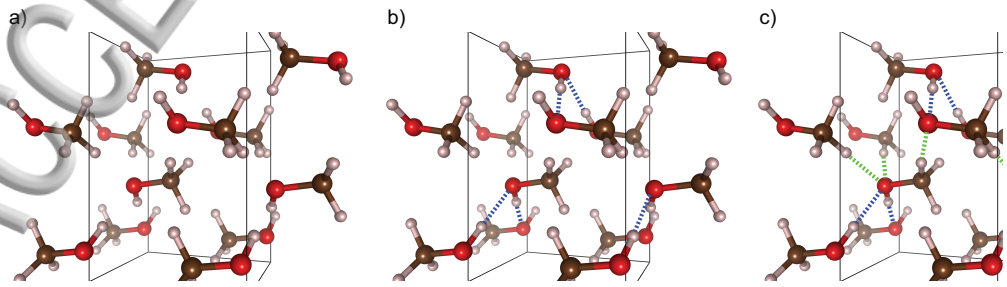

FIG. 2. α-methanol with a) no explicit hydrogen bonds, b) explicit O-H$\cdots$O hydrogen bonds only and c) adding C-H$\cdots$O hydrogen bonds to b). A single unit cell contains four methanol molecules.

ACCEPTED MANUSCRIPT

AIP Publishing

This manuscript was accepted by J. Chem. Phys. Click here to see the version of record.

IV. POTENTIAL APPLICATIONS

When used to predict solvent effects, the accuracy of the structures produced by our approach is highly dependant on the guessed positions of the solvent molecules. For the structures where case by case reasoning is impossible, using a genetic algorithm (e.g: as implemented in the Atomic Simulation Environment[93, 94]) to maximize the number of hydrogen bonds in the structure of interest can yield reasonable starting points, enabling the use of explicit bonding in automated tasks. An active area of research like water confinement in carbon nanotubes of varying diameters [95], where the hydrogen bonding is the single most important factor for structure determination, would likely benefit from cheap, systematic structure elucidation. Appropriate modifications to the ASE code are underway to allow for solvent rigid motions during optimizations, and a proof of concept script is available on github (https://github.com/DCoupry/GenAlgHbond). However, the main application of explicit hydrogen bonding is case by case reasoning for a reasonable pre-optimization, followed by electronic structure methods. More complex properties of the structures, like frequencies (for which UFF was not parametrized) or the effects of H-bond anisotropy, fall squarely outside of the scope of this paper.

V. CONCLUSIONS

We report a simple approach to explicitly treat hydrogen bonds within the Universal Force Field. The approach does not require any definition of atomic charges, which is a great technical (codes such as GULP and ADF do not include long-range and non-bonded electrostatic interactions within UFF) and practical (the definition of atomic charges is not well-defined) advantage. Moreover, it avoids the computationally costly electrostatic interaction term and is equally applicable for periodic and non-periodic systems.

The specification of hydrogen bonds with negligible (0.001) bond order increases the coordination number of the acceptor site by one, which is compatible with the definition of the main acceptor sites O and N, as well as with under-coordinated metal sites as present in metal-organic frameworks.

The approach was validated for the hydrogen bonded complexes in the S22 database. We show that it is very effective for both framework-framework hydrogen bonds and importantly, for framework-adsorbate bonds. Using this approach, both the large pore and narrow pore structures of MIL-53 can be calculated accurately with UFF. We show the generality of the approach by also applying it to hydrogen bonded crystal structures and host-guest inclusion complexes, including the environmentally and commercially important clathrates,

Finally, we note that this approach does not require any implementation, and thus works in any software that includes a UFF implementation. We propose this method to be most useful for pre-optimization and screening of static hydrogen-bonded systems.

VI. ACKNOWLEDGEMENTS

This work was partially supported by the European Commission (ERC StG C3ENV (GA 256962), MSC-EID PROPAGATE (316897), DFG FOR 2433 FlexMOF (HE 3543/31-1) and the VolkswagenStiftung.

