Natural polarizability and flexibility via explicit valency: The case of water

Seyit Kale[1] and Judith Herzfeld[2,a)]
[1]*Graduate Program in Biophysics and Structural Biology, Brandeis University, Waltham, Massachusetts 02454-9110, USA*
[2]*Department of Chemistry, Brandeis University, Waltham, Massachusetts 02454-9110, USA*

(Received 18 December 2011; accepted 6 February 2012; published online 29 February 2012)

As the dominant physiological solvent, water drives the folding of biological macromolecules, influences conformational changes, determines the ionization states of surface groups, actively participates in catalytic events, and provides "wires" for long-range proton transfer. Elucidation of all these roles calls for atomistic simulations. However, currently available methods do not lend themselves to efficient simulation of proton transfer events, or even polarizability and flexibility. Here, we report that an explicit account of valency can provide a unified description for the polarizability, flexibility, and dissociability of water in one intuitive and efficient setting. We call this approach LEWIS, after the chemical theory that inspires the use of valence electron pairs. In this paper, we provide details of the method, the choice of the training set, and predictions for the neat ambient liquid, with emphasis on structure, dynamics, and polarization. LEWIS water provides a good description of bulk properties, and dipolar and quadrupolar responses. © *2012 American Institute of Physics*. [http://dx.doi.org/10.1063/1.3688228]

I. INTRODUCTION

Where computer simulations are concerned, it is important that the model fit the application with respect to both the level of detail and the computational burden. For water, the range of models is broad, from the continuum dielectric[1] to high-level quantum mechanics.[2] Within this range, empirical atomistic representations occupy an important niche with respect to length scale and efficiency.[3,4] The most basic, "rigid water" construct, with fixed OH bond length, HOH angle, and distribution of partial charges, dates to the pre-computer era[5] and still forms the basis of modern mainstream models, such as the TIPnP's,[3,6–8] or the SPC's,[9,10] while flexible variants,[11–13] with variable bond lengths and bond angles, provide improved descriptions of bulk liquid properties.

Recently, more ambitious water models have addressed the polarizability of water.[14–18] Water undergoes substantial internal charge rearrangements in response to neighboring molecules. For example, the molecular dipole moment increases by as much as ∼60% from the gas phase to the liquid.[19] These changes are usually modeled by adding oscillating, distributed, or variable charges at some extra computational expense. The last decade has seen promising validation and growing acceptance of polarizable models.[20]

However, polarizable models remain incapable of breaking bonds. For simulations of reactions, the gold standard is quantum mechanics (QM) where the solution to Schrödinger's equation is numerically approximated for a many-body problem of nuclei and electrons. In principle, *ab initio* QM methods, can represent both polarization and reactions in arbitrary settings. Liquid water, in neutral,[2,21,22] protonated[23] and deprotonated states,[24–26] and ice[27] have been rigorously studied by QM methods, and one study on autoionization[28] has managed to bypass computational time limitations by using transition state theory. However, the QM computational burden limits applications to relatively small systems, even with frugal basis sets and the use of density functional theory (DFT).

Empirical approaches, which can simulate far larger systems for far longer times, have generally restricted water reactivity to "protonation." Several hydronium or excess proton models have been reported.[29] In addition, flexible[30] or even polarizable[31] multi-state empirical valence bond methods can be applied accurately to excess protons in the bulk,[31,32] at interfaces,[33] and in biologically relevant media.[34] Empirical valence bond (EVB) methods are based on Warshel's general theory of valence states[35] where an arbitrary system configuration is expressed as a sum of multiple valence states weighted by energy optimized coefficients. While EVB is applied in the region of interest, other water molecules are typically described by a simple model. EVB states have also been developed for deprotonated systems[37] by using a charged ring model for hydroxide.[36] However, the excess proton and proton vacancy constructs have yet to be integrated.

Recently, we have unified polarization, flexibility, protonation, and deprotonation in one efficient and self-consistent construct by taking explicit account of valency.[38] Our approach is related to that of Stillinger's central force model[39] and polarization model[90] in the sense that carefully designed pairwise potentials allow full dissociability. However, we obviate extra polarization constructs by taking explicit account of valency. Specifically, we unpack each water molecule into a $+6e$ charged core (i.e., the combination of a $+8e$ oxygen nucleus and a $-2e$ pair of $1s$ electrons) surrounded by four $-2e$ charged valence electron pairs and two protons bearing $+1e$ charges.[38] These LEWIS particles are all independently mobile and represented by point locations that give the

a)Author to whom correspondence should be addressed. Electronic mail: herzfeld@brandeis.edu. Tel: 781-736-2538.

electron pairs a pseudo-classical nature. For efficiency, interactions are pairwise only. Quantum effects (i.e., electron delocalization and Pauli exclusion and dispersion) are incorporated implicitly in the deviations of the interactions from Coulombic at relevant distances. The functional forms and parameterization are optimized against a training set that includes: the structures of neutral, protonated and deprotonated monomers, and dimers; monomer proton affinities; the hydrogen bond energy; the rotation[7] and stretching[40] landscape of the neutral dimer; the proton hopping barrier in the deprotonated dimer;[41] and the ambient density of bulk water. Reference data are taken from the literature, with priority given to experimental results. High-level quantum mechanical calculations are considered only when experimental values are not available.

In this paper, we present the analytical forms of the LEWIS potentials, describe the molecular mechanics setup, and report ambient liquid state results with emphasis on liquid structure, dynamics, and polarizability. Because polarization in LEWIS is an incidental benefit of freely mobile protons and valence electron pairs, it does not play a direct role in training the potentials. The response occurs simply via subtle intramolecular rearrangements that involve stretched covalent bonds and an expanded bond angle. Reorganization of electron pairs also contributes, resulting in accurate dipolar and quadrupolar changes. In this respect, LEWIS addresses a longstanding problem of simple flexible and polarizable empirical water models that have difficulties in reproducing the experimentally observed[42] and quantum mechanically supported[43] changes of the water geometry in the liquid state.[11,14,44,45]

