

Optimal finite-range atomic basis sets for liquid water and ice

Fabiano Corsetti¹, M-V Fernández-Serra², José M Soler³
and Emilio Artacho^{1,4,5,6}

¹ CIC nanoGUNE, 20018 Donostia-San Sebastián, Spain

² Dept. of Physics and Astronomy, Stony Brook University, Stony Brook, New York 11794-3800, USA

³ Dept. de Física de la Materia Condensada, Universidad Autónoma de Madrid, 28049 Madrid, Spain

⁴ Theory of Condensed Matter, Cavendish Laboratory, University of Cambridge, Cambridge CB3 0HE, United Kingdom

⁵ Basque Foundation for Science Ikerbasque, 48011 Bilbao, Spain

⁶ Donostia International Physics Center, 20018 Donostia-San Sebastián, Spain

E-mail: f.corsetti@nanogune.eu

Abstract. Finite-range numerical atomic orbitals are the basis functions of choice for several first principles methods, due to their flexibility and scalability. Generating and testing such basis sets, however, remains a significant challenge for the end user. We discuss these issues and present a new scheme for generating improved polarization orbitals of finite range. We then develop a series of high-accuracy basis sets for the water molecule, and report on their performance in describing the monomer and dimer, two phases of ice, and liquid water at ambient and high density. The tests are performed by comparison with plane-wave calculations, and show the atomic orbital basis sets to exhibit an excellent level of transferability and consistency. The highest-order bases (quadruple- ζ) are shown to give accuracies comparable to a plane-wave kinetic energy cutoff of at least ~ 1000 eV for quantities such as energy differences and ionic forces, as well as achieving significantly greater accuracies for total energies and absolute pressures.

PACS numbers: 71.15.Ap, 71.15.Mb, 31.15.aq, 31.15.xr

1. Introduction

Many large-scale density-functional theory (DFT) methods make use of localized basis functions, since locality in some form is an essential prerequisite for developing a method that scales linearly with system size [1]. In particular, the use of strictly localized orbitals [2] results in the Hamiltonian being formally sparse, without needing to impose a cutoff tolerance on the matrix elements.

Numerical atomic orbitals (NAOs) of finite range [2] have been found to be particularly well-suited for DFT calculations, due to the fact that they are very flexible, and, therefore, only a small number of them is usually needed to obtain accurate results [3, 4, 5, 6, 7, 8]. The transferability of the NAO basis between different systems has also been found to be quite reasonable [3, 4].

The most important drawback of using NAOs is the lack of a systematic way of improving the basis to arbitrary precision, such as what can easily be achieved in plane-wave (PW) methods [9] by increasing the kinetic energy cutoff, or in real-space-grid methods [10, 11] by decreasing the grid spacing. This is an important shortcoming of the method, since it means that it is not possible to assess the accuracy of the calculation with respect to the basis set with any degree of certainty. In contrast, it is standard practice in studies using PW methods to perform preliminary convergence tests for the kinetic energy cutoff on representative samples of the system in question, in order to tune the cutoff to be used in the production calculations to a given degree of accuracy.

In this paper, we aim to demonstrate best practice in the generation of transferable NAO basis sets of increasing accuracy for ordered and disordered bulk material, and to show how it is possible to determine with confidence the level of accuracy of the basis by careful comparison with PW calculations.

As a test case, we focus on water, in the liquid and two solid phases. Water is a material of fundamental importance in many different areas of science, and is currently of great interest to practitioners of *ab initio* techniques, since the subtle interplay between different interaction mechanisms that gives rise to its unusual properties (e.g., its notoriously complex phase diagram [12], and the large number of anomalies in the liquid phase [13, 14]) is still not well understood (see, e.g., Refs. [15, 16]). However, *ab initio* molecular dynamics (AIMD) studies using PW methods have been limited by the small system sizes and simulation times that are accessible due to computational cost, while the accuracy of existing NAO basis sets for such systems has never been systematically verified, as we propose here.

Using the water monomer as our starting point, we construct double-, triple-, and quadruple- ζ basis sets for hydrogen and oxygen. We make use of a previously-documented confinement potential for the valence orbitals [4], and propose a simple new form for the polarization orbitals. We find this new confinement scheme to give a good control over the overall shape of the polarization orbitals by varying a single parameter, making it ideal for variational optimization. We test our basis sets on the water dimer, ice I_c and VIII, and liquid water, showing our most accurate bases to achieve the level of precision of high-quality PW calculations for energies, pressures, and forces, and all bases to be highly transferable between systems. Therefore, our basis sets can be used with confidence for ambitious future studies of aqueous systems, as they enable cost-efficient large-scale simulations of high accuracy.

2. Computational methods

We use the SIESTA [17] code for generating the NAO bases and performing the DFT calculations with them, and the ABINIT