ACCEPTED MANUSCRIPT

This manuscript was accepted by J. Chem. Phys. Click here to see the version of record.

[1] H. Li, M. Eddaoudi, M. O'Keeffe, and O. M. Yaghi, Nature **402**, 276 (1999).
[2] Y. G. Chung, J. Camp, M. Haranczyk, B. J. Sikora, W. Bury, V. Krungleviciute, T. Yildirim, O. K. Farha, D. S. Sholl, and R. Q. Snurr, Chem. Mater. **26**, 6185 (2014).
[3] S. O. Odoh, C. J. Cramer, D. G. Truhlar, and L. Gagliardi, Chem. Rev. **115**, 6051 (2015).
[4] A. Schneemann, V. Bon, I. Schwedler, I. Senkovska, S. Kaskel, and R. A. Fischer, Chem. Soc. Rev. **43**, 6062 (2014).
[5] N. C. Burtch, H. Jasuja, and K. S. Walton, Chem. Rev. **114**, 10575 (2014).
[6] J. Canivet, A. Fateeva, Y. Guo, B. Coasne, and D. Farrusseng, Chem. Soc. Rev. **43**, 5594 (2014).
[7] E.-Y. Choi, K. Park, C.-M. Yang, H. Kim, J.-H. Son, S. W. Lee, Y. H. Lee, D. Min, and Y.-U. Kwon, Chem. Eur. J. **10**, 5535 (2004).
[8] E.-C. Yang, H.-K. Zhao, B. Ding, X.-G. Wang, and X.-J. Zhao, Cryst. Growth Des. 7, 2009 (2007).
[9] N. Stock and S. Biswas, Chem. Rev. **112**, 933 (2012).
[10] J. B. DeCoste, G. W. Peterson, B. J. Schindler, K. L. Killops, M. A. Browe, and J. J. Mahle, J. Mater. Chem. A **1**, 11922 (2013).
[11] H. Furukawa, F. Gándara, Y.-B. Zhang, J. Jiang, W. L. Queen, M. R. Hudson, and O. M. Yaghi, J. Am. Chem. Soc. **136**, 4369 (2014).
[12] P. St. Petkov, G. N. Vayssilov, J. Liu, O. Shekhah, Y. Wang, C. Wöll, and T. Heine, ChemPhysChem **13**, 2025 (2012).
[13] K. Tan, N. Nijem, Y. Gao, S. Zuluaga, J. Li, T. Thonhauser, and Y. J. Chabal, CrystEngComm **17**, 247 (2015).
[14] P. Küsgens, M. Rose, I. Senkovska, H. Fröde, A. Henschel, S. Siegle, and S. Kaskel, Microporous Mesoporous Mater. **120**, 325 (2009).
[15] J. H. Cavka, S. Jakobsen, U. Olsbye, N. Guillou, C. Lamberti, S. Bordiga, and K. P. Lillerud, J. Am. Chem. Soc. **130**, 13850 (2008).
[16] D. Saha and S. Deng, J. Phys. Chem. Lett. **1**, 73 (2010).
[17] J. A. Greathouse and M. D. Allendorf, J. Am. Chem. Soc. **128**, 10678 (2006).
[18] Y. Cheng, A. Kondo, H. Noguchi, H. Kajiro, K. Urita, T. Ohba, K. Kaneko, and H. Kanoh, Langmuir **25**, 4510 (2009).
[19] I. Bezverkhyy, G. Ortiz, G. Chaplais, C. Marichal, G. Weber, and J.-P. Bellat, Microporous Mesoporous Mater. **183**, 156 (2014).
[20] J.-H. Wang, M. Li, and D. Li, Chem. Eur. J. **20**, 12004 (2014).
[21] T. A. Vu, G. H. Le, C. D. Dao, L. Q. Dang, K. T. Nguyen, Q. K. Nguyen, P. T. Dang, H. T. K. Tran, Q. T. Duong, T. V. Nguyen, and G. D. Lee, RSC Advances **5**, 5261 (2015).
[22] M. A. Addicoat, N. Vankova, I. F. Akter, and T. Heine, J. Chem. Theory Comput. **10**, 880 (2014).