II. LEWIS MODEL

A. Potentials

Upon testing various combinations of different functional forms we arrived at the following set of six pairwise potentials for interactions between our three types of particles, valence pairs (V), hydrogen nuclei (H), and oxygen cores (O):

$$U_{\mathrm{VV}} = \frac{q_{\mathrm{V}}^2}{\left(r^8 + \rho_{\mathrm{VV1}}^4 r^4 + \rho_{\mathrm{VV2}}^8\right)^{1/8}} + \frac{\kappa_{\mathrm{VV}}}{1 + \left(\frac{r}{\rho_{\mathrm{VV3}}}\right)^6}, \tag{1}$$

$$U_{\mathrm{VO}} = \frac{q_{\mathrm{V}} q_{\mathrm{O}}}{\left(r^8 + \rho_{\mathrm{VO1}}^4 r^4 + \rho_{\mathrm{VO2}}^6 r^2 + \rho_{\mathrm{VO3}}^8\right)^{1/8}} + \left(\frac{\rho_{\mathrm{VO4}}}{r}\right)^6, \tag{2}$$

$$U_{\mathrm{VH}} = \frac{q_{\mathrm{V}} q_{\mathrm{H}}}{\left(r^8 + \rho_{\mathrm{VH1}}^2 r^6 + \rho_{\mathrm{VH2}}^4 r^4 + \rho_{\mathrm{VH3}}^6 r^2 + \rho_{\mathrm{VH4}}^8\right)^{1/8}}, \tag{3}$$

$$U_{\mathrm{OH}} = \frac{q_{\mathrm{O}} q_{\mathrm{H}}}{r} - \frac{\kappa_{\mathrm{OH}}}{1 + \tau_{\mathrm{OH}}\left(\frac{r}{\rho_{\mathrm{OH}}}\right)^3 + \left(\frac{r}{\rho_{\mathrm{OH}}}\right)^6}, \tag{4}$$

$$U_{\mathrm{OO}} = \frac{q_{\mathrm{O}}^2}{r} - \frac{\kappa_{\mathrm{OO}}}{1 + \tau_{\mathrm{OO}}\left(\frac{r}{\rho_{\mathrm{OO}}}\right)^3 + \left(\frac{r}{\rho_{\mathrm{OO}}}\right)^6}, \tag{5}$$

$$U_{\mathrm{HH}} = \frac{q_{\mathrm{H}}^2}{r}, \tag{6}$$

where r represents the interparticle distance, q represents the particle charge (+6 for O, +1 for H, and −2 for V), and other symbols represent parameters adjusted to provide the best agreement with the physical data in the training set. Note that the leading electrostatic term is softened in the three cases involving valence pairs, and not otherwise. This softening reflects the diffuse and responsive nature of the electron pairs. Additional terms reflect Pauli exclusion (between valence pairs in the VV interaction and between valence pairs and the 1*s* electrons in the VO interaction) and dispersive attractions (in the OO and OH interactions). Dispersion is finite in the short range and follows an inverse sixth power in the long range.[46]

The choice of the potential forms is based on a rigorous evaluation of over ∼300 different combinations. Both the mathematical forms and the training set were allowed to coevolve in light of the results of typically several hundred optimization trials for each.

1. Coulomb softening is found to be well represented by triple-square roots with multiple parameters. The use of a four-parameter variant in the VH potential was crucial to address the rotational barrier of the water dimer. Less-flexible forms typically overestimate this barrier, yielding a "frozen" liquid due to virtually permanent hydrogen bonds.
2. The explicit dispersion terms were needed primarily to fit the thermodynamic properties and the geometries of the ionic dimers. The shifted polynomial form, with a finite value at zero distance, allowed for large attractions at bonding distances and inverse sixth power decay at long distances.
3. Exclusion terms are important among explicit valence electron pairs and between valence electron pairs and implicit core electrons. Core-core exclusions, analogous to, e.g., the inverse 12th power polynomial of Lennard-Jonnes, were not helpful.

B. Potential optimization

The best parameters are obtained using a double-layer optimization program. The outer layer explores *potential* parameters, seeking those that minimize weighted deviations from structural and thermodynamic data in the training set. These deviations are calculated by the inner layer after exploring the *structural* parameters of species in the training set to find those that minimize the energies given the current set of *potential* parameters. Structure optimizations are expedited by making use of all applicable molecular symmetries (including symmetries of electron pairs). Minimum energy configurations are then validated in Monte Carlo[47] simulations without the symmetry constraints.

The weights given to elements of the training set in the outer optimization are varied from one run to another in order to facilitate a full exploration of the parameter space. Taking the weight of the H_2 molecular bond length

TABLE I. Model parameters used in Eqs. (1)–(6). Distances (ρ's) are in Ångstroms, energies (κ's) are in atomic units, and τ's are unitless.

Parameter	Value	Parameter	Value
ρ_{VV1}	0.12820	ρ_{VV2}	0.52780
ρ_{VV3}	0.72342	κ_{VV}	1.95019
ρ_{VH1}	0.09122	ρ_{VH2}	1.07407
ρ_{VH3}	0.56994	ρ_{VH4}	0.59761
ρ_{VO1}	0.59188	ρ_{VO2}	0.42553
ρ_{VO3}	0.47725	ρ_{VO4}	0.07272
ρ_{OH}	1.08820	ρ_{OO}	2.94886
κ_{OH}	1.98485	κ_{OO}	0.22145
τ_{OH}	1.54600	τ_{OO}	8.05300

TABLE II. LEWIS fit to monomer and dimer features. r_{OH} and r_{HH} denote covalent bond lengths, θ_{HOH} the HOH bond angle, and PA the proton affinity. All reference values are experimental, except for the O-O distance of the deprotonated dimer, which is taken from a quantum mechanical calculation.

Molecule	Feature	Unit	Model	Reference
H_2O	r_{OH}	Å	0.94	0.958[a]
	θ_{HOH}	deg	104.7	104.4[a]
	PA	kJ/mol	617.8	691.0[b]
H_3O^+	r_{OH}	Å	0.96	0.976[c]
	θ_{HOH}	deg	109.2	111.3[c]
OH^-	r_{OH}	Å	1.01	0.964[d]
	PA	kJ/mol	1671.2	1635.1[e]
H_2	r_{HH}	Å	0.66	0.741[f]
H^-	PA	kJ/mol	1667.3	1675.3[g]
$(H_2O)_2$	r_{OO}	Å	2.85	2.976[h]
	θ_{donor}	deg	−53.4	−51.0[h]
	$\theta_{acceptor}$	deg	41.3	57.0[h]
$H_5O_2^+$	r_{OO}	Å	2.38	2.40[i]
$H_3O_2^-$	r_{OO}	Å	2.38	2.47[j]

[a]Reference 80.
[b]Reference 81.
[c]Reference 82.
[d]Reference 83.
[e]Reference 84.
[f]Reference 85.
[g]Reference 86.
[h]Reference 87.
[i]Reference 88.
[j]Reference 41.

