Predictions of the Properties of Water from First Principles

Robert Bukowski, *et al.*

Science **315**, 1249 (2007);
DOI: 10.1126/science.1136371

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005.

Council (project nos. 12X-1253 and K2002X-72X-10358-10A); Swedish Cancer Research Foundation (project no. 3911); Torsten och Ragnar Söderbergs Foundation; King Gustaf V:s and Drottning Victorias Foundation; Axel Linders Foundation; the Regional Developmental Board and the Chamber of Commerce and Industry in Western Sweden; Sahlgrenska Academy; the Neurological Foundation of New Zealand; the Health Research Council of New Zealand; the National Research Centre for Growth and Development, New Zealand; A.-M. Alborn, and J. Lind for the technical support; and J. Faijerson, M. Komitova, and G. Kuhn for scientific and technical advice.

Supporting Online Material
www.sciencemag.org/cgi/content/full/1136281/DC1
Materials and Methods
Figs. S1 to S5
Table S1

13 October 2006; accepted 24 January 2007
Published online 15 February 2007;
10.1126/science1136281
Include this information when citing this paper.

Predictions of the Properties of Water from First Principles

Robert Bukowski,[1] Krzysztof Szalewicz,[1] Gerrit C. Groenenboom,[2] Ad van der Avoird[2]

A force field for water has been developed entirely from first principles, without any fitting to experimental data. It contains both pairwise and many-body interactions. This force field predicts the properties of the water dimer and of liquid water in excellent agreement with experiments, a previously elusive objective. Precise knowledge of the intermolecular interactions in water will facilitate a better understanding of this ubiquitous substance.

Water has been extensively studied on account of its ubiquity and importance for so many aspects of human activity. The deceptively simple water molecule forms one of the most complex liquids and solids, and investigations continue to focus on all forms of pure water, including small clusters [e.g., (*1*, *2*)], bulk liquid [e.g., (*3*–*6*)], and a multitude of ice polymorphs. Similarly numerous are theoretical analyses of water, all of which require knowledge of the intermolecular potential (the derivatives of this potential give the force field that governs the dynamics). Most such investigations use empirical pair potentials fitted to reproduce certain measured bulk properties in Monte Carlo (MC) or molecular dynamics (MD) simulations of water. These "effective" potentials account for the important many-body interactions in water by (nonphysical) deformations of the true pair potential. The well-known result is that such potentials poorly describe the water dimer, give very inaccurate second virial coefficients, and fail to reproduce experimental spectra of small water clusters. Therefore, studies of molecular-scale properties of water with empirical potentials, such as the molecular jump mechanism of water reorientation (*7*), may suffer from an inadequate representation of the force field. Another known drawback of empirical potentials is that the quality of their predictions deteriorates quickly beyond the range of thermodynamic parameters used in the fitting procedure. Moreover, there does not appear to be any systematic method to improve the predictive accuracy of these potentials.

Another way of obtaining the force fields—which does not require prior knowledge of any experimental data—is by quantum mechanical ab initio calculations. Such an approach can provide the most reliable foundation for an understanding of water and other substances. However, the accuracy of ab initio force fields is limited by unavoidable approximations in the level of theory and incompleteness of basis sets.

A first-principles approach can proceed either "on the fly" (i.e., by computing the electronic energy for each configuration of the N molecules present in a simulation of bulk water, where N at least equals 32) or by computing the interaction energy as a sum of relatively simple analytic many-body potentials (pair, three-body, etc.) fitted beforehand to ab initio data. The former approach can be realized only with the fastest electronic structure methods such as density functional theory (DFT). Although for some time the published results suggested that this approach reproduced the properties of liquid water very well, recent work has shown that this agreement was due to fortuitous choices of the DFT functionals (*8*). Even for the same functional, different values of some parameters in the simulations may lead to markedly different predictions [e.g., the results in (*9*) and (*10*)]. Moreover, the current DFT approaches cannot describe the dispersion component of the intermolecular interaction energy, which is nonnegligible in water.