[23] D. E. Coupry, M. A. Addicoat, and T. Heine, Journal of Chemical Theory and Computation **12**, 5215 (2016).
[24] A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard, and W. M. Skiff, J. Am. Chem. Soc. **114**, 10024 (1992).
[25] T. Loiseau, C. Serre, C. Huguenard, G. Fink, F. Taulelle, M. Henry, T. Bataille, and G. Férey, Chem. Eur. J. **10**, 1373 (2004).
[26] F. Salles, S. Bourrelly, H. Jobic, T. Devic, V. Guillerm, P. Llewellyn, C. Serre, G. Ferey, and G. Maurin, J. Phys. Chem. C **115**, 10764 (2011).
[27] J. Cirera, J. C. Sung, P. B. Howland, and F. Paesani, J. Chem. Phys. **137**, (2012).
[28] N. Guillou, F. Millange, and R. I. Walton, Chem. Commun. **47**, 713 (2011).
[29] G. Ortiz, G. Chaplais, J.-L. Paillaud, H. Nouali, J. Patarin, J. Raya, and C. Marichal, J. Phys. Chem. C **118**, 22021 (2014).
[30] J. M. Salazar, G. Weber, J. M. Simon, I. Bezverkhyy, and J. P. Bellat, J. Chem. Phys. **142**, (2015).
[31] G. Ferey, Zeitschrift für anorganische und allgemeine Chemie **638**, 1897 (2012).
[32] T. Heine, H. F. Dos Santos, S. Patchkovskii, and H. A. Duarte, J. Phys. Chem. A **111**, 5648 (2007), pMID: 17402718, http://pubs.acs.org/doi/pdf/10.1021/jp068988s.
[33] I. Lula, Â. L. Denadai, J. M. Resende, F. B. de Sousa, G. F. de Lima, D. Pilo-Veloso, T. Heine, H. A. Duarte, R. A. Santos, and R. D. Sinisterra, Peptides **28**, 2199 (2007).
[34] N. L. Allinger, Journal of the American Chemical Society **99**, 8127 (1977).
[35] N. L. Allinger, Y. H. Yuh, and J. H. Lii, Journal of the American Chemical Society **111**, 8551, 8566, 8576 (1989).
[36] J.-H. Lii and N. L. Allinger, Journal of Physical Organic Chemistry **7**, 591 (1994).
[37] R. S. Paton and J. M. Goodman, Journal of Chemical Information and Modeling **49**, 944 (2009).
[38] A. K. Rappe and W. A. Goddard, The Journal of Physical Chemistry **95**, 3358 (1991), http://dx.doi.org/10.1021/j100161a070.
[39] M. G. Martin, Molecular Simulation **39**, 1212 (2013).
[40] N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, and G. R. Hutchison, Journal of Cheminformatics **3**, 33 (2011).
[41] A. C. T. van Duin, S. Dasgupta, F. Lorant, and W. A. Goddard, *The Journal of Physical Chemistry A*, J. Phys. Chem. A **105**, 9396 (2001).
[42] Y. Bouteiller, B. Tremblay, and J. P. Perchard, Chem. Phys. **386**, 29 (2011).
[43] E. A. Serov, M. A. Koshelev, T. A. Odintsova, V. V. Parshin, and M. Y. Tretyakov, Phys. Chem. Chem. Phys. **16**, 26221 (2014).
[44] Y. Song, H. Chen, C. Zhang, Y. Zhang, and Y. Yin, Eur. Phys. J. D **68**, 242 (2014).

