

A tight binding model for water

A. T. Paxton^{a)} and J. J. Kohanoff

Atomistic Simulation Centre, School of Mathematics and Physics, Queen's University Belfast, Belfast BT7 1NN, United Kingdom

(Received 28 September 2010; accepted 15 November 2010; published online 28 January 2011)

We demonstrate for the first time a tight binding model for water incorporating polarizable oxygen atoms. A novel aspect is that we adopt a “ground up” approach in that properties of the monomer and dimer only are fitted. Subsequently we make predictions of the structure and properties of hexamer clusters, ice-XI and liquid water. A particular feature, missing in current tight binding and semiempirical Hamiltonians, is that we reproduce the almost two-fold increase in molecular dipole moment as clusters are built up toward the limit of bulk liquid. We concentrate on properties of liquid water, particularly dielectric constant and self diffusion coefficient, which are very well rendered in comparison with experiment. Finally we comment on the question of the contrasting densities of water and ice which is central to an understanding of the subtleties of the hydrogen bond. © 2011 American Institute of Physics. [doi:10.1063/1.3523983]

I. INTRODUCTION

We are motivated to construct a model of water for use in atomistic simulation within the framework of the self consistent polarizable ion tight binding (TB) theory.^{1,2} The intention is that this should bridge the outstanding gap between the local density functional theory (DFT) and classical molecular mechanics models. The risk is that the model will prove too slow computationally to replace existing, highly effective classical models while being unable to capture enough of the physics and chemistry (especially of the hydrogen bond) to provide a useful imitation of the first principles approach. To address the latter concern we have adopted a “ground up” strategy in which parameters of the TB model are fitted first to the monomer and lastly to the dimer, so that all the properties of the solid and liquid are *predictions* of the model. The former concern may be settled by observing at once that our model is capable of propagating 500 self consistent molecular dynamics time steps for a unit cell of 128 water molecules in one CPU hour over four threads of a commodity multicore 2.4 GHz processor—12 ps per day using a time step of 1 fs.

The structure of the paper is this. We begin in Sec. II by describing TB models constructed intuitively using parameters that are selected with little adjustment from published TB models, in particular drawing on Walter Harrison's seminal scheme.³ This should give the reader confidence that the TB approach contains the essential physics. In Sec. III we demonstrate a model that is constructed by tuning the parameters using a complex genetic algorithm that provides us with the most accurate rendering of the properties of the monomer, in particular its polarizability and force constants, and the dimer, in particular its geometry and the shape of the energy versus length of the hydrogen bond as deduced from high level quantum chemical calculations. We apply this model first to the structure and energetics of the water hexamer and to ice (Secs. IV and V); and in Sec. VI we demon-

strate its predictive power in respect of the properties of liquid water, especially the radial distribution function, dielectric constant and self diffusion coefficient. We discuss our results in Sec. VII in which we also consider the question of the relative densities of ice and water and we conclude in Sec. VIII.

II. INTUITIVE MODELS

The self consistent polarizable ion tight binding theory was proposed in the first instance to address the atomic and electronic structure of ceramic oxides whose properties are dominated by the highly polarizable oxygen anion.^{1,4} Later it was applied to problems involving polarizability in molecular physics.² The method is described in detail in a recent textbook⁵ and lecture article.⁶ We should point out a close similarity between our theory and that of the SCC-DFTB method.⁷ The principal difference is that we go beyond the point charge approximation and allow the self consistent development of point charge multipoles; this is accompanied by associated higher angular momentum components of the electrostatic potential allowing us to describe crystal field effects in a self consistent manner. The second difference is in our approach to the choice of model parameters, which we do largely by rational intuition rather than direct computation and fitting to density functional total energies.⁷

A. Monomer

1. A simple non-self-consistent model

One might construct a very simple, non self consistent model for the monomer in the spirit of Harrison's “solid state table.”³ In this way we place a $1s$ orbital on the hydrogen atoms and $2s$ and $2p$ orbitals on the oxygen with off diagonal Hamiltonian matrix elements, or *hopping integrals* being

$$V_{\ell\ell'x} = \alpha \eta_{\ell\ell'x} \frac{\hbar^2}{2m} \frac{1}{r^2},$$

^{a)}Electronic mail: tony.paxton@qub.ac.uk.

in which $\eta_{ss\sigma} = -1.4$ and $\eta_{sp\sigma} = 1.84$. Here, r is the bond length and m the electron mass. χ stands for either σ or π bonding. The factor α is ours, since we have found that $\alpha = \frac{1}{2}$ produces a HOMO–LUMO gap in better agreement with experiment. Specifically in order to suppress intermolecular O–H interactions and generally to achieve short ranged hopping, we adopt a modified scaling known as GSP, after its authors,⁸ such that

$$V(r) = V(r_0) \left(\frac{r_0}{r}\right)^n \exp \left[n \left(-\left(\frac{r}{r_c}\right)^{n_c} + \left(\frac{r_0}{r_c}\right)^{n_c} \right) \right]. \quad (1)$$

The virtue of the GSP scaling is that it has the same value at r_0 (but not slope) as the Harrison form if we use $n = 2$ while it decays exponentially with distance under the control of the cut off parameters, r_c and n_c ; r_0 is conventionally the equilibrium bond length, but as with all the parameters this could depend on the orbital. For this reason there are implicit $\{\ell\ell'\chi\}$ subscripts on the parameters in (1) that we have suppressed. In this intuitive model we set $n = 2$, $n_c = 4$ and $r_c = 1.8r_0$; r_0 is our target bond length, 1.8094 bohr. The *diagonal* matrix elements of the Hamiltonian in the solid state table³ are taken to be Hartree–Fock term values, namely $\varepsilon_s = -1$ Ry on the H-atom, the values for oxygen being $\varepsilon_s = -2.142$ Ry and $\varepsilon_p = -1.038$ Ry. We should point out that in this, and indeed all our subsequent models, the bond angle is dominated by the $s-p$ splitting on the oxygen atom. This is clear since in the absence of on-site $s-p$ hybridization the bond angle would be exactly 90° . Then as ε_s is raised relative to ε_p the bond angle increases accordingly.⁹ It is notable that the bond angle is very accurately rendered using the atomic term values.

2. Intuitive point charge model

We will find that charge transfer which leads to an electrostatic repulsion between the H-atoms has a very small influence on the bond angle. Turning to the charge transfer, this is where the self consistency appears at the simplest level. After solution of the Schrödinger equation using the Harrison TB Hamiltonian, examination of the eigenvectors will reveal Mulliken charges at the atomic sites. One might ignore these, or compute their contribution to the energy using elementary electrostatics. Since charge transfer will always serve to lower the energy, if the electrostatic potential arising due to Mulliken charges is added to the Hamiltonian and a further solution to the Schrödinger equation is made then, as this process is repeated, charge will continue to accumulate at the more electronegative site resulting in a Coulomb catastrophe. This is corrected by a term in the Hamiltonian, an energy proportional to the *charge squared*, such that the *inter-site* Coulomb energy is balanced by the on-site *Hubbard energy* so called.¹⁰ These two are expected largely to cancel in many cases so that it is admissible to neglect these and stick to a non-self-consistent model.¹⁰ On the other hand the self consistent inclusion of charge transfer is exactly what is needed to describe mixed covalent ionic bonding as we expect to find in water. Therefore at the next level of an intuitive model we specify Hubbard– U parameters which we choose to be 1 Ry on the H-atoms and which we adjust on the oxygen until the self consistent dipole moment has the value, 1.86 Debye (D), that