The many-body expansion of the potential requires ab initio calculations on dimers, trimers, and larger clusters. If the many-body expansion converges sufficiently rapidly, such calculations can be restricted to only a few small clusters and can make use of accurate electronic structure methods. In practice, the interaction energy must already be well reproduced at the three-body level, because calculations of the complete four-body potential for water would be too time-consuming. An important advantage of using the many-body expansion approach is that it can be systematically improved by extending the level of theory, using larger basis sets, calculating more grid points, and improving the form of the fitting function. A large number of ab initio water dimer potentials have been published [e.g., (*11*–*16*)]. However, so far no first-principles approach has simultaneously reproduced the experimental results for both the water dimer and the condensed phases of water. Here we present substantial progress toward this goal.

We developed a pair potential for water on the basis of the coupled cluster method with single, double, and noniterative triple excitations [CCSD(T)], generally regarded as the most accurate of practically applicable electronic structure methods. We have used the same set of 2510 carefully selected grid points as in the SAPT-5s potential developed previously by our group (*12*). The H_2O monomer was assumed to be rigid in the averaged ground-state vibrational geometry, as in (*12*). For each grid point, we first computed the interaction energies with the use of second-order perturbation theory based on the Møller-Plesset partitioning of the Hamiltonian (MP2). The MP2 energies were computed in augmented triple- and quadruple-zeta quality basis sets supplemented by midbond functions. The energies were then extrapolated to the complete basis set (CBS) limit by means of the well-established extrapolation formula that assumes a convergence rate proportional to the inverse third power of the cardinal number of the basis set. The CCSD(T) contribution beyond the MP2 level was computed in the triple-zeta quality basis set and added to the extrapolated MP2 values. We also performed single-point calculations in a quintuple-zeta quality basis. The basis set convergence patterns indicate that the uncertainty of the computed interaction energies is about 0.07 kcal/mol. For comparison, the uncertainty of the SAPT-5s potential was 0.3 kcal/mol. The accuracy of the current calculations is virtually the same as that of the most extensive published ab initio work [e.g., (*17*)]. However, the cited calculations have been performed only for a few selected geometries of the dimer, whereas we have obtained the complete six-dimensional potential surface.

[1]Department of Physics and Astronomy, University of Delaware, Newark, DE 19716, USA. [2]Theoretical Chemistry, Institute for Molecules and Materials, Radboud University Nijmegen, Toernooiveld 1, 6525 ED Nijmegen, Netherlands.

We fitted the computed interaction energies by an analytic potential function. The form of this function is analogous to that used for the SAPT-5s potential (*12*), except for an added "polarization" term to streamline molecular simulations. This term also fits the induction energy more accurately than did the site-site expansion used in SAPT-5s. A single polarizable center, almost identical to that used in the SAPT three-body potential of (*18*, *19*), was placed 0.1946906 Å away from the oxygen atom toward the hydrogens and had an isotropic polarizability of 9.922 bohr^3. During the fit of the asymptotic coefficients of the potential [in a manner analogous to (*12*)], the contribution of the polarization term was subtracted from the long-range interaction energies. The root-mean-square deviation (RMSD) of the fit for the points with negative energies was 0.09 kcal/mol. The analytic polarizable potential produced by this fit is hereafter termed CC-pol. The functional form of CC-pol represents a good compromise between the accuracy of reproducing the computed points and the simplicity needed for molecular simulations.

The CC-pol potential was used in nearly exact quantum mechanical calculations to predict the spectrum of the water dimer. The intermolecular vibrations of this dimer are strongly anharmonic, and its ro-vibrational levels are split by quantum mechanical tunneling between eight equivalent equilibrium geometries (*20*). The largest splitting $a(K)$ is due to the acceptor switching tunneling; smaller splittings i_1 and i_2 originate from the donor-acceptor interchange tunneling. The vibration-rotation-tunneling (VRT) levels for states with overall rotation quantum numbers J and K equal to 0, 1, and 2 were computed by an approach described in (*21*) that is similar to a coupled-channel scattering method, well converged with respect to the number of channels included. It has been shown (*22*) by calculations with various water potentials from the literature that the VRT levels of the water dimer are extremely sensitive to such details as the shape of the anharmonic potential wells and the barriers between these wells.