ACCEPTED MANUSCRIPT

This manuscript was accepted by J. Chem. Phys. Click here to see the version of record.

[45] Z. P. Wang, P. M. Dinh, P. G. Reinhard, and E. Suraud, Laser Physics **24**, 106004 (2014).
[46] E. Ronca, L. Belpassi, and F. Tarantelli, ChemPhysChem **15**, 2682 (2014).
[47] K. Morokuma and L. Pedersen, J. Chem. Phys. **48**, 3275 (1968).
[48] K. Morokuma and J. R. Winick, J. Chem. Phys. **52**, 1301 (1970).
[49] J. G. C. M. van Duijneveldt-van de Rijdt and F. B. van Duijneveldt, J. Chem. Phys. **97**, 5019 (1992).
[50] J. A. Odutola and T. R. Dyke, J. Chem. Phys. **72**, 5062 (1980).
[51] T. R. Dyke, K. M. Mack, and J. S. Muenter, J. Chem. Phys. **66**, 498 (1977).
[52] P. Jurecka, J. Sponer, J. Cerny, and P. Hobza, Phys. Chem. Chem. Phys. **8**, 1985 (2006).
[53] J. Witte, M. Goldey, J. B. Neaton, and M. Head-Gordon, J. Chem. Theory Comput. **11**, 1481 (2015).
[54] J. Gale and A. Rohl, Mol. Simul. **29**, 291 (2003).
[55] J. D. Gale, Zeitschrift fur Kristallographie **220**, 552 (2005).
[56] "demonnano," http://physics.jacobs-university.de/theine/research/deMon (2009).
[57] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, "Gaussian 09 Revision D.1," Gaussian Inc. Wallingford CT (2009).
[58] F. H. Allen, Acta Crystallographica Section B **58**, 380 (2002).
[59] D. W. Tomlin, T. M. Cooper, S. M. Cline, J. M. Hughes, and W. W. Adams, Acta Crystallographica Section C **53**, 1815 (1997).
[60] C. M. Gramaccioli and R. E. Marsh, Acta Crystallographica **21**, 594 (1966).
[61] X. Li, B.-L. Wu, C.-Y. Niu, Y.-Y. Niu, and H.-Y. Zhang, *Crystal Growth & Design*, Cryst. Growth Des. **9**, 3423 (2009).
[62] Y. Zhang, M. K. Saha, and I. Bernal, CrystEngComm **5**, 34 (2003).
[63] C. Bromant, H. Flemig, W. Nika, I. Pantenburg, and G. Meyer, Zeitschrift für anorganische und allgemeine Chemie **632**, 851 (2006).
[64] M. S. Grigorev, A. A. Bessonov, A. I. Yanovsk, Y. T. Struchkov, and N. N. Krot, Radiokhimiya(Russ.)(Radiochem.) **33**, 46 (1991).
[65] M.-L. Guo and C.-H. Guo, Acta Crystallographica Section C **62**, m7 (2006).
[66] M. Mizutani, N. Maejima, K. Jitsukawa, H. Masuda, and H. Einaga, Inorg. Chim. Acta **283**, 105 (1998).
[67] X.-Y. Huang, K.-F. Yue, J.-C. Jin, J.-Q. Liu, C.-J. Wang, Y.-Y. Wang, and Q.-Z. Shi, Inorg. Chem. Commun. **13**, 338 (2010).
[68] C.-J. Wu and J.-M. Shi, Chinese Journal of Structural Chemistry **23**, 1161 (2004).
[69] A. Rodriguez-Dieguez, A. Salinas-Castillo, A. Sironi, J. M. Seco, and E. Colacio, CrystEngComm **12**, 1876 (2010).
[70] H.-N. Wang, X. Meng, C. Qin, X.-L. Wang, G.-S. Yang, and Z.-M. Su, Dalton Trans. **41**, 1047 (2012).
[71] W. Li, H.-P. Jia, Z.-F. Ju, and J. Zhang, Dalton Trans. , 5350 (2008).
[72] S.-R. Zheng, S.-L. Cai, Z.-Z. Wen, J. Fan, and W.-G. Zhang, Polyhedron **38**, 190 (2012).