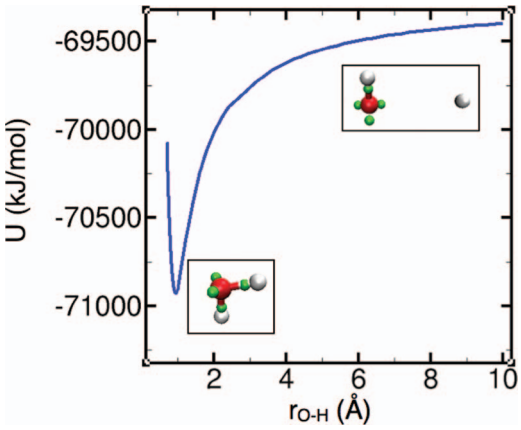

FIG. 1. Monte Carlo results for the potential energy of a water molecule as one of its protons is moved toward and away from the oxygen. In the insets, oxygen is rendered in red, protons in white, and electron pairs in green. The depth of the minimum relative to the dissociation limit ($r_{O\text{-}H} \to \infty$) corresponds to the proton affinity of hydroxide.

as 1, other structural features typically have weights of 4-12 for H_2O and H_3O^+, 2–5 for OH^-, and 6–10 for the three water dimers. Proton affinities have relative weights of 2-3 for H_2O and OH^-, and 0.05–0.1 for H^-. The hydrogen bond enthalpy in the neutral dimer generally carries a relative weight of less than 0.5 (see discussion of trade-offs below). Although dimer dissociation and rotation landscapes are sampled at 8–10 equally spaced positions along the reaction coordinate, most weight eventually focuses on a few crucial positions around a prominent energy barrier or wall, with weights comparable to those used for the energy minimum. Short (200 ps – 1ns) NPT simulations at 300 K and 1 atm are run on promising parameter sets to determine liquid water density.

The parameter values for the potentials are listed in Table I, the degree to which the interactions accommodate the features of monomers and dimers is shown Table II, the monomer dissociation curve is shown in Fig. 1, and results for small clusters are shown in Table III. The fits in Table II and the predictions in Fig. 1 and Table III show that strictly pairwise potentials are capable of approximating a wide range of phenomena.

TABLE III. LEWIS predictions for small neutral water clusters and comparison with QM results. (Refs. 75 and 89) Here n denotes the number of monomers and n_{hb} the number of hydrogen bonds. The formation enthalpy is given by $\Delta E = (E_{cluster} - n\, E_{H2O})$, and the energy per hydrogen bond is given by $\Delta E_{hb} = (\Delta E / n_{hb})$. For O-O distances ($r_{OO}$), ranges are given. QM distances are taken from the structures provided by one of the authors (S.X.) of the QM references. The twisted prism hexamer of LEWIS is compared with the straight prism hexamer predicted by QM. The latter has nine hydrogen bonds.

Cluster	n	n_{hb}	$-\Delta E$ (kJ/mol)	$-\Delta E_{hb}$ (kJ/mol)	r_{OO} (Å)	$r_{OO,QM}$ (Å)	$-\Delta E_{QM}$ (kJ/mol)
$(H_2O)_2$	2	1	25.27	25.27	2.85	2.91	20.85
$(H_2O)_3$	3	3	72.71	24.24	2.74–2.76	2.79	66.28
$(H_2O)_4$	4	4	137.2	34.29	2.69–2.70	2.73	115.68
$(H_2O)_5$	5	5	188.7	37.73	2.63–2.67	2.71–2.72	151.90
$(H_2O)_{6,cycle}$	6	6	235.4	39.23	2.63–2.64	2.71	187.61
$(H_2O)_{6,prism}$	6	7	234.6	33.52	2.62–2.84	2.65–2.95	192.01
$(H_2O)_{8,\ D2d}$	8	12	361.3	30.11	2.70–2.87	2.67–2.84	305.39
$(H_2O)_{8,S4}$	8	12	361.3	30.10	2.69–2.91	2.66–2.84	304.72

C. Molecular dynamics

Newton's equations of motion are solved using a time step of 0.2 fs and electron pair mass of 1 amu. Nuclei are given the masses of their most common isotopes (m_O = 16 amu and m_H = 1 amu). Temperature is maintained at 300 K using stochastic velocity rescaling[48] with a time constant of 0.1 ps. Electrons and nuclei are subject to separate thermostats to avoid spontaneous energy transfers between the two groups. Pressures are maintained at 1 atm using an isotropic Parrinello-Rahman barostat[49] that does not distinguish between intra- and intermolecular degrees of freedom. The coupling constant is 20 ps. Stronger coupling can lead to unphysical acceleration of electron pairs. Residual center of mass motion is subtracted every 100 steps. Simulations are run using GROMACS, version 4.5.3,[50] where LEWIS interactions are introduced as tabulated potentials. NPT and NVT ensembles are run in single precision leap-frog algorithm[51] and NVE in double precision Velocity-Verlet.[52] The neighbor list radius is 11 Å and the lists are updated every 25 steps (5 fs). Trajectories were saved every 100 fs.

An 11 ns simulation was run with 500 water monomers in a cubic box with periodic boundaries. The run began with conjugate gradient energy minimization of randomly distributed monomers following a 1 ns NPT simulation and a 10 ns NVT production run. A separate 2 ns NVE run was performed to benchmark energy conservation. Two separate 5 ns NPT simulations of the 500 H_2Os box (at 1 atm, and 300 K or 295 K) were run to estimate the heat capacity at constant pressure.

Analyses use GROMACS subroutines as well as homemade scripts. Diffusion constants are calculated from the slopes of mean square displacements using the Einstein relation.[53] Dielectric properties are calculated using the conventional dipole fluctuation formula[54] where monomers are wrapped at the periodic boundaries. The heat capacity per monomer (at 297.5 K) was obtained by dividing the total energy difference ($H = H_{300\,K} - H_{295\,K}$) by the temperature difference (T = 5 K).