The CC-pol dimer levels agree very well with the levels deduced from the measured spectra, both for the ground state and for the excited intermolecular vibrations (Fig. 1). The only substantial error occurs for the interchange splitting $a(0) + a(1)$, which is 19% too large. This observable is particularly sensitive to monomer flexibility effects that decrease its value (*23*). Thus, rigid-monomer potentials such as CC-pol should overestimate this splitting.

The CC-pol prediction of the dimer spectra is markedly better than those of the earlier ab initio potentials: ASP-W, ASP-S, SAPT-5s, TTM2.1, and SDFT-5s (Table 1). The empirical TIP4P and MCY-KW potentials fail completely, even though the latter is polarizable and was partly fitted to ab initio data. The agreement of the CC-pol spectra with experiment represents an improvement over the predictions of the VRT(ASP-W)III potential, containing parameters fitted to these spectra. Only SAPT-5st, which was tuned to reproduce the measured splitting $a(0) + a(1)$, and VRT(MCY-5f) give somewhat better agreement with the measured dimer levels. The VRT(MCY-5f) potential is 12-dimensional (i.e., it includes flexible monomers), and all of its parameters were optimized to fit the H_2O dimer spectra. It has not yet been tested in simulations of liquid water (see below).

The second virial coefficient computed from CC-pol agrees very well with the compilation of experimental results by Harvey and Lemmon (*24*) (Fig. 2). For temperatures lower than 500 K, it proves even more accurate than SAPT-5s, which was sufficiently accurate to guide the experimental determination (*24*).

The CC-pol potential was also applied together with the SAPT three-body potential from (*18*, *19*) in NVE (constant number of particles,