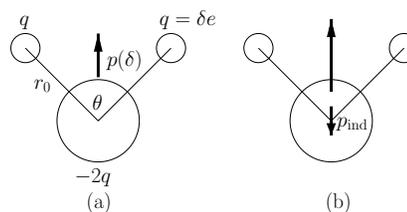


FIG. 1. A cartoon of the water monomer. The target bond angle is $\theta = 104.26^\circ$, and the bond length is $r_0 = 1.809$ bohr. In (a), the dipole moment arising from charge transfer is indicated and it is a simple matter to adjust the charge transfer, δ , by a choice of Hubbard– U to achieve the target dipole moment, 1.86 D. If we admit that the oxygen is *polarizable* (b) then the self consistent problem can also be solved. The charge transfer is increased by reducing U so that a larger charge transfer derived dipole moment exists which is in turn partially canceled by the induced dipole, again to achieve the target dipole moment. See the text for details of how this is done.

is observed. We can now summarize this, our *point charge model*, by reference to Fig. 1(a). A fractional amount, δ of an electron is transferred from each H-atom to the oxygen; the resulting charges give rise to a dipole moment,

$$p(\delta) = 2r_0\delta e \cos \frac{1}{2}\theta, \quad (2)$$

which, once we have fixed our target bond angle, θ , and bond length, depends only upon δ as $p(\delta) = 5.65\delta$ D. Hence since we know our target dipole moment it is an easy matter to adjust our one free parameter, the oxygen Hubbard– U , to obtain a point charge TB model for the electronic structure. There remains to fix the standard pair potential, which in the language of the tight binding bond model is said to account for the all terms except for the bandstructure energy in the Harris–Foulkes functional.^{6,11} We adopt the GSP form (1) for this allowing two free parameters, namely, the prefactor which we call $A(r_0)$ and the exponent n . We fit these exactly to the bond length and symmetric vibrational force constant from experiment, while setting $n_c = 6$ and $r_c = 2.9$ bohr. This completes the description of our point charge model.

3. Intuitive dipole model

Let us now admit that the oxygen anion is a polarizable species and the dipole moment created by the charge transfer will result in an electric field acting at the oxygen atomic site. This will *induce* a dipole moment which we model as a point dipole

$$p_{\text{ind}} = -\alpha_{\text{O}}(\delta) \frac{2e\delta}{r_0^2} \cos \frac{1}{2}\theta, \quad (3)$$

in which α_{O} is a notional polarizability of the oxygen anion. If we knew what this was, and in principle it should depend on the oxidation state of the anion (as indicated by its dependence upon δ), then our *dipole model* would follow immediately without any further guessing of parameters. We would write using (2) and (3)

$$\frac{p_{\text{ind}}}{p(\delta)} = -\frac{\alpha_{\text{O}}(\delta)}{r_0^3},$$

while

$$p(\delta) + p_{\text{ind}} = 1.86 \text{ D},$$

which could be solved if we knew $\alpha_{\text{O}}(\delta)$. In fact examining the known polarizabilities of the neutral first row atoms, nitrogen through neon, we found it to be a rather linear function of the valence leading to an interpolation formula,

$$\alpha_{\text{O}}(\delta) \approx 0.8 - 0.46\delta \quad (\text{\AA}^3). \quad (4)$$

Unfortunately using the resulting value of α_{O} leads to a polarization catastrophe: The bond angle tends to zero to allow the electric field to maximize and induce the largest possible dipole moment at the oxygen site. Therefore, to construct a sensible intuitive model, we need to guess a smaller polarizability and we choose 0.25\AA^3 . Thereafter it should be straight forward to readjust the oxygen Hubbard- U until the target dipole moment is achieved. However the bond angle is reduced in this process from the same origin that gives rise to the polarization catastrophe and it is necessary to raise the oxygen ε_s . But it is simple enough to adjust these two parameters simultaneously, after which fitting the pair potential to r_0 and the symmetric vibrational force constant finalizes the form of our intuitive *dipole model*.

4. Crystal field

We should make it clear at this point that our self consistent TB theory does not deal directly with oxygen polarizability. Instead the action of the electric field at an atomic site is captured by a term in the Hamiltonian in addition to the non self consistent Hamiltonian such that we have

$$H = H_0 + H',$$

and H' amounts to a self consistent field having the form

$$H'_{\mathbf{R}L'\mathbf{R}L''} = U_{\mathbf{R}} q_{\mathbf{R}} \delta_{L'L''} + \sum_L V_{\mathbf{R}L} \Delta_{\ell'\ell''\ell} C_{L'L''L}, \quad (5)$$

and the polarizability is expressed through new parameters, $\Delta_{\ell'\ell''\ell}$ which are generalizations of the parameters $\langle r^\ell \rangle$ of crystal field theory.^{12,13} In (5) \mathbf{R} labels atomic sites and L is a composite angular momentum index $L = \{\ell m\}$; $q_{\mathbf{R}}$ is the Mulliken charge transfer at site \mathbf{R} and $U_{\mathbf{R}}$ is the Hubbard- U at that site. The second term acts to split the otherwise degenerate atomic energy levels ε_ℓ through the appearance of on-site, off diagonal matrix elements of the Hamiltonian. $C_{L'L''L}$ are Gaunt integrals (A1) and $V_{\mathbf{R}L}$ are expansion coefficients of the electrostatic potential (energy) seen by an electron at site \mathbf{R} , this potential having been expanded into spherical waves¹⁴

$$V_{\mathbf{R}}(\mathbf{r}) = \sum_L V_{\mathbf{R}L} r^\ell Y_L(\mathbf{r}).$$

Poisson's equation relates the electrostatic potential to the multipole moments $Q_{\mathbf{R}'L'}$ of the Mulliken charge transfer

$$V_{\mathbf{R}L} = e^2 \sum_{\substack{\mathbf{R}'L' \\ \mathbf{R}' \neq \mathbf{R}}} \tilde{B}_{\mathbf{R}L\mathbf{R}'L'} Q_{\mathbf{R}'L'},$$

in which \tilde{B} is a sort of generalized Madelung matrix, equal to $|\mathbf{R} - \mathbf{R}'|^{-1}$ for point charges. The general case is

TABLE I. Predictions of the intuitive point charge and dipole models for the water monomer. The “genetic model” is described in Sec. III. Experimental data are taken from the CRC Handbook (Ref. 15). E_{gap} is the HOMO-LUMO gap.

Model	δ	α_{O} (\AA^3)	ν_1	ν_2	ν_3	$\alpha_{\text{H}_2\text{O}}$ (\AA^3)	E_{coh} (Ry)	E_{gap} (Ry)
Point	0.33	—	1.029	0.099	1.002	2.3	0.77	0.81
Dipole	0.45	0.25	1.029	0.104	0.840	1.8	0.82	1.03
Genetic	0.47	0.27	1.029	0.065	1.061	1.5	0.76	0.66
Exp.	—	—	1.029	0.100	1.062	1.4	0.75	0.82

described in the Appendix. Finally, the multipole moments are obtained from the self consistent eigenvectors, c , of the Hamiltonian^{1,5,6}

$$Q_{\mathbf{R}L} = \sum_{L'L''} \sum_n f_n \bar{c}_{\mathbf{R}L'}^n c_{\mathbf{R}L''}^n \Delta_{\ell'\ell''\ell} C_{L'L''L},$$

where f_n is the occupation number of state n . It is notable that the crystal field parameters play a dual role as multipole strengths in the theory.

For the case of our water model the only polarizability parameter we need is the quantity Δ_{spp} at the oxygen site. The interpretation of $\Delta_{\ell'\ell''\ell}$ is that the ℓ -component of the potential (in the case $\ell = 1$ the electric field) causes an on-site coupling in the Hamiltonian of the ℓ' and ℓ'' orbitals (in this case the s and p are coupled by the crystal field). In its role as multipole strength parameter, it describes the contributions of the on-site ℓ' and ℓ'' components of the eigenvectors to the development of the ℓ -pole moment of the charge (in this case the dipole). This reflects the well-known requirement to mix s and p orbitals on-site if one is to develop a dipole moment of the charge.