D. Long-range electrostatics

We compensate for long-range interactions by using a novel version of force shifting, also known as the shifted force gradient, as described earlier.[55] This pairwise scheme modifies original potentials such that their first two derivatives go smoothly to zero at the cutoff. For example, for an original interaction potential U in Eq. (1)–(6), the long-range compensated potential U_c and its associated compensated force F_c are given by the relations,

$$U_c(r_{ij}) = \begin{cases} U(r_{ij}) - U(r_c) - (r_{ij} - r_c)\dfrac{dU(r_c)}{dr} - \dfrac{1}{2}(r_{ij} - r_c)^2\dfrac{d^2U(r_c)}{dr^2} & r_{ij} \le r_c, \\ 0 & r_{ij} > r_c, \end{cases} \tag{7}$$

$$F_c(r_{ij}) = \begin{cases} -\dfrac{dU(r_{ij})}{dr} + \dfrac{dU(r_c)}{dr} + (r_{ij} - r_c)\dfrac{d^2U(r_c)}{dr^2} & r_{ij} \le r_c, \\ 0 & r_{ij} > r_c, \end{cases} \tag{8}$$

where r_{ij} denotes the distance between particles i and j, and r_c denotes the cutoff distance. LEWIS is parameterized for r_c = 9 Å, which outperforms smooth particle mesh Ewald by ∼ 8%-9% in CPU time.[55] Note that $F_c = -\nabla U_c$ and compensated forces are energy-conserving (also see Sec. III).

III. RESULTS

A. Structure and dynamics of ambient liquid water

The LEWIS description of the ambient neat liquid is similar to that obtained by Car-Parrinello using the HCTH functional, which was reported to perform well in predicting liquid water properties.[56,57] The radial distribution function, g_{OO} (Fig. 2), correlates well with the most recent model-independent diffraction data[58] in terms of first peak position and height, while other features are somewhat shifted and exaggerated as in the Car-Parrinello structure. The peaks of g_{OH} and g_{HH} behave similarly (Fig. 2). Consistent with this excess order, diffusion is somewhat slow, with $D_{H2O} = 0.11$ Å^2 ps^{-1} compared to the experimental $D_{H2O} = 0.23$ Å^2 ps^{-1}.[59] Slow

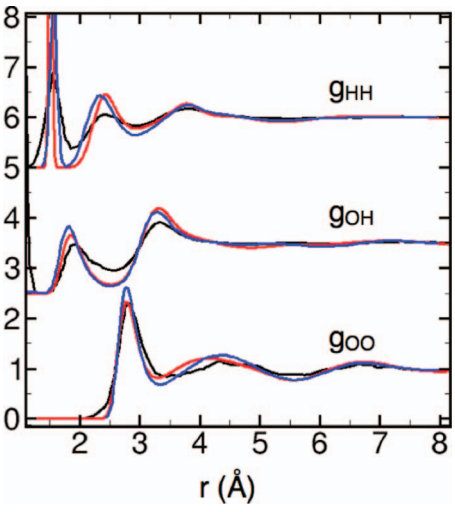

FIG. 2. Radial distribution functions of ambient water at 300 K: LEWIS (red), Car-Parrinello with the HCTH functional (blue), (Ref. 56) and experimental (black) (Ref. 58).

diffusion is common in polarizable models.[18] Coordination numbers at $R_c = 3.3$ Å (n_c) and $R_{c2} = 5.5$ Å (n_{c2}) are 3.89 and 22.22, respectively (vs. experimental $n_c = 4.26$ and $n_{c2} = 22.39$ (Refs. 11 and 60). At 300 K and 1 atm, the density (ρ) is ~0.995 g/cm^3 and the heat of vaporization (ΔH_{vap}) is 33.24 kJ/mol (vs. experimental $\rho = 0.997$ g/cm^3 (Ref. 61) and $\Delta H_{vap} = 44.02$ kJ/mol (Ref. 62) at T = 298.15 K). Similar deviations in ΔH_{vap} typically occur in quantum mechanical calculations[63] and in empirical models that are not optimized for this feature.[4] The estimated heat capacity at constant pressure (see *Methods*) is more than twice the experimental value, as expected given the additional classical degrees of freedom[12,15,64] inherent in the LEWIS electron pairs.

B. Polarization and dielectric response

LEWIS predicts a distribution of monomer dipole moments in the neat liquid with an average of $\langle\mu\rangle$ ~2.64 D and a width at half maximum of ~0.54 D (Fig. 3, right panel), consistent with the range of currently accepted experimental estimates[65] and with DFT predictions of 2.66 D.[22] The LEWIS gas phase dipole moment of $\mu = 2.01$ D is higher than the experimental 1.85 D,[66] but less so than in rigid, non-polarizable models such as TIP3P (2.35 D) (Ref. 7) or SPC (2.27 D) (Ref. 9), where the dipoles are optimized between the gas and liquid state values.

The traceless quadrupole moment tensor Q is calculated according to

$$Q_{\alpha\beta} = \frac{1}{2}\sum_i q_i \left(3r_{i\alpha}r_{i\beta} - r_i^2\delta_{\alpha\beta}\right), \tag{9}$$

where $\boldsymbol{r}_i$ is the position of particle i relative to the center of mass, α and β denote x, y, or z coordinates, δ is the Kronecker delta function, and the sum over i goes through all particles of the molecule (see, e.g., Wu *et al.*[11]). Following convention we set the molecule in the yz-plane with the sum of the two OH bond vectors along the z-axis. In the gas-phase LEWIS predicts $Q_{xx} = -2.55$, $Q_{yy} = 2.58$, and $Q_{zz} = -0.033$ DÅ which agree well with the experimental values, −2.50, 2.63, and −0.13 DÅ, respectively.[67] In the liquid state, MP2,[65] DFT,[68] and coupled-cluster calculations[69] all predict a 6%–13% increase in the breadth of the tensor. A similar effect is also observed in LEWIS which predicts mean diagonal quadrupole moments in the liquid state of $Q_{xx} = -2.86$, $Q_{yy} = 2.92$, and

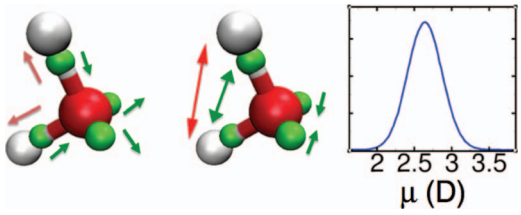

FIG. 3. Flexible polarization response of LEWIS water in liquid at 300 K. White, red, and green spheres represent hydrogen, oxygen, and valence pair particles, respectively. Arrows in the left panel indicate direction of changes in distances from the oxygen, and arrows in the middle panel identify the changes in angles. Red arrows refer to protons, and green arrows to electron pairs. The right panel shows the distribution of dipole moments in the liquid.