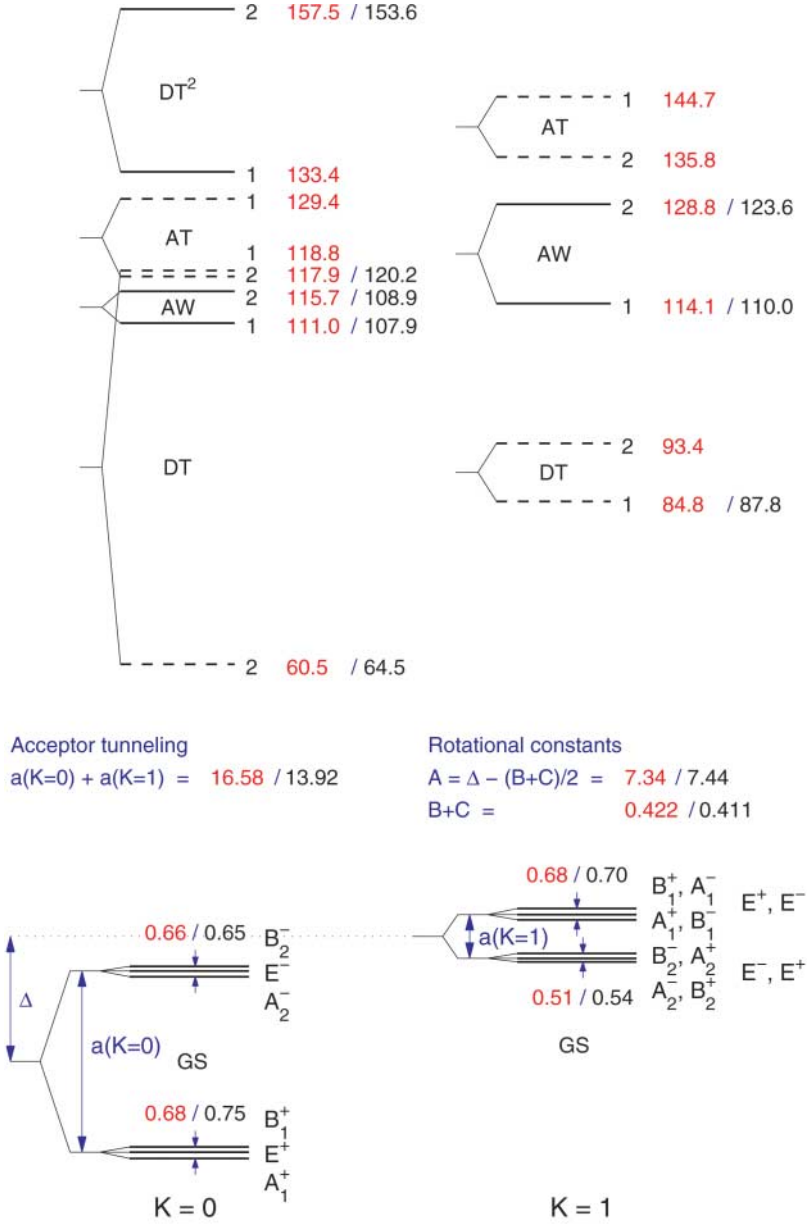

Fig. 1. VRT levels of the H_2O dimer (in cm^{-1}) from converged calculations with the CC-pol ab initio potential (first number, in red), in comparison with experimental data (*31*, *32*) (second number, in black). Virtually all the measured levels have been shown except for some data for the ground state nearly redundant with those presented. The symbols $A_i^{\pm}$ and $B_i^{\pm}$ are standard spectroscopic notations for VRT states [see (*21*)]. Abbreviations for modes: GS, ground state; DT, donor torsion; AW, acceptor wag; AT, acceptor twist; DT^2, donor torsion overtone. The individual GS splittings $a(K)$ for $K = 0$ and $K = 1$ could not be measured, so their sum is used in the comparison. The excited levels are drawn at a smaller energy scale than the GS levels and the small splittings are omitted, that is, the origins 1 and 2 (the averages of the $A_1^{\pm}$, $B_1^{\pm}$ and $A_2^{\pm}$, $B_2^{\pm}$ levels, respectively) are shown.

volume, and energy) MD simulations of liquid water in ambient conditions. The three-body potential used was fitted to ab initio calculations for the water trimer. It is the only such potential that includes three-body exchange terms in addition to polarization effects. The four- and higher-body effects were computed with the polarization model included in CC-pol (applying proper subtractions to avoid double counting of two- and three-body induction effects). Up to 512 molecules were used and the velocities were scaled during the equilibration to achieve the ambient temperature. Most calculations were performed with 256 molecules in a box size of 19.71 Å, corresponding to a density of 1 g/cm^3. The time step was 1 fs and the simulation time was 30 ps, including a 10-ps equilibration period. The radial atom-atom distribution functions g_{XY} obtained in these simulations are shown in Fig. 3. The simulations with only the pair potential result in g_{XY} values quite different from the measured ones, thus revealing the crucial role of pair-nonadditive forces in water. Including the many-body nonadditive potential with the pair-only CC-pol potential (CC-pol+NB curves) strongly improves agreement with experiment. In particular, the positions of the minima and maxima on the g_{OO} curve, which are measured most accurately, are in near-perfect agreement. The g_{OH} and g_{HH} functions from CC-pol+NB calculations reproduce the experiment equally well. The CC-pol+NB(ind) model—which represents both the pair and many-body polarization effects just by the simple polarization term in CC-pol—also gives g_{XY} functions that agree very well with experiment. This observation is encouraging because many-body effects beyond the polarization model are very difficult to compute. We note that the agreement is likely as good as possible considering the neglected quantum and monomer-flexibility effects. In particular, there are indications (*10*) that the latter make the first minimum in g_{OO} deeper, which would improve the agreement with experiment, although quantum effects (*25*) may partly cancel this trend.