5. Predicted properties of the monomer

The question now is, what is the predictive power of these two models? Table I gives an answer. For each model we show the charge transfer, δ , larger in the dipole model since its derived dipole moment is to be partially canceled by the induced moment to achieve the target experimental value. The symmetric force constant associated¹⁶ with the vibrational frequency ν_1 is fitted in the pair potential, but the angular and asymmetric frequencies are *predictions*. First we note that ν_2 is well rendered and we remind the reader that the angular forces in TB arise from the $sp\sigma$ matrix elements, transforming like the Slater-Koster table.¹⁸ The predictions of the asymmetric force constant are revealing. It is observed, and also predicted in the local density functional approximation, that the asymmetric mode is stiffer than the symmetric. Intuitively one expects the opposite: since the symmetry is broken there should be greater opportunity to relax the electronic structure and hence the curvature of the energy-stretch curve ought to be smaller. In fact this wrong result is predicted in our models and the effect is exacerbated in the dipole model for which the direction of oxygen site polarization is free to rotate and hence lower the energy. Therefore both models at the intuitive level invert the ordering of ν_1 and ν_3 and the dipole model is of course the worse culprit. The total

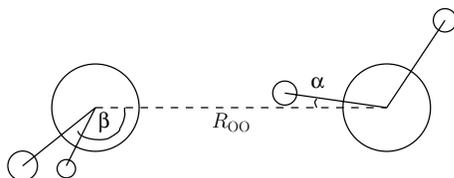


FIG. 2. A cartoon of the water dimer. Strictly the two hydrogen atoms at the left are eclipsed and β is the angle between the O–O bond and the plane of the left hand molecule. In the instance that $\alpha = 0$ we talk of a *linear* hydrogen bond.

polarizability of the monomer is well known and can be calculated simply by applying an electric field and recording the resulting dipole moment.² We find that the polarizability is in reasonable agreement with experiment, the dipole model being quite accurate in this regard. (Note that even in our point charge model the molecule is polarizable, through interatomic charge transfer in the electric field.) The cohesive energy of the monomer with respect to its atoms is rather well reproduced in either model.

B. Dimer

The curious structure of the water dimer *in vacuo* is rather well known. We show in Fig. 2 a cartoon of the geometry with the principal length and bond angles indicated. In order to adapt our intuitive models to describe the dimer we need to make TB parameters for the O–O interactions. We note that the O–O distance, marked R_{OO} in Fig. 2 is not much different from that in a typical ceramic oxide, such as zirconia; therefore as a first guess, with some modifications we try the hopping parameters that we previously used in molecular statics and dynamics calculations in that material.^{4,19} In that work we employed no O–O pair potential since these anions were second neighbours in the fluorite lattice. For the intuitive model for water we choose a simple pair potential of the form

$$\varphi(r) = A r^{-m} e^{-pr}, \quad (6)$$

having A , m and p all greater than or equal to zero. We adjusted these three parameters by eye to obtain a reasonable fit of the energy versus O–O bond length (at fixed angles β and α) calculated with the GAMESS program^{20,21} using coupled clusters at the CCSD(T) level.²² We can then ask, how well do the intuitive models predict the angles β and α shown in Fig. 2 which have been calculated by Klopper *et al.*²³ and found to be $\alpha = 5.5^\circ$ and $\beta = 124.4^\circ$, while $R_{OO} = 5.50$ bohr. We use molecular statics to relax the water dimer and we find in the point charge model, $R_{OO} = 5.34$ bohr, $\alpha = 13.13^\circ$ and $\beta = 172.7^\circ$; while the dipole model predicts, $R_{OO} = 5.50$ bohr, $\alpha = 5.9^\circ$ and $\beta = 102.5^\circ$. This is very encouraging. First the models do not predict that the two molecules line up with their monomer dipoles aligned antiparallel—a configuration one might well regard as natural from an elementary electrostatic viewpoint, and one that DFT in the generalized gradient approximation²⁴ (GGA) predicts to be remarkably close (but higher) in energy than the observed orientation of Fig. 2. Second, the dipole model at least makes a good rendering of the angles β and α considering the

subtlety of the structure. Furthermore the fact that the dipole model is evidently superior to the point charge model in this regard argues strongly for the correctness in accounting for the oxygen polarizability in the construction of a TB model for water.

III. FITTED MODELS

We hope that the reader is now convinced that the self consistent polarizable ion TB theory provides a proper framework for the description of the structure, bonding and energetics of the water monomer and dimer. In that case there should be no objection to the next development which is to seek refined Hamiltonian matrix elements, pair potentials and their bond length dependences via a thorough search in the parameter space. To this end, we have adopted Schwefel’s multi-membered evolution strategy.^{25,26} As we have seen in Sec. II a set of TB parameters may be used to predict certain properties. Indeed any set of parameters will give rise to values of a number of chosen properties which can be compared to target values taken from experiment or coupled cluster calculations. A suitably weighted sum of squares of differences becomes an *objective function* and the aim is to find that set of parameters which gives the smallest value of the objective function, subject to certain constraints on the parameters so that they continue to be physically motivated. This last point is very important—we cannot accept a model whose parameters violate certain basic truths; for example $\varepsilon_p - \varepsilon_s > 0$, $V_{ss\sigma} < 0$ and so on. We also have an instinct for the approximate sizes and appropriate scaling laws for the hopping integrals which are informed by canonical band theory.⁹ The procedure is this. A computer script is written that, for a given set of TB parameters, calculates the required set of properties and evaluates the objective function. The evolution algorithm repeatedly calls this script with TB parameters that belong to successive parents and offspring until a global minimum is hopefully found. In the process, a huge set of parameters is tested and we are free to browse this set to select attractive looking models and to investigate how certain properties are governed by particular parameters.

A. Monomer

In the case of the monomer, the bond length, dipole moment, frequencies and polarizability are used in the fitting and the resulting model is added here to Table I. We note that the correct ordering of ν_1 and ν_3 is achieved, at the expense of an angular force constant that is significantly too small. We believe that for the simulations of liquid water, to be described in Sec. VI below, the polarizability of the monomer plays a large role and it is therefore important at this stage that this is now properly rendered in the “genetic model,” so called.

B. Dimer

It is a feature of our “ground up” philosophy toward the construction of TB models that, having established a model for the monomer, we do not further adjust its parameters; and

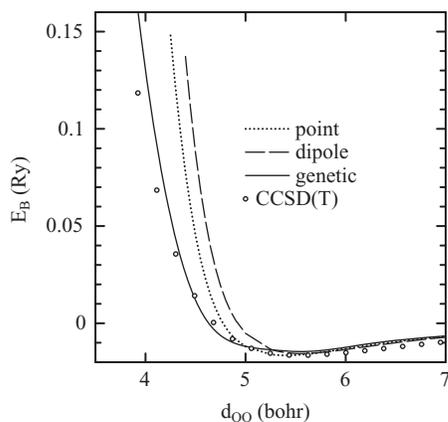


FIG. 3. Cohesive energy versus O–O distance in the dimer at fixed $\alpha = 5.5^\circ$ and $\beta = 124.4^\circ$, using the two intuitive and the fitted models, compared to quantum chemical calculations at the CCSD(T) level using the GAMESS suite of programs (Refs. 20 and 21).

continue to the fitting of the oxygen–oxygen interactions only to the properties of the dimer. Our objective function is constructed using the relaxed geometry of the dimer, in particular, R_{OO} , α and β obtained from a full geometry optimization. In addition we do calculations of the cohesive energy at a set of O–O distances at fixed values of α and β , namely their target values as found in the literature.²³ Five such data points are included in the objective function. Fig. 3 shows the energy that we have calculated^{20,21} in coupled clusters at the CCSD(T) level compared to the two intuitive models and our final “genetic” model. We note that the exponential pair potential (6), which is typical for standard solid state TB models is too steeply repulsive compared to the CCSD(T) curve. This is clearly a peculiar signature of the hydrogen bond. We have chosen a softer pair potential based on the very first, quadratic, pair potential proposed by Jim Chadi to describe the harmonic properties of semiconductor $s-p$ bonding,²⁷

$$\varphi(r) = U_1\epsilon + U_2\epsilon^2,$$

where $\epsilon = (r - r_0)/r_0$ is the fractional change in bond length relative to some reference distance, r_0 . Of course a parabola does not describe the CCSD(T) curve in Fig. 3 at large O–O distances; we replace the potential $\varphi(r)$ with a fifth degree polynomial beyond some distance, r_1 , which is constructed to be continuous and twice differentiable at r_1 and to vanish smoothly at a cut-off distance, r_c . This is consistent with modern practice in the bond order potentials.²⁸ The behavior of this curve at long distances is dictated by electrostatics and van der Waals interactions. The latter are explicitly absent in our model while the electrostatics are included through the self consistent Hamiltonian (Sec. II A 4).