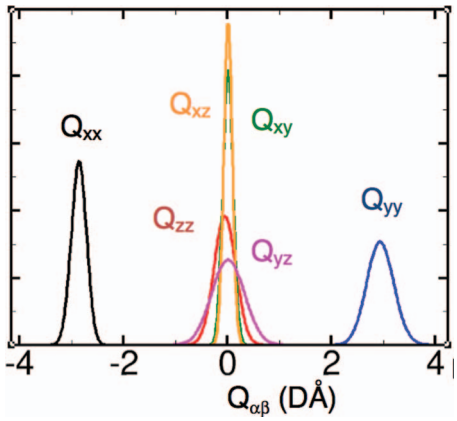

FIG. 4. Distribution of all six elements of the quadrupole tensor at 300 K. Note that normalization imposes an inverse relation between peak heights and widths. For the principal components, e.g., this implies that $\sigma_{Qxx} < \sigma_{Qzz} < \sigma_{Qyy}$. The greatest variance is observed in the molecular (yz) plane.

$Q_{zz} = -0.060$ DÅ and off-diagonal elements close to zero. It is also noteworthy that the six quadrupole elements exhibit different variances with the greatest being in the molecular plane (Fig. 4). To the best of our knowledge, no experimental data are yet available for comparison.

In the liquid state, the distribution of bond angles and bond lengths is broad (Fig. 5), with average values expanded by ~0.4° and ~0.01 Å relative to the monomer. This trend is consistent with experimental results,[42] but is not captured by most flexible models.[14,15] Geometry-dependent electronic polarization has been suggested as a requirement for capturing this effect correctly.[11,43,44] Consistent with Wannier function analyses,[19] the LEWIS electrons assume a more spherical distribution in the liquid state, in that bonding pairs are drawn somewhat closer the oxygen nucleus while lone pairs drift somewhat further away, and the angular arrangement shifts towards tetrahedrality (Fig. 3). While the average physical displacements are small in magnitude, the impact on the overall polarization is substantial due to the full ionic charges of the particles. The dielectric constant ε_0 at 300 K converges to ~71.9 (Fig. 6), and finite and infinite Kirkwood factors to $G_k \sim 2.99$ and $g_k \sim 2.01$, respectively (vs. experimental $\varepsilon_0 = 78.49$ (Ref. 70 and $g_k = 2.90$ [71] at T = 298.15 K).

C. Conservation of total energy

One concern about reactive methods is energy conservation.[30,31,72] While the precision of a non-reactive force field is typically limited by the integrator, most reactive simulations involve calculations that may affect "completeness" such as basis states expansions[31] or massive multi-body interactions.[73] LEWIS with long-range compensation (see *Methods*) enjoys the simplicity of smooth pairwise potentials, which yield good energy conservation, even in tabulated forms. Total energy drift in a system of 500 ambient water molecules is within half a kJ/mol ns^{-1} and may be reduced further by the use of analytical forms. As noted in *Methods*, we monitor the temperatures of individual particle groups, i.e., T_O, T_H, and T_V, to detect spontaneous energy transfers that may slow dynamics and affect observables. In our

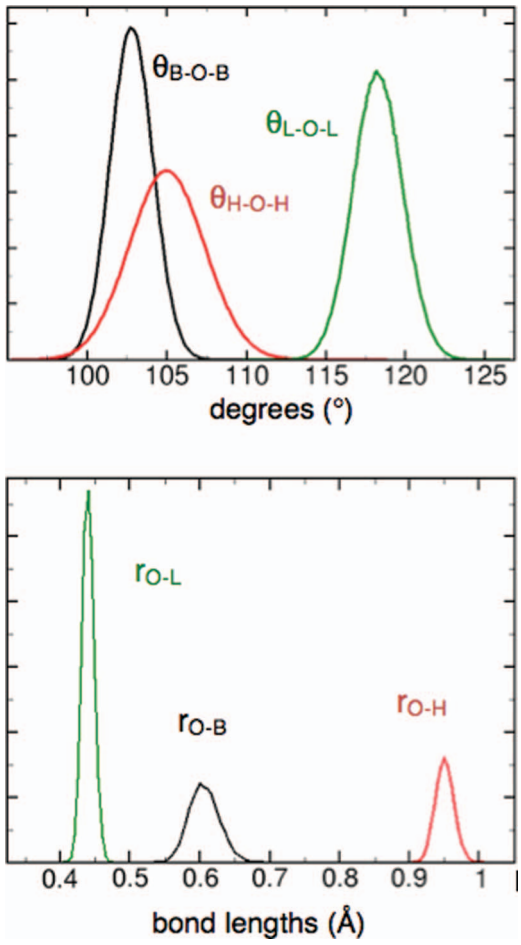

FIG. 5. Distribution of intramolecular angles and distances in the liquid state. Here, "L" denotes a lone pair and "B" a bonding pair. The molecular angle, H-O-H, has a broader spread than the angles involving electron pairs.

experience, constant energy simulations, as well as badly configured NPT and NVT setups, are prone to this effect.