[73] F. Wang, Y.-X. Tan, H. Yang, Y. Kang, and J. Zhang, Chem. Commun. **48**, 4842 (2012).
[74] T. Kawasaki, C. Kachi-Terajima, T. Saito, and T. Kitazawa, Bull. Chem. Soc. Jpn. **81**, 268 (2008).
[75] H. Xu and Z. Xu, *Metal Organic Frameworks*, Microporous Mesoporous Mater. **157**, 33 (2012).
[76] M. C. Das, S. K. Ghosh, E. C. Sañudo, and P. K. Bharadwaj, Dalton Trans. , 1644 (2009).
[77] J. An, S. J. Geib, and N. L. Rosi, *Journal of the American Chemical Society*, J. Am. Chem. Soc. **132**, 38 (2010).
[78] C. M. Gramaccioli, Acta Crystallographica **21**, 600 (1966).
[79] B. Torrie, S.-X. Weng, and B. Powell, Molecular Physics **67**, 575 (1989).
[80] K. J. Tauer and W. N. Lipscomb, Acta Crystallographica **5**, 606 (1952).
[81] B. Gbaruko, J. Igwe, P. Gbaruko, and R. Nwokeoma, Journal of Petroleum Science and Engineering **56**, 192 (2007).
[82] E. D. Sloan, Nature **426**, 353 (2003).
[83] A. Lenz and L. Ojamäe, The Journal of Physical Chemistry A **115**, 6169 (2011).
[84] L. Yeo and K. D. Harris, Canadian Journal of Chemistry **77**, 2105 (1999).
[85] Y. L. Chang, M. A. West, F. W. Fowler, and J. W. Lauher, J. Am. Chem. Soc. **115**, 5991 (1993).
[86] S. Apel, S. Nitsche, K. Beketov, W. Seichter, J. Seidel, and E. Weber, J. Chem. Soc., Perkin Trans. 2 , 1212 (2001).
[87] S. W. Gordon-Wylie, E. Teplin, J. C. Morris, M. I. Trombley, S. M. McCarthy, W. M. Cleaver, , and G. R. Clark, Crystal Growth & Design **4**, 789 (2004).
[88] S. G. Telfer, N. D. Parker, R. Kuroda, T. Harada, J. Lefebvre, and D. B. Leznoff, Inorganic Chemistry **47**, 209 (2008).
[89] D. A. Fowler, C. R. Pfeiffer, S. J. Teat, C. M. Beavers, G. A. Baker, and J. L. Atwood, CrystEngComm **16**, 6010 (2014).
[90] E. A. Engel, B. Monserrat, and R. J. Needs, Phys. Rev. X **5**, 021033 (2015).
[91] I. Olovsson and D. H. Templeton, Acta Crystallographica **12**, 832 (1959).
[92] H. J. Lindner, F. W. Lichtenthaler, K. Fujita, C. Yang, D.-Q. Yuan, and Y. Nogami, Acta Crystallographica Section E **59**, o387 (2003).
[93] S. R. Bahn and K. W. Jacobsen, Comput. Sci. Eng. **4**, 56 (2002).
[94] L. B. Vilhelmsen and B. Hammer, J. Chem. Phys. **141**, 044711 (2014), http://dx.doi.org/10.1063/1.4886337.

ACCEPTED MANUSCRIPT

This manuscript was accepted by J. Chem. Phys. Click here to see the version of record.

[95] T. A. Pascal, W. A. Goddard, and Y. Jung, Proceedings of the National Academy of Sciences **108**, 11794 (2011), http://www.pnas.org/content/108/29/11794.full.pdf.

ACCEPTED MANUSCRIPT

AIP Publishing

This manuscript was accepted by J. Chem. Phys. Click here to see the version of record.

ACCEPTED MANUSCRIPT

AIP Publishing

This manuscript was accepted by J. Chem. Phys. Click here to see the version of record.

ACCEPTED MANUSCRIPT