All of the parameters of our three models are gathered into Table II. The properties of the dimer are displayed in Table III.

C. The hydrogen bond

It is worthwhile to make a few preliminary remarks concerning the hydrogen bond at this stage of the development. In a sense the water dimer in its geometry of Fig. 2 is an archety-

TABLE II. Parameters of our three models—intuitive point charge and dipole models and the model fitted by genetic evolution. See the text for meanings of the various symbols. All quantities are in atomic Rydberg units.

Model:	Point	Dipole	Genetic
On site parameters			
ϵ_s (H)	−1	−1	−1
ϵ_s (O)	−2	−1.45	−1.51
ϵ_p (O)	−1.038	−1.038	−1.20
U (H)	1	1	1.08
U (O)	0.885	0.77	1.16
Δ_{spp}	—	−0.84	−0.924
O–H hopping integrals			
$V_{ss\sigma}$	−0.428	−0.428	−0.348
$V_{sp\sigma}$	0.550	0.550	0.313
$n_{ss\sigma}$	2	2	1.48
$n_{sp\sigma}$	2	2	1.98
n_c	4	4	4.04
r_c/r_0	1.8	1.8	1.92
O–H pair potential			
A	0.848	0.937	0.552
n	3.190	3.045	3.362
n_c	6	6	6.04
r_c	2.9	2.9	3.04
O–O hopping integrals			
$V_{ss\sigma}$	−0.072	−0.072	−0.080
$V_{sp\sigma}$	0.084	0.084	0.050
$V_{pp\sigma}$	0.06	0.06	0.00012
$V_{pp\pi}$	−0.01	−0.01	−0.004
$n_{ss\sigma}$	2	2	2
$n_{sp\sigma}$	2	2	2
$n_{pp\sigma}$	3	3	3
$n_{pp\pi}$	3	3	3
n_c	6	6	4
d_0	5.6	5.6	5
r_c	9	9	6.8
O–O pair potential			
A	10^5	1.5×10^5	—
m	9.7	6	—
p	0	1.2	—
U_1	—	—	0.010
U_2	—	—	0.647
r_0	—	—	5.992
r_1	—	—	5.494
r_c	—	—	6.110

pal example; and one may ask what is the nature of such a bond and what determines the values of R_{OO} , α and β ? The repulsive force which we model with a pair potential must arise in real life from the closed shell repulsion and valence–core overlap. Our view is that the principal attractive force comes about from the mutual static polarizability of the two monomers and that in addition the point charge electrostatics plays a role in the angular disposition of the molecules. On the other hand we recall again that the O–O distance is only moderately smaller than that in a typical metal oxide which can

TABLE III. Properties predicted, or fitted, of the water dimer. Target values are taken from Klopper *et al.* (Ref. 23)

Model:	Point	Dipole	Genetic	Target
R_{OO} (bohr)	5.3423	5.5011	5.5091	5.4991
α (deg.)	13.1	5.9	3.0	5.5
β (deg.)	172.7	102.5	113.7	124.4
E_{coh} (mRy)	-18.2	-16.8	-15.1	-15.9

be modeled as mixed ionic-covalent bonding.¹ For this reason we include the *sp* and *pp* bonding integrals in our model. If we leave these out, we can make a plausible model, but we find we cannot reproduce the angles well. One can imagine that negotiation between the σ and π bonds is what results in the final choice of the angle β .

IV. THE HEXAMER

The nature of the hydrogen bond as manifested in the dimer can be further studied in the structure and properties of the hexamer *in vacuo*. Four isomeric structures have been identified,²⁹ and remarkably these are practically degenerate in energy even though their shapes and the number of hydrogen bonds are very different.²⁹ The four isomers are displayed in Fig. 4 which shows the structures after geometry optimization using the genetic fitted TB model of Sec. III. As in the dimer, the hydrogen bonds are not linear: again one sees the angle equivalent to α in the dimer is some 5° . The exception to this is the cyclic hexamer which shows almost exactly linear hydrogen bonds. Significantly this is the least stable of the four. Although the TB model overestimates the energy differences with respect to the most stable, the prism, by as much as an order of magnitude it is very significant that the ordering predicted using quantum chemistry at the MP2 level²⁹ is reproduced by our model. This is especially noteworthy since of 18 quantum chemistry methods employed by Santra *et al.*²⁹ only quantum Monte Carlo, MP2 and CCSD(T) predict the prism to be the most stable isomer; moreover Hartree-Fock, the hybrid functional B3LYP, and the BLYP function predict the ordering quite reversed, namely the book is most stable

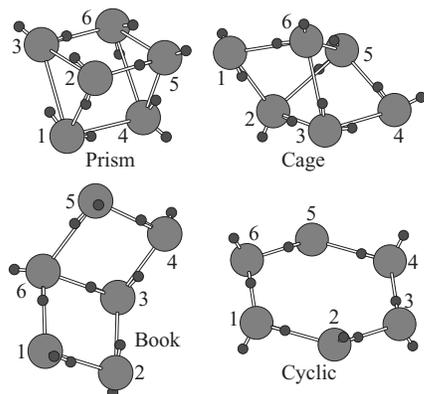


FIG. 4. Structures of the four hexamer isomers considered in the text, relaxed using the fitted genetic TB model. The labeling of the oxygen atoms is such that the O-O bond lengths can be read off in Table V.

TABLE IV. Dipole moments on individual water molecules found in the monomer, dimer, trimer ... hexamer using our fitted genetic, and intuitive point charge TB models and compared to available similar results using the GGA to density functional theory (Ref. 31). Note in connection with remarks in Sec. VI that the point charge model does display the expected increase in dipole moment with size of cluster, but this is non monotonic, the trimer being an exception to the trend, and the magnitude of the effect is probably exaggerated by the anomalously large polarizability of the monomer (Table I).

Cluster size:	1	2	3	4	5	6	...	Liq.
Point	1.87	2.07	2.33	2.31	2.16	2.25		
		2.13	2.33	2.12	2.38	2.31		
			2.33	2.23	2.10	2.54		
				2.34	2.39	2.39		
					2.37	2.33		
						2.45		
Mean	1.87	2.10	2.33	2.25	2.28	2.38	...	2.51
Genetic	1.86	2.00	2.13	2.14	2.15	2.23		
		2.00	2.15	2.12	2.15	2.46		
			2.14	2.26	2.31	2.55		
				2.33	2.46	2.73		
					2.46	2.18		
						2.50		
Mean	1.86	2.00	2.14	2.21	2.31	2.44	...	3.28
GGA	1.87	2.1	2.4	—	—	—	...	2.95

followed by the cyclic, the cage and prism in that order. Our TB model predicts that the binding energy of the prism hexamer, relative to the isolated monomers is -0.132 Ry. On the other hand the binding energy relative to three isolated dimers is -0.087 Ry. This reflects the additional binding due first to the establishment of three more hydrogen bonds; and second, in our view having greater significance, the attraction due to the mutual polarization of the water monomers in forming the hexamer from dimers. In fact we have found, in common with previous first principles calculations,^{30,31} that the dipole moments of component monomers *increase* as they are assembled into successively larger clusters. This is shown in Table IV. We will return to this point when we discuss the structure of liquid water. We should note that unlike density functional theory, the tight binding model provides an unambiguous measure of individual molecular multipole moments; this does not however detract from the fact that these are, in principle, unmeasurable and therefore ill-defined quantities.