D. Computational efficiency

While the proliferation of sub-molecular particles is rarely desirable, especially in large-scale systems, the amphiproticity and polarizability provided by LEWIS justifies a reasonable computational cost. In its current sub-optimized implementation (see *Methods*), single-precision LEWIS is ∼5-fold slower than TIP5P per time-step, as tested on a 2.26 GHz 8 Core MacPro workstation running a cubic box of 500 water molecules subject to periodic boundaries. A similar initial performance was observed for the polarizable (but unreactive) AMOEBA description of water.[16] Obviously, there is room for improvement as tabulated potentials can be memory intensive and the inter-nuclear potential forms are simple enough that analytical evaluation may be faster than spline interpolation.[74]

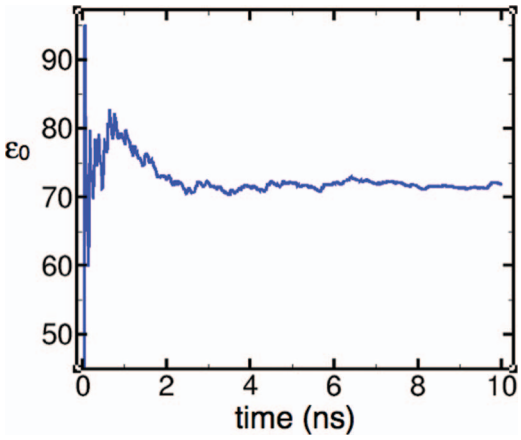

FIG. 6. Time evolution of the dielectric constant ε_0 of water at 300 K. Reasonable convergence occurs within 10 ns.

IV. CONCLUDING REMARKS

In LEWIS, we have developed a chemically intuitive and reasonably simple model that can capture essential features of molecular reactivity, polarizability, and flexibility in a seamless manner. Indeed, it is conceptually difficult to separate polarization from a dissociation or association event. Especially for the hydrides of an electronegative element such as oxygen, strong bond polarization typically signals the beginning of a reaction, while geometric rearrangements typically accommodate the approach of a nucleophile or electrophile. For example, the bond angle of water increases by 5°–7° upon protonation. Such changes should occur continuously in order to capture meaningful transition states and obey energy conservation.

The structures and dynamics that LEWIS predicts for liquid water are comparable to DFT predictions which are orders of magnitude more demanding in computational power. However, like any model, LEWIS has its limitations, some inherent in the underlying approach, and others due to the impossibility of exhaustively exploring its novel space of potential functions and parameters. With regard to the underlying approach, LEWIS makes two key assumptions in the interests of efficiency: a pseudo-classical treatment of the electrons in pairs, and a neglect of multi-body interactions. The former more than doubles the number of degrees of freedom that can be thermally activated, a major consequence of which is a proportional increase in the heat capacity. The pointillist representation of the electron distribution also fails to account for delocalization. A notable case is hydroxide where the lone pair arrangement enforces three-fold symmetry, a poor representation of the proposed donut-shape of the lone pairs.[25] The main consequence of eschewing multi-body interactions is difficulty in matching the dihedral angles of the deprotonated dimer[41] and hydrogen peroxide.[76]

Within the LEWIS approach, we encountered a number of trade-offs in the space of potential functions that we were able to explore. For example, the overestimation of water dimer enthalpy (see Table III), also an issue in common water models such as TIP3P or SPC,[6] resulted from a reduced weight given to this feature in the interest of better fits to other properties (see Sec. II B). Also noteworthy in this category is the behavior of hexameric water clusters. An earlier parameterization of LEWIS, based on a strictly gas phase training set,[38] favored cage-like hexamers over ring-like ones, in agreement with high-level QM calculations.[75] On the other

hand, the LEWIS parameters reported here do the opposite (see Table III). However, the energy differences involved are fairly small and it is unrealistic to expect chemical accuracy from an empirical model for reactivity, especially in its early stages. It remains that the compromises made in the interests of a highly efficient construct have presented fewer trade-offs than initially expected. This is due to the detailed nature of the potential forms and the training sets that have been employed.

Our follow-up studies with LEWIS will focus on protonic defects in bulk water (i.e., hydrated hydronium and hydrated hydroxide). Our understanding of the solvation of these ions and the relevant proton transfer dynamics has significantly increased over the past decade thanks to both neutron diffraction experiments[77] and *ab initio* simulations.[24,78,79] However, both diffraction experiments and high-level theory require extremely high ion concentrations. This is unfortunate because, at least in the case of hydroxide, the solvation structure is significantly concentration dependent.[79] LEWIS is not so constrained; with its ability to simulate thousands of water molecules on a common desktop platform, overlap between the ion solvation shells can be drastically reduced.

More generally, the approach presented here suggests a new direction for incorporating reactivity in routine molecular mechanics. Current work is extending the model to the hydrides of other elements of the second row of the periodic table with the goal of efficient simulations of organic reactions. We expect that such simulations will suggest reaction pathways that can then be explored by QM if more quantitatively accurate results are needed.

ACKNOWLEDGMENTS

We thank Peter Jordan and Michael Francis Hagan for helpful discussions, Mason Kramer for technical advice, and Vijay Pande for pointing us to Gromacs. This work was supported by the National Institutes of Health (NIH) Grant No. R01EB001035. Additional computational support was provided by the Brandeis HPC.