The g_{XY} functions obtained from CC-pol reproduce the experimental data virtually as well as the empirical potentials fitted to bulk water observables. We computed RMSDs from experiment for three characteristic points on each curve: first maximum, first minimum, and second maximum. This quantity amounts to 0.13 for CC-pol and 0.15 for TIP4P (*26*).

The empirical potentials fitted to spectral data were recently used in simulations of liquid water with the many-body effects described only by a simple classical polarization model (*27*). The empirical VRT(ASP-W)III potential performed rather well. Thus, among all published potentials, VRT(ASP-W)III provides the best overall (i.e., for both clusters and liquid water) agreement with experiment. Although VRT(ASP-W)III gave radial distribution functions in good agreement with experiment, overall the CC-pol potential compares more favorably. The RMSD for the three characteristic points on the three g_{XY} VRT(ASP-W)III curves is 0.21. In particular, the first peak on the VRT(ASP-W)III g_{OO} curve is about 19% too low, whereas it is within 5% of experiment in the CC-pol simulation. The 12-dimensional VRT(MCY-5f) potential has not been used in simulations so far. Its rigid form with added polarization effects, VRT(MCY-5r/pol), performed poorly in simulations, producing pair distribution curves similar to those given by the two-body-only CC-pol (i.e., the dotted curves in Fig. 3). Also, SAPT-5st performed rather poorly in liquid water simulations, similarly to SAPT-5s.

We also computed the internal energy, the self-diffusion coefficient, and the coordination number of liquid water. The internal energy for the CC-pol+NB model is –10.89 kcal/mol, compared to the experimental value of –9.92 ± 0.3 kcal/mol [see (*28*)]. Our value is in still better agreement with experiment if the correction for quantum effects, equal to 0.86 kcal/mol (*25*), is added. The classical and quantum values for the TIP4P potential are –9.81 and –8.95 kcal/mol, respectively (*25*). For the diffusion coefficient, a value of 2.4 × 10^{-5} cm^2/s was obtained, in perfect agreement with experiment (*29*). This agreement is partly fortuitous, as quantum effects may increase the theoretical value by about 50% (*25*). The TIP4P potential gives the classical and quantum diffusion coefficients of 3.6 × 10^{-5} cm^2/s and 5.5 × 10^{-5} cm^2/s, respectively (*25*). The so-called coordination number was determined by integrating the function g_{OO} up to the first minimum. The result for the CC-pol+NB model is 5.6, whereas the experimental curve (*30*) gives 4.8. We have not calculated the number of hydrogen bonds for the CC-pol models, but such numbers are known for the SAPT-5s models (*18, 19*). We found that the number of hydrogen bonds is well correlated with the coordination number because hydrogen bonds lead to a more structured liquid and narrower g_{OO} peaks. For example, for the SAPT-5s+3B, SAPT-5s+NB, and SAPT-5s models, the numbers of hydrogen bonds are 3.31, 3.34, and 2.77, whereas the coordination numbers are 7.9, 8.6, and 11.3, respectively. Because the relation between the number of hydrogen bonds and the coordination number was found to be approximately linear, we

Table 1. Root-mean-square relative percentage errors in various properties of H_2O dimer from calculations with different potentials, compared with experimental data from (*31*, *32*) and references therein. The following properties are used in this analysis: ground-state rotational constants A and $B + C$, ground state tunneling splittings $a(0) + a(1)$ (acceptor switch), i_1 and i_2 (donor acceptor interchange), and frequencies of the intermolecular vibrations DT, AW, AT, and DT^2 for $K = 0$ (see Fig. 1). For more information on these properties, see (*21*).