The relaxed structures are compared with the MP2 calculations by displaying the O-O hydrogen bond lengths in Table V. Our relaxed structures compare extremely well with the quantum chemistry results; and we remark that the good agreement and the improvement in our model over the GGA is because we have rather carefully obtained agreement over a wide range of O-O bond lengths with the CCSD(T) curve in Fig. 3. It is well known that the local density functional theory, even in gradient corrected form, has difficulty with the energetics of the hydrogen bond and so we have been able to leap-frog these problems in our TB model by comparing to the CCSD(T) not the GGA.

TABLE V. Properties of the four hexamer isomers considered in the text after geometry optimization using the fitted genetic TB model. Energy differences, ΔE , are with respect to the Prism isomer. The results labeled “MP2” are taken from Santra *et al.* (Ref. 29).

	Prism		Cage		Book		Cyclic	
	TB	MP2	TB	MP2	TB	MP2	TB	MP2
ΔE (mRy)	0	0	2.8	0.2	6.9	1.5	11.8	5.0
O–O bond lengths (bohr):								
1–2	4.975	5.280	5.055	5.205	4.970	5.094	5.009	5.134
2–3	5.639	5.522	4.914	5.047	4.955	5.093	5.009	5.134
3–1	5.597	5.545	—	—	—	—	—	—
3–4	—	—	5.208	5.265	5.045	5.298	5.009	5.134
4–5	5.382	5.240	5.003	5.225	5.061	5.249	5.008	5.134
5–6	5.564	5.424	5.542	5.520	5.299	5.261	5.009	5.134
6–4	5.627	5.556	—	—	—	—	—	—
1–4	5.533	5.512	—	—	—	—	—	—
2–5	4.926	5.028	5.570	5.573	—	—	—	—
3–6	5.013	5.185	5.566	5.494	5.520	5.544	—	—
6–1	—	—	4.998	5.149	4.972	5.081	5.009	5.134

V. THE STRUCTURE OF ICE

We now turn to the solid state to examine how our model reproduces the structure of ice. This is not the place to make an exhaustive study of the many solid state phases, nor to attempt a prediction of the phase diagram, we will leave that to future work. Most solid phases have disordered hydrogen atom sites, but the phase Ice-XI is thought to be “proton-ordered” so we test our model against this structure. Hirsch and Ojamäe³² have identified a number of putative structures of Ice Ih of which Ice-XI is found to have lowest energy. In Fig. 5 we show the structure of Ice-XI after geometry optimization using our genetic TB model. We find its cohesive energy with respect to the isolated water monomers is 39 mRy per monomer, or 12.2 kcal/mol. This result is consistent with the experimental lattice energy of 13.3 kcal/mol.³² We also find that the optimized structure of the proton-ordered Ice-Ih “phase number 6” of Hirsch and Ojamäe³² has an energy higher than Ice-XI by 0.16 kcal/mol which may be compared with the DMol3/BLYP calculation result of 0.05 kcal/mol.³² It is significant that classical force field models find, conversely, that “structure number 6” is more stable than Ice-XI. Our optimized Ice-XI has a density at 0°K of 0.97 g/cm³ compared

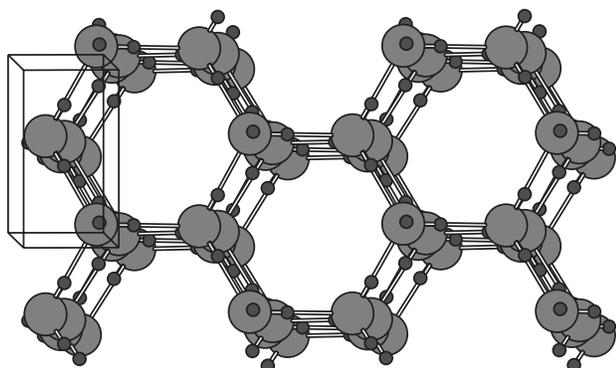


FIG. 5. The structure of ice-XI after geometry optimization within our fitted genetic TB model. A unit cell is marked by straight lines.

to the experimental value of 0.93 g/cm³ at 5°K.³³ On the other hand the density of Ice-XI as calculated in a GGA can be as large as 1.00 g/cm³ as we report in Table VI which also shows our calculated lattice constants of Ice-XI compared to GGA calculations and experiment.

VI. THE STRUCTURE AND PROPERTIES OF LIQUID WATER

Finally we apply our TB models to predict properties of liquid water. For this we take a cubic box of side 29.75 bohr containing 128 water molecules, having periodic boundary conditions and previously equilibrated in molecular dynamics (MD) using the SIESTA program and the BLYP GGA.⁴⁶ We further equilibrated for 100 ps in an NVT ensemble at 27°K and 1 g/cm³ before an additional 100 ps to gather statistics in an NVE ensemble. We have also made a simulation in an NPT ensemble at 27°K and zero external pressure for a total of 100 ps, taking statistics from the last 50 ps of simulation time. We employ the reversible integrators with Liouville

TABLE VI. Lattice constants and density of Ice-XI at 0°K calculated in our genetic and intuitive point charge TB models compared to GGA predictions and experiment. Densities of liquid water are also shown for comparison (in g/cm³), and discussed below in Sec. VI. BLYP (Refs. 35 and 36) and PBE (Ref. 24) denote specific gradient corrected density functionals. Similar data for ice Ih have been reported recently (Ref. 34).

	Ice-XI				Liquid
	<i>a</i>	<i>b</i>	<i>c</i>	ρ	ρ
TB (genetic)	8.422	14.506	13.668	0.967	0.926
TB (point)	8.488	14.754	13.900	0.928	1.030
BLYP	8.227 ^a	14.394 ^a	13.619 ^a	0.995 ^a	0.80 ^b
PBE	8.282 ^c	14.405 ^c	13.536 ^c	1.000 ^c	0.88 ^b
experiment	8.438 ^d	14.850 ^d	13.780 ^d	0.935 ^d	1.00

^aReference 32.

^bReference 37.

^cReference 38.

^dMeasurements at 5°K (Ref. 33).

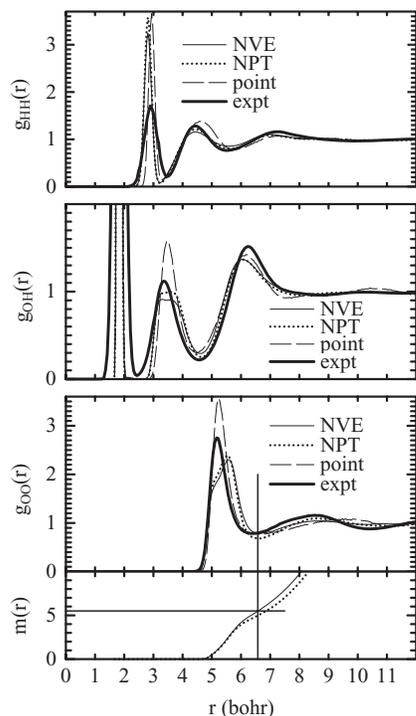


FIG. 6. Calculated radial distribution functions in liquid water, compared to experimental neutron diffraction data (Ref. 39). NVE and NPT refer to our genetic dipole model simulations in microcanonical and canonical ensembles. NVE is done at a density of 1 g/cm^3 . The average density in NPT simulation is displayed in Table VI. We also show NVE simulations in our intuitive point charge model. The lowest panel shows the integrated quantity $m(r) = \int_0^r g_{OO}(r')r'^2 dr'$ and we mark the position of the first minimum to show the average number of O-O bonds. As is generally known this is larger than the number, 4, associated with the hexagonal ices, such as Ice-XI shown in Fig. 5.

operators developed by Martyna *et al.*⁴⁷ using just a single Nosé-Hoover thermostat and particle and barostat relaxation times of 100 fs, and 1 ps, respectively. We use atomic masses of oxygen and hydrogen (not deuterium) and time steps of 1 fs in NVT and NVE and 0.25 fs in NPT ensembles.