[1] C. J. Cramer and D. G. Truhlar, Chem. Rev. **99**(8), 2161 (1999).
[2] I. F. W. Kuo, C. J. Mundy, M. J. McGrath, J. I. Siepmann, J. VandeVondele, M. Sprik, J. Hutter, B. Chen, M. L. Klein, F. Mohamed, M. Krack, and M. Parrinello, J. Phys. Chem. B **108**(34), 12990 (2004).
[3] W. L. Jorgensen and J. Tirado-Rives, Proc. Natl. Acad. Sci. U.S.A. **102**(19), 6665 (2005).
[4] B. Guillot, J. Mol. Liq. **101**(1–3), 219 (2002).
[5] J. D. Bernal and R. H. Fowler, J. Am. Chem. Soc. **1**(8), 515 (1933).
[6] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**(2), 926 (1983).
[7] M. W. Mahoney and W. L. Jorgensen, J. Chem. Phys. **112**(20), 8910 (2000).
[8] J. L. F. Abascal, E. Sanz, R. G. Fernandez, and C. Vega, J. Chem. Phys. **122**(23), 234511 (2005).
[9] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gusteren, and J. Hermans, in *Intermolecular Forces*, edited by B. Pullman (Reidel, Dordrecht, 1981), pp. 331.
[10] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, J. Phys. Chem. **91**(24), 6269 (1987).
[11] Y. J. Wu, H. L. Tepper, and G. A. Voth, J. Chem. Phys. **124**(2), 024503 (2006).
[12] M. Levitt, M. Hirshberg, R. Sharon, K. E. Laidig, and V. Daggett, J. Phys. Chem. B **101**(25), 5051 (1997).
[13] S. Habershon, T. E. Markland, and D. E. Manolopoulos, J. Chem. Phys. **131**(2), 024501 (2009).
[14] J. W. Ponder and D. A. Case, Adv. Protein Chem. **66**, 27 (2003).
[15] P. Y. Ren and J. W. Ponder, J. Phys. Chem. B **107**(24), 5933 (2003).
[16] P. Y. Ren and J. W. Ponder, J. Phys. Chem. B **108**(35), 13427 (2004).
[17] M. Sprik and M. L. Klein, J. Chem. Phys. **89**(12), 7556 (1988); P. Cieplak, P. Kollman, and T. Lybrand, J. Chem. Phys. **92**(11), 6755 (1990).
[18] H. A. Stern, F. Rittner, B. J. Berne, and R. A. Friesner, J. Chem. Phys. **115**(5), 2237 (2001).
[19] P. L. Silvestrelli and M. Parrinello, Phys. Rev. Lett. **82**(16), 3308 (1999).
[20] J. W. Ponder, C. J. Wu, P. Y. Ren, V. S. Pande, J. D. Chodera, M. J. Schnieders, I. Haque, D. L. Mobley, D. S. Lambrecht, R. A. DiStasio, M. Head-Gordon, G. N. I. Clark, M. E. Johnson, and T. Head-Gordon, J. Phys. Chem. B. **114**(8), 2549 (2010).
[21] P. L. Silvestrelli and M. Parrinello, J. Chem. Phys. **111**(8), 3572 (1999).
[22] K. Laasonen, M. Sprik, M. Parrinello, and R. Car, J. Chem. Phys. **99**(11), 9080 (1993).
[23] M. E. Tuckerman, D. Marx, M. L. Klein, and M. Parrinello, Science **275**(5301), 817 (1997); S. S. Iyengar, T. J. F. Day, and G. A. Voth, Int. J. Mass Spectrom. **241**(2-3), 197 (2005).
[24] M. E. Tuckerman, A. Chandra, and D. Marx, Acc. Chem. Res. **39**(2), 151 (2006).
[25] M. E. Tuckerman, D. Marx, and M. Parrinello, Nature (London) **417**(6892), 925 (2002).
[26] D. Asthagiri, L. R. Pratt, J. D. Kress, and M. A. Gomez, Proc. Natl. Acad. Sci. U.S.A. **101**(19), 7229 (2004).
[27] S. Yoo, X. C. Zeng, and S. S. Xantheas, J. Chem. Phys. **130**(22), 2211020 (2009).
[28] P. L. Geissler, C. Dellago, D. Chandler, J. Hutter, and M. Parrinello, Science **291**(5511), 2121 (2001).
[29] R. E. Kozack and P. C. Jordan, J. Chem. Phys. **96**(4), 3131 (1992); L. Ojamae, I. Shavitt, and S. J. Singer, *ibid.* **109**(13), 5547 (1998); S. V. Shevkunov and A. Vegiri, Mol. Phys. **98**(3), 149 (2000).
[30] Y. Wu, H. Chen, F. Wang, F. Paesani, and G. A. Voth, J. Phys. Chem. B **112**(23), 7146 (2008).
[31] G. Brancato and M. E. Tuckerman, J. Chem. Phys. **122**(22), 224507 (2005).
[32] U. W. Schmitt and G. A. Voth, J. Phys. Chem. B **102**(29), 5547 (1998).
[33] M. K. Petersen, S. S. Iyengar, T. J. F. Day, and G. A. Voth, J. Phys. Chem. B **108**(39), 14804 (2004).
[34] G. A. Voth, Front. Biosci. **8**, S1384 (2003).
[35] A. Warshel and R. M. Weiss, J. Am. Chem. Soc. **102**(20), 6218 (1980).
[36] I. S. Ufimtsev, A. G. Kalinichev, T. J. Martinez, and R. J. Kirkpatrick, Chem. Phys. Lett. **442**(1–3), 128 (2007).
[37] I. S. Ufimtsev, A. G. Kalinichev, T. J. Martinez, and R. J. Kirkpatrick, Phys. Chem. Chem. Phys. **11**(41), 9420 (2009); S. T. Roberts, P. B. Petersen, K. Ramasesha, A. Tokmakoff, I. S. Ufimtsev, and T. J. Martinez, Proc. Natl. Acad. Sci. U.S.A. **106**(36), 15154 (2009).
[38] S. Kale, J. Herzfeld, S. Dai, and M. Blank, J. Biol. Phys. **38**(1), 49 (2012).
[39] H. L. Lemberg and F. H. Stillinger, J. Chem. Phys. **62**(5), 1677 (1975); F. H. Stillinger and A. Rahman, *ibid.* **68**(2), 666 (1978).
[40] C. J. Burnham and S. S. Xantheas, J. Chem. Phys. **116**(4), 1479 (2002).
[41] C. C. M. Samson and W. Klopper, J. Mol. Struct.: THEOCHEM **586**, 201 (2002).
[42] K. Ichikawa, Y. Kameda, T. Yamaguchi, H. Wakita, and M. Misawa, Mol. Phys. **73**(1), 79 (1991).
[43] C. J. Burnham and S. S. Xantheas, J. Chem. Phys. **116**(12), 5115 (2002).
[44] J. Jeon, A. E. Lefohn, and G. A. Voth, J. Chem. Phys. **118**(16), 7504 (2003).
[45] P. K. Mankoo and T. Keyes, J. Chem. Phys. **129**(3), 034504 (2008).
[46] A. Koide, J. Phys. B **9**(18), 3173 (1976).
[47] N. Metropolis and S. Ulam, J. Am. Stat. Assoc. **44**(247), 335 (1949); N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**(6), 1087 (1953).
[48] G. Bussi, D. Donadio, and M. Parrinello, J. Chem. Phys. **126**(1), 014101 (2007).
[49] M. Parrinello and A. Rahman, J. Appl. Phys. **52**(12), 7182 (1981); S. Nose and M. L. Klein, Mol. Phys. **50**(5), 1055 (1983).
[50] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen, J. Comput. Chem. **26**(16), 1701 (2005); B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl, J. Chem. Theory Comput. **4**(3), 435 (2008).
[51] R. W. Hockney, S. P. Goel, and J. W. Eastwood, J. Comput. Phys. **14**(2), 148 (1974).
[52] L. Verlet, Phys. Rev. **159**(1), 98 (1967); W. C. Swope, H. C. Andersen, P. H. Berens, and K. R. Wilson, J. Chem. Phys. **76**(1), 637 (1982).
[53] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, New York, 1987).