Potential	Rotational constants	Tunneling splittings	Vibrational frequencies
MCY-KW*	9.1	87	37
TIP4P†	14	81	23
ASP-W‡	7.4	325	17
ASP-S‡	5.2	110	15
TTM2.1§	16	72	12
VRT(ASP-W)III‖	6.6	38	7.1
SAPT-5s¶	4.9	26	7.8
SDFT-5s#	2.9	35	6.6
CC-pol	2.1	12	4.4
VRT(MCY-5f)**	3.8	11	1.7
SAPT-5st††	1.7	2.3	5.8

*From energy levels calculated in (*22*) with potential of (*33*). †From energy levels calculated in present work with potential of (*26*). ‡From energy levels calculated in (*22*) with potential of (*11*). §From energy levels calculated in present work using the potential of (*15*) with monomer geometries fixed in the equilibrium position. ‖From energy levels calculated in (*31*, *32*, *34*) using ASP-W potential form with parameters fitted to spectral data, except for rotational constants calculated in present work. ¶From energy levels calculated in (*21*) with potential of (*12*), excited levels calculated in present work. #From energy levels calculated in (*16*) and in present work with potential of (*16*). **From energy levels calculated in (*31*, *32*, *35*) using an extended MCY potential form with all parameters fitted to spectral data. ††From energy levels calculated in (*21*) with potential of (*12*) tuned to reproduce the measured interchange splitting $a(0) + a(1)$.

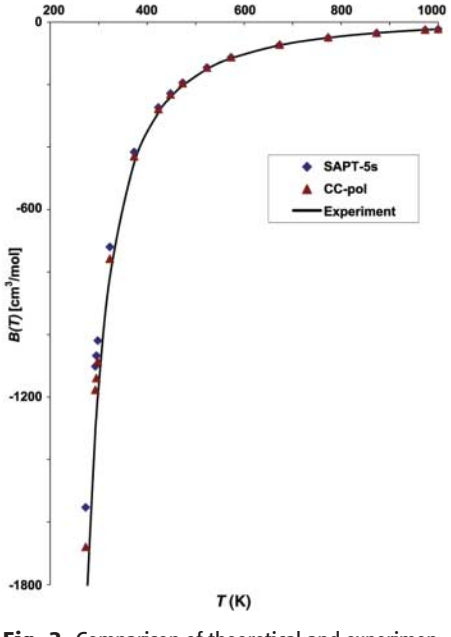

Fig. 2. Comparison of theoretical and experimental second virial coefficients. Theoretical data were computed including quantum effects to infinite order. Experimental data are from (*24*).

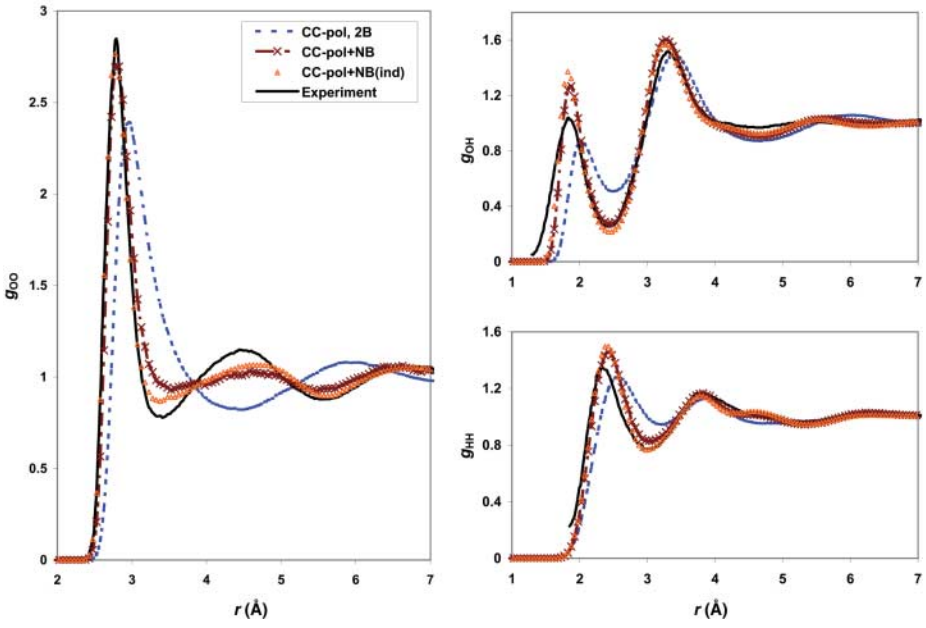

Fig. 3. Atom-atom radial distribution functions from MD simulations based on the CC-pol potential. See text for explanations of acronyms. The experimental curves are from (*30*).

could extrapolate this dependence to obtain 3.8 hydrogen bonds for the CC-pol+NB model. Thus, this model supports the standard picture of water, wherein each molecule is on average almost tetrahedrally coordinated, rather than the lower coordination geometry proposed in (*3*, *5*). With the use of only the CC-pol pair potential, the simulations gave a very large coordination number, 11.9 (the extrapolated number of hydrogen bonds equal to 2.7). Hence, the many-body interactions play an important role in determining the structure of liquid water, particularly in predicting tetrahedral coordination.

The ab initio water pair potential developed in this work recovers well a diverse range of experimental data from water dimer to liquid water. It predicts dimer spectra and second virial coefficients that not only agree well with existing experimental data, but also can be considered to complement experiments in spectral or temperature ranges inaccessible to measurement. When the CC-pol potential together with an earlier representation of three-body forces (*18*, *19*) was used in simulations of liquid water, the predictions agreed well with the neutron and x-ray diffraction data. These predictions, made entirely from first principles, are of comparable accuracy to results of simulations with empirical potentials fitted to liquid water experimental data.

We believe that the ab initio force field presented here will find numerous applications in predicting the properties of water. It can be used, for example, to resolve the current controversies about the coordination of water molecules in the liquid (*3*–*6*). The analysis of the temporal structures in MD simulations should provide the ultimate picture of liquid water. The force field can also be used to investigate the numerous polymorphic forms of ice. Important applications can be made in extreme regimes where empirical potentials fail completely, such as supercritical, overcooled, or confined water.

Further improvements in the first-principles predictions for water should take account of the monomer flexibility and of quantum effects in the liquid simulations. For the former case, the first step has recently been attained (*23*). Quantum effects in molecular simulations can be accounted for by either path-integral MC or centroid MD (*25*) methods.

References and Notes

1. K. Liu, J. Cruzan, R. Saykally, *Science* **271**, 929 (1996).
2. J. K. Gregory, D. C. Clary, K. Liu, M. G. Brown, R. Saykally, *Science* **275**, 814 (1997).
3. Ph. Wernet *et al.*, *Science* **304**, 995 (2004); published online 1 April 2004 (10.1126/science.1096205).
4. J. D. Smith *et al.*, *Science* **306**, 851 (2004).
5. A. Nilsson *et al.*, *Science* **308**, 793a (2005).
6. J. D. Smith *et al.*, *Science* **308**, 793b (2005).
7. D. Laage, J. T. Hynes, *Science* **311**, 832 (2006); published online 25 January 2006 (10.1126/science.1122154).
8. J. VandeVondele *et al.*, *J. Chem. Phys.* **122**, 014515 (2005).
9. M. Sharma, Y. Wu, R. Car, *Int. J. Quantum Chem.* **95**, 821 (2003).
10. M. Allesch, E. Schwegler, F. Gygi, G. Galli, *J. Chem. Phys.* **120**, 5192 (2004).
11. C. Millot, A. J. Stone, *Mol. Phys.* **77**, 439 (1992).
12. E. M. Mas *et al.*, *J. Chem. Phys.* **113**, 6687 (2000).
13. C. J. Burnham, S. S. Xantheas, *J. Chem. Phys.* **116**, 1500 (2002).
14. C. J. Burnham, S. S. Xantheas, *J. Chem. Phys.* **116**, 5115 (2002).
15. G. S. Fanourgakis, S. S. Xantheas, *J. Phys. Chem. A* **110**, 4100 (2006).
16. R. Bukowski, K. Szalewicz, G. C. Groenenboom, A. van der Avoird, *J. Chem. Phys.* **125**, 044301 (2006).
17. G. S. Tschumper *et al.*, *J. Chem. Phys.* **116**, 690 (2002).
18. E. M. Mas, R. Bukowski, K. Szalewicz, *J. Chem. Phys.* **118**, 4386 (2003).
19. E. M. Mas, R. Bukowski, K. Szalewicz, *J. Chem. Phys.* **118**, 4404 (2003).
20. R. S. Fellers, C. Leforestier, L. B. Braly, M. G. Brown, R. J. Saykally, *Science* **284**, 945 (1999).
21. G. C. Groenenboom *et al.*, *J. Chem. Phys.* **113**, 6702 (2000).
22. R. S. Fellers, L. B. Braly, R. J. Saykally, C. Leforestier, *J. Chem. Phys.* **110**, 6306 (1999).
23. K. Szalewicz, G. Murdachaew, R. Bukowski, O. Akin-Ojo, C. Leforestier, in *Lecture Series on Computer and Computational Science: ICCMSE 2006*, G. Maroulis, T. Simos, Eds. (Brill Academic, Leiden, Netherlands, 2006), vol. 6, pp. 482–491.
24. A. H. Harvey, E. W. Lemmon, *J. Phys. Chem. Ref. Data* **33**, 369 (2004).
25. L. Hernández de la Peña, P. G. Kusalik, *J. Chem. Phys.* **121**, 5992 (2004).
26. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **79**, 926 (1983).
27. N. Goldman, C. Leforestier, R. J. Saykally, *Philos. Trans. R. Soc. London Ser. A* **363**, 493 (2005).
28. W. L. Jorgensen, *J. Chem. Phys.* **77**, 4156 (1982).
29. R. Mills, *J. Phys. Chem.* **77**, 685 (1973).
30. A. K. Soper, *Chem. Phys.* **258**, 121 (2000).
31. F. N. Keutsch, N. Goldman, H. A. Harker, C. Leforestier, R. J. Saykally, *Mol. Phys.* **101**, 3477 (2003).
32. F. N. Keutsch *et al.*, *J. Chem. Phys.* **119**, 8927 (2003).
33. S. Kuwajima, A. Warshel, *J. Phys. Chem.* **94**, 460 (1990).
34. N. Goldman *et al.*, *J. Chem. Phys.* **116**, 10148 (2002).
35. C. Leforestier, F. Gatti, R. Fellers, R. Saykally, *J. Chem. Phys.* **117**, 8710 (2002).

36. We thank C. Leforestier for many discussions of the water dimer spectra. Supported by NSF grants CHE-0239611 and CHE-0555979.

16 October 2006; accepted 26 January 2007
10.1126/science.1136371

The Focusing of Electron Flow and a Veselago Lens in Graphene *p-n* Junctions

Vadim V. Cheianov,[1]* Vladimir Fal'ko,[1] B. L. Altshuler[2,3]

The focusing of electric current by a single *p-n* junction in graphene is theoretically predicted. Precise focusing may be achieved by fine-tuning the densities of carriers on the *n*- and *p*-sides of the junction to equal values. This finding may be useful for the engineering of electronic lenses and focused beam splitters using gate-controlled *n-p-n* junctions in graphene-based transistors.

There are many similarities between optics and electronics. Rays in geometrical optics are analogous to classical trajectories of electrons, whereas electron de Broglie waves can interfere. The electron microscope is one example of the technological implementation of this similarity. The analogy with optics may also hold considerable potential for semiconductor elec-