First we show radial distribution functions (RDF) in Fig. 6. These show the same sorts of features in comparison to experiment as found in previous calculations using a GGA.^{37,45} In particular the first neighbor O-H peak is narrower than experiment, due probably to the quantum nature of protons that is usually neglected in MD simulations. The most critical RDF is that of O-O bonds since this describes the complex of hydrogen bonds responsible for many of the subtle properties of water. The RDF from our intuitive point charge model is very similar to a recent calculation⁴⁸ using the SCC-DFTB method,⁷ which as discussed in Sec. II above is equivalent in construction to our point charge model. The point charge model suffers from a nearest neighbor O-O peak which is too sharp and too high and a lack of structure in the RDF beyond first neighbors; that is to say, overstructured in the first solvation shell and understructured in more distance shells. Both these shortcomings are remedied by our genetic dipole model at the expense of the first neighbour O-O peak falling beyond the experiment indicating an average hydrogen bond length that is too long. It is notable that this average length is quite insensitive to the simulation box vol-

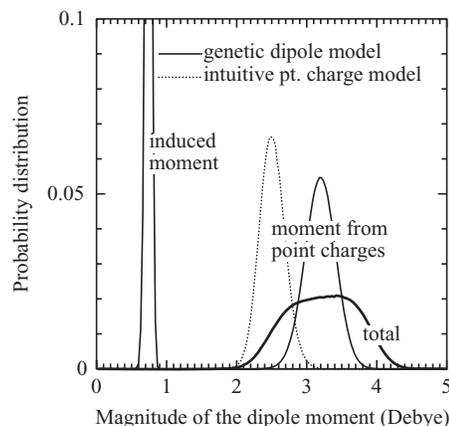


FIG. 7. Distribution of the magnitude of the dipole moment in an NVE molecular dynamics. In the case of the intuitive point charge model there is just the dipole moment arising from point charge transfer, indicated by a dotted line. In the genetic dipole model the point charge contribution is larger, consistent with the picture for the monomer, shown in Fig. 1; however, unlike in the monomer, the partial cancellation from the oppositely oriented induced dipole moment is not complete (see Fig. 8) so that the total dipole moment is on average larger than in the point charge model and is consistent with measurement and GGA calculations.

ume, since the NVE is done at a density of 1 g/cm^3 , while the average density in the NPT simulation was 0.926 g/cm^3 (see Table VI).

We use Figs. 7 and 8 to illustrate the point that we have made earlier (see Tables IV and VII) that the molecular dipole moment increases significantly as the water is built up in clusters from individual molecules having a dipole moment of 1.86 D to a total dipole moment in liquid water of about 3 D.

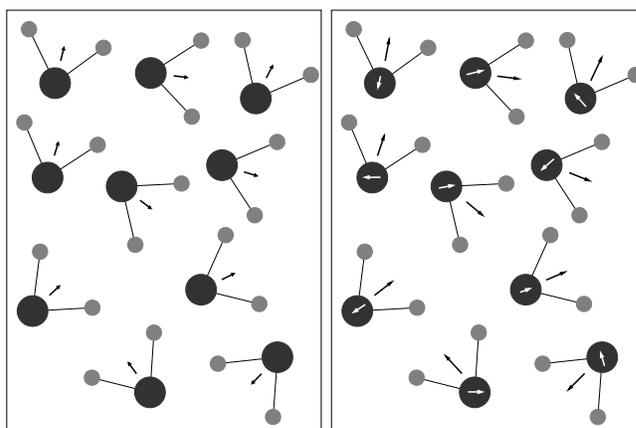


FIG. 8. A cartoon to illustrate the incomplete cancellation of the point charge transfer dipole moment in liquid water. The left hand sketch shows molecules in the point charge model whose dipoles are necessarily aligned along the centerline of the molecule and whose magnitude is close to that in the monomer. In the dipole model with induced dipoles, the right hand cartoon explains how the molecular dipole moment is increased: in the monomer (see Fig. 1) the induced dipole moment is constrained to point opposite to the point charge transfer moment, leading to a reduction in the total moment. Conversely in the liquid (and also in the solid and clusters discussed above) the induced moment “measures” the total electric field at the oxygen site, not just that due to the two closest hydrogen ions; in this way the partial cancellation of the monomer is suppressed and the resulting total moment is closer to that of the point charge transfer which is larger than the corresponding quantity in the point charge model.

TABLE VII. Some properties of liquid water calculated by our intuitive point charge and genetic dipole models, compared to experiment and published GGA calculations. \bar{p}_{tot} and \bar{p}_{ind} are total and induced average dipole moments, taken from Fig. 7, $\epsilon(0)$ is the static dielectric constant and D_{self} is the self diffusion coefficient. PBE and BLYP have the meanings of Table VI.

Model	\bar{p}_{tot} (D)	\bar{p}_{ind} (D)	$\epsilon(0)$	D_{self} $10^{-5} \text{ cm}^2/\text{s}$
Experiment	2.95 ± 0.2^a	–	$78,^b 68^c$	2.2^d
Genetic (27 °C)	3.28	0.74	86.7	3.0
Point (27 °C)	2.51	–	58.7	3.5
PBE	$2.95,^e 3.09^f$	–	$67 \pm 6,^f 75^g$	1.6^f
BLYP				$0.25,^h 0.55^i$

^a27 °C, Ref. 40.

^b27 °C, Refs. 41 and 42.

^c57 °C, Ref. 42.

^d27 °C, Ref. 43.

^e45 °C, Ref. 31.

^f57 °C, Ref. 44.

^gExtrapolated to 27 °C from data in Ref. 44.

^h35 °C, Ref. 45.

ⁱ32 °C, Ref. 46.

It is a particular feature of our dipole model that this effect is properly captured, in contrast to our point charge model in which the molecular dipole moments do not grow monotonically with cluster size (see Table IV) and do not reach the expected ~ 3 D in the bulk limit. Moreover most of the increase in dipole moment is probably a result of the polarizability of the monomer being nearly a factor of two too large in the intuitive point charge model (see Table I). The same failing, and to a much greater degree, is apparent in semiempirical schemes such as AM1 and PM3 (see Table 4 of Ref. 49). In the present context it is important to take note of a new polarizable atom NDDO scheme;⁵⁰ however we do not know whether this corrects the size effect on the monomer moments. The quantitative rendering of the dipole moments as a function of environment as evidenced in Table IV is probably the one principal advantage conveyed by using a model with polarizable ions. Classical models usually establish a molecular dipole moment of the *monomer* to about 2–3 D at the outset of the construction of the model.

Table VII lists some properties of liquid water arising from our simulations, compared to previous GGA calculations and experimental results. The static dielectric constant is calculated using the method of Sharma *et al.*⁴⁴ and the self diffusion coefficient is extracted as the linear slope of the mean square displacement. We should note that both models give a reasonable prediction of the dielectric constant, the genetic being rather better, which we attribute to the oxygen polarizability. The intuitive point charge model overestimates D_{self} which is consistent with recent SCC-DFTB simulations,⁴⁸ although these latter overestimate by a great deal more, finding $D_{\text{self}} = 11.1 \pm 0.4$ and $6.5 \pm 0.2 \times 10^{-5} \text{ cm}^2/\text{s}$ in two different modifications of the SCC-DFTB model, which are very much greater than our $3.5 \times 10^{-5} \text{ cm}^2/\text{s}$. This is very encouraging, firstly indicating that our intuitive point charge model is better in this regard than the SCC-DFTB model; more importantly by extending the theory to allow oxygen polarizability results in a significant improvement in the rendering of the self diffusion in the liq-

uid. In contrast to a GGA calculation,⁴⁴ we reproduce D_{self} quite accurately without recourse to artificial elevation of the temperature.

VII. DISCUSSION

A. General remarks on the new model

The central result of this work is that it is possible to construct TB models of water in a “ground up” approach; that is, by fitting parameters to observed and calculated properties of the monomer and dimer alone the model is found to be sufficiently robust and transferable to be able to predict many of the known properties of water. We believe that the rather transparent construction that we have outlined in Sec. II allows the models to provide some insight into the very subtle properties of water, many of these being manifestations of the hydrogen bond. In Sec. III C we made some observations on the nature of the hydrogen bond, asserting that its principal two components are an attractive force arising from the mutual static polarizability of the monomers countered by a short ranged repulsion which we model with the usual pair potential of TB theory. Of secondary importance is the additional attraction due to orbital overlap which is modeled by Hamiltonian hopping matrix elements as in all electronic structure localized orbitals methods, including the density functional description. We have not addressed the question of van der Waals forces. While we expect these to be weaker and to decay more rapidly with O–O distance than the static polarizability, there is recent evidence from DFT studies that their inclusion significantly improves upon the local density approximation.^{29,37} One could argue that by careful fitting to the CCSD(T) curve in Fig. 3 we have “renormalized” our pair potential to include some “ $1/r^6$ -like” attraction as would be done in a classical potential. Conversely, as in the DFT, we could extend our TB theory to include dynamic polarizability explicitly.¹⁴ Whether this can lead to an improved description of the hydrogen bond must be left for future work.

B. The density of water and ice

The most striking and anomalous feature of water is the fact that ice is less dense than its liquid, leading to the possibility of life on earth. Therefore we would very much like our models to display this. Unfortunately as we have seen in Table VI our genetic dipole model gives an ice density of 0.97 g/cm^3 at 0°K . We find in an NPT simulation at -3°C that this decreases somewhat to 0.96 g/cm^3 but this is still greater than the average density of the liquid in the NPT simulations at 27°C , namely 0.926 g/cm^3 . On the other hand this situation is nothing like as bad as the GGA densities reported in the literature.^{32,37,38} These find the density of Ice-XI to be as much as 1 g/cm^3 while the liquid density is as low as 0.8 g/cm^3 in the BLYP functional and 0.88 g/cm^3 in the PBE functional. This situation is rescued using a functional including dynamic van der Waals forces but at the expense of a poor rendering of the O–O RDF.³⁷ From the DFT point of view the question of the densities of ice and liquid water is clearly not yet resolved. In the TB framework it is most surprising

TABLE VIII. O–O bond lengths in the structures studied here using our two models (in atomic units of length). We have used the cyclic hexamer as an example. In the case of Ice-XI and liquid water these are averages.

Model	Dimer	Hexamer	Ice-XI	Liquid
Genetic	5.51	5.00	5.13	5.52
Point	5.34	5.01	5.22	5.26

and at the same time very encouraging that the intuitive point charge model correctly predicts that ice will float on water. This raises two questions. (i) Is the inclusion of oxygen polarization an unnecessary and uncontrollable complexity? In answer, we can certainly see from the above that the genetic dipole model performs better than the point charge model in respect of the RDF and properties of the monomer and dimer. On the other hand we have not refined the point charge model but have kept it at the “intuitive” level. (ii) Could a point charge model be constructed that will adopt the benefits of the dipole model and retain the correct difference between the liquid and ice densities? This must be a subject of further research but we suspect not. First we know from the failings of the SCC-DFTB simulations,⁴⁸ in respect of both O–O RDF and diffusivity that the TB description can fail at the point charge level, although our intuitive point charge model corrects the worse shortcomings of the SCC-DFTB. Second, and this is a point we wish to reemphasize, it is *only by including oxygen polarizability* that the trend of increasing molecular dipole moment with cluster size into the limit of the solid and liquid can be reproduced (Table IV). We believe that it is this one element that elevates our TB model into direct competition with the very much more (computer) time consuming DFT models.

We can use some of the results above to speculate upon the question, why is ice less dense than water? The conventional wisdom has it that hexagonal ice having four hydrogen bonds to each molecule is less densely packed than the liquid in which, as we see from the lowest panel in Fig. 6, each molecule has on average between five and six neighbouring molecules. However this cannot be the whole story because this is true also in our genetic model and indeed published GGA simulations (see Table VI) in which ice is (wrongly) more dense than water. It is important to note that, in contrast to the solid state, the average O–O bond length in liquid water in our models depends on the model parameters, but not strongly on the size of the confining box, that is, on the volume. This is evident from the first peaks of $g_{OO}(r)$ from the NVE and NPT simulations in Fig. 6. Some light is thrown on the matter by Table VIII which shows (average) hydrogen bond lengths from our calculations using the genetic dipole and intuitive point charge models. The point charge model gives the bond length and hence density of Ice-XI much more accurately compared to experiment than the genetic dipole model. Furthermore since the first peak of $g_{OO}(r)$ in the intuitive point charge model is close to the experiment (Fig. 6) the consequence is that this model, perhaps fortuitously, predicts the right ordering of the densities of ice and water. This may well indicate that this will be correctly rendered by any model that simultaneously predicts *both* the density of ice *and*

the position of the first peak of $g_{OO}(r)$. However it is not obvious that such a model can be easily found without specifically fitting these two properties—this is of course possible for classical models.

VIII. CONCLUSIONS

We have demonstrated models for water at “intuitive” and fitted levels in the point charge transfer and polarizable ion tight binding approximations. They have the benefit of quite clear physical and chemical insight by virtue of the way they are constructed, and at the same time possess quite remarkable predictive power. This is evident from the “ground up” approach that we have adopted in which only properties of the monomer (in respect of O–H interactions) and the dimer (in the modeling of the hydrogen bond) are fitted. Properties of clusters, ice and liquid water are then predicted with good quantitative agreement with experiment and published GGA calculations. Indeed in some notable respects our models are superior to the GGA, in particular in the density of ice-XI and the liquid (Table VI) and in the self-diffusion coefficient, which are particularly poorly described by the PBE and BLYP functionals.

We expect our model to find a niche where nanosecond simulations of some thousands of particles are called for while retaining a proper self consistent quantum mechanical description of the chemical bond. In particular, we believe that the model will provide insight into the complexities of the hydrogen bond and in the role of polar protic solvents in chemical reactions.

ACKNOWLEDGMENTS

We are grateful to C. Walsh and J. Armstrong for contributions to the early stages, and A. Y. Lozovoi to the later stages of this work. We have enjoyed many stimulating discussions with R. M. Lynden-Bell. We thank M.-V. Fernández-Serra for providing an equilibrated sample of 128 water molecules; and B. Santra for the hexamer structure parameters. Financial support was provided by the UK EPSRC under grant number EP/G012156/1.

APPENDIX: GENERALIZED MADELUNG MATRIX

The generalized Madelung matrix \tilde{B} is proportional to the KKR structure constant matrix, B ,^{51,52}

$$\tilde{B}_{\mathbf{R}'L'\mathbf{R}L} = \frac{4\pi}{(2\ell + 1)!!(2\ell' + 1)!!} B_{\mathbf{R}'L'\mathbf{R}L},$$

where

$$B_{\mathbf{R}'L'\mathbf{R}L} = 4\pi \sum_{L''} (-1)^\ell (2\ell'' - 1)!! C_{L'L''} K_{L''}(\mathbf{R} - \mathbf{R}')$$

and

$$K_L(\mathbf{r}) = r^{-\ell-1} Y_L(\mathbf{r})$$

is the solid Hankel function.

$$C_{L''L'L} = \iint d\Omega Y_{L''} Y_{L'} Y_L \quad (\text{A1})$$

are the Gaunt integrals for real spherical harmonics.¹⁴

- ¹M. W. Finnis, A. T. Paxton, M. Methfessel, and M. van Schilfgaarde, *Phys. Rev. Lett.* **81**, 5149 (1998).
- ²A. T. Paxton, T. N. Todorov, and A. M. Elena, *Chem. Phys. Lett.* **483**, 154 (2009).
- ³W. A. Harrison, *Electronic Structure and the Properties of Solids* (W. H. Freeman, San Francisco, 1980).
- ⁴S. Fabris, A. T. Paxton, and M. W. Finnis, *Phys. Rev. B* **63**, 094101 (2001).
- ⁵M. W. Finnis, *Interatomic Forces in Condensed Matter* (Oxford University, Oxford, 2003).
- ⁶A. T. Paxton, in "Multiscale Simulation Methods in Molecular Sciences," (NIC series, Vol. 42, Jülich Supercomputing Centre) pp. 145–174. Available on-line at <http://www.fz-juelich.de/nic-series/volume42>.
- ⁷M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai, and G. Seifert, *Phys. Rev. B* **58**, 7260 (1998).
- ⁸L. Goodwin, A. J. Skinner, and D. G. Pettifor, *Europhys. Lett.* **9**, 701 (1989).
- ⁹D. G. Pettifor, *Bonding and Structure of Molecules and Solids* (Oxford University, Oxford, 1995).
- ¹⁰W. A. Harrison, *Phys. Rev. B* **31**, 2121 (1985).
- ¹¹A. P. Sutton, M. W. Finnis, D. G. Pettifor, and Y. Ohta, *J. Phys.: Condens. Matter* **21**, 35 (1988).
- ¹²D. S. McClure, "Phonons in perfect lattices and in lattices with point imperfections" (Oliver and Boyd, Edinburgh, 1966), Chap. 13.
- ¹³A. M. Stoneham, *Theory of Defects in Solids* (Clarendon, Oxford, 1975).
- ¹⁴A. J. Stone, *The Theory of Intermolecular Forces* (Clarendon, Oxford, 1996).
- ¹⁵D. R. Lide, ed., *CRC Handbook of Chemistry and Physics—83rd edition* (Chemical Rubber Publishing Company, Boca Raton, Florida, 2002).
- ¹⁶Formulas relating force constant and frequency can be found in the book by Herzberg (Ref. 17), Sect. II, 4: Eqs. (II, 244).
- ¹⁷G. Herzberg, *Infrared and Raman Spectra of Polyatomic Molecules* (D. Van Nostrand Company, Inc., New York, 1945).
- ¹⁸J. C. Slater and G. F. Koster, *Phys. Rev.* **94**, 1498 (1954).
- ¹⁹S. Fabris, A. T. Paxton, and M. W. Finnis, *Phys. Rev. B* **61**, 6617 (2000).
- ²⁰M. W. Schmidt, K. K. Baldrige, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, and M. Dupuis, and J. A. Montgomery, Jr., *J. Comput. Chem.* **14**, 1347 (1993).
- ²¹M. S. Gordon and M. W. Schmidt, "Theory and applications of computational chemistry, the first forty years" (Elsevier, Amsterdam, 2005), Chap. 41.
- ²²The CCSD(T) calculations were done with the basis set aug-cc-vPTZ, that is, augmented, correlation-consistent, valence polarization triple zeta.
- ²³W. Klopper, J. G. C. M. van Duijneveldt-van de Rijdt, and F. B. van Duijneveldt, *Phys. Chem. Chem. Phys.* **2**, 2227 (2000).
- ²⁴J. P. Perdew, K. Burke, and M. Ernzerhof, *Phys. Rev. Lett.* **77**, 3865 (1996); *Phys. Rev. Lett.* **78**, 1396 (1997).
- ²⁵H.-P. Schwefel, *Numerische Optimierung von Computer-Modellen mittels der Evolutionsstrategie, Interdisciplinary Systems Research*, Vol. 26 (Birkhäuser, Basle, 1977).
- ²⁶H.-P. Schwefel, *Evolution and Optimum Seeking: The Sixth Generation* (John Wiley, New York, 1993).
- ²⁷D. J. Chadi, *Phys. Rev. Lett.* **41**, 1062 (1978).
- ²⁸S. Znam, D. Nguyen-Manh, D. G. Pettifor, and V. Vitek, *Philosophical Magazine* **83**, 415 (2003).
- ²⁹B. Santra, A. Michaelides, M. Fuchs, A. Tkatchenko, C. Filippi, and M. Scheffler, *J. Chem. Phys.* **129**, 194111 (2008).
- ³⁰L. Delle Site, A. Alavi, and R. M. Lynden-Bell, *Mol. Phys.* **96**, 1683 (1999).
- ³¹P. L. Silvestrelli and M. Parrinello, *Phys. Rev. Lett.*, **82**, 3308 (1999).
- ³²T. K. Hirsch and L. Ojamäe, *J. Phys. Chem.* **108**, 15856 (2004).
- ³³C. M. B. Line and R. W. Whitworth, *J. Chem. Phys.* **104**, 10008 (1996).
- ³⁴P. J. Feibelman, *Phys. Chem. Chem. Phys.* **10**, 4688 (2008).
- ³⁵A. D. Becke, *Phys. Rev. A* **38**, 3098 (1988).
- ³⁶C. Lee, W. Yang, and R. G. Parr, *Phys. Rev. B* **37**, 785 (1988).
- ³⁷J. Wang, G. Roman-Perez, J. M. Soler, E. Artacho, and M.-V. Fernández-Serra, *J. Chem. Phys.* **134**, 024516 (2011).
- ³⁸K. Umemoto, R. M. Wentzcovitch, and S. Baroni S. de Gironcoli, *Phys. Rev. Lett.* **92**, 105502 (2004).
- ³⁹A. K. Soper, *Chem. Phys.* **258**, 121 (2000).
- ⁴⁰A. V. Gubskaya and P. G. Kusalik, *J. Chem. Phys.* **117**, 5920 (2002).
- ⁴¹J. N. Murrell and A. D. Jenkins, *Properties of Liquids and Solutions* (John Wiley and Sons, Chichester, England, 1994).
- ⁴²D. P. Fernández, Y. Mulev, A. R. H. Goodwin, and J. M. H. L. Sengers, *J. Phys. Chem. Ref. Data* **24**, 33 (1995).
- ⁴³D. Eisenberg and W. Kauzmann, *The Structure and Properties of Water* (Oxford University, London, 1969).
- ⁴⁴M. Sharma, R. Resta, and R. Car, *Phys. Rev. Lett.* **98**, 247401 (2007).
- ⁴⁵H.-S. Lee and M. E. Tuckerman, *J. Chem. Phys.* **126**, 164501 (2007).
- ⁴⁶M.-V. Fernández-Serra and E. Artacho, *J. Chem. Phys.* **121**, 11136 (2004).
- ⁴⁷G. J. Martyna, M. E. Tuckerman, D. J. Tobias, and M. L. Klein, *Mol. Phys.* **87**, 1117 (1996).
- ⁴⁸C. M. Maupin, B. Aradi, and G. A. Voth, *J. Phys. Chem. A* **114**, 6922 (2010).
- ⁴⁹G. Monard, M. I. Bernal-Uruchurtu, A. van der Vaart, K. M. Merz, and M. F. Ruiz-López, *J. Phys. Chem. A* **109**, 3425 (2005).
- ⁵⁰D. T. Chang, G. K. Schenter, and B. C. Garrett, *J. Chem. Phys.* **128**, 164111 (2008).
- ⁵¹A. R. Williams, J. Kübler, and C. D. Gelatt, *Phys. Rev. B* **19**, 6094 (1979).
- ⁵²O. K. Andersen, *The Electronic Structure of Complex Systems*, edited by P. Phariseau and W. M. Temmerman (Plenum, New York, 1984), p. 11.