[54]M. Neumann, Mol. Phys. **50**(4), 841 (1983).
[55]S. Kale and J Herzfeld, J. Chem. Theory Comput. **7**(11), 3620 (2011).
[56]S. Izvekov and G. A. Voth, J. Chem. Phys. **123**(4), 044505 (2005).
[57]A. D. Boese, N. L. Doltsinis, N. C. Handy, and M. Sprik, J. Chem. Phys. **112**(4), 1670 (2000).
[58]K. T. Wikfeldt, M. Leetmaa, M. P. Ljungberg, A. Nilsson, and L. G. M. Pettersson, J. Phys. Chem. B **113**(18), 6246 (2009).
[59]K. Krynicki, C. D. Green, and D. W. Sawyer, Faraday Discuss. **66**, 199 (1978).
[60]A. K. Soper, Chem. Phys. **258**(2–3), 121 (2000).
[61]G. S. Kell, J. Chem. Eng. Data **12**(1), 66 (1967).
[62]W. Wagner and A. Pruss, J. Phys. Chem. Ref. Data **31**(2), 387 (2002).
[63]J. C. Grossman and L. Mitas, Phys. Rev. Lett. **94**(5), 056403 (2005).
[64]D. Eisenberg and W. Kauzmann, *The Structure and Properties of Water* (Oxford University Press, New York, 1969).
[65]Y. Q. Tu and A. Laaksonen, Chem. Phys. Lett. **329**(3–4), 283 (2000).
[66]S. A. Clough, Y. Beers, G. P. Klein, and L. S. Rothman, J. Chem. Phys. **59**(5), 2254 (1973).
[67]J. Verhoeven and A. Dymanus, J. Chem. Phys. **52**(6), 3222 (1970).
[68]L. Delle Site, A. Alavi, and R. M. Lynden-Bell, Mol. Phys. **96**(11), 1683 (1999).
[69]J. Kongsted, A. Osted, K. V. Mikkelsen, and O. Christiansen, Chem. Phys. Lett. **364**(3-4), 379 (2002).
[70]M. Uematsu and E. U. Franck, J. Phys. Chem. Ref. Data **9**(4), 1291 (1980).
[71]M. Neumann, Mol. Phys. **57**(1), 97 (1986).
[72]D. Marx and J. Hutter, in *Modern Methods and Algorithms of Quantum Chemistry*, edited by J. Grotendorst (John von Neumann Institute of Computing, Julich, NIC Series, 2000), Vol. 1, pp. 301.
[73]A. C. T. van Duin, S. Dasgupta, F. Lorant, and W. A. Goddard, J. Phys. Chem. A. **105**(41), 9396 (2001).
[74]D. van der Spoel, E. Lindahl, B. Hess, A. R. van Buuren, E. Apol, P. J. Meulenhoff, D. P. Tieleman, A. L. T. M. Sijbers, K. A. Feenstra, R. van Drunen, and H. J. C. Berendsen, Gromacs User Manual version 4.5.3, www.gromacs.org (2010).
[75]S. S. Xantheas, C. J. Burnham, and R. J. Harrison, J. Chem. Phys. **116**(4), 1493 (2002).
[76]W. C. Oelfke and W. Gordy, J. Chem. Phys. **51**(12), 5336 (1969).
[77]A. Botti, F. Bruni, S. Imberti, M. A. Ricci, and A. K. Soper, J. Chem. Phys. **121**(16), 7840 (2004); A. Botti, F. Bruni, S. Imberti, M. A. Ricci, and A. K. Soper, J. Chem. Phys. **120**(21), 10154 (2004); A. Botti, F. Bruni, S. Imberti, M. A. Ricci, and A. K. Soper, J. Mol. Liq. **117**(1–3), 81 (2005).
[78]D. Marx, M. E. Tuckerman, J. Hutter, and M. Parrinello, Nature (London) **397**(6720), 601 (1999); D. Marx, A. Chandra, and M. E. Tuckerman, Chem. Rev. **110**(4), 2174 (2010).
[79]B. Chen, J. M. Park, I. Ivanov, G. Tabacchi, M. L. Klein, and M. Parrinello, J. Am. Chem. Soc. **124**(29), 8534 (2002).
[80]P. Jensen, J. Mol. Spectrosc. **133**(2), 438 (1989).
[81]E. P. L. Hunter and S. G. Lias, J. Phys. Chem. Ref. Data **27**(3), 413 (1998).
[82]T. J. Sears, P. R. Bunker, P. B. Davies, S. A. Johnson, and V. Spirko, J. Chem. Phys. **83**(6), 2676 (1985).
[83]J. C. Owrutsky, N. H. Rosenbaum, L. M. Tack, and R. J. Saykally, J. Chem. Phys. **83**(10), 5338 (1985).
[84]M. J. S. Dewar and K. M. Dieter, J. Am. Chem. Soc. **108**(25), 8075 (1986).
[85]D. R. Lide, *CRC Handbook of Chemistry and Physics*, 76th ed. (CRC, Boca Raton, FL, 1995).
[86]J. E. Bartmess, J. A. Scott, and R. T. Mciver, J. Am. Chem. Soc. **101**(20), 6046 (1979).
[87]J. A. Odutola and T. R. Dyke, J. Chem. Phys. **72**(9), 5062 (1980).
[88]T. D. Fridgen, T. B. McMahon, L. MacAleese, J. Lemaire, and P. Maitre, J. Phys. Chem. A **108**(42), 9008 (2004).
[89]S. S. Xantheas and E. Apra, J. Chem. Phys. **120**(2), 823 (2004).
[90]F. H. Stillinger and C. W. David, J. Chem. Phys. **69**(4), 1473 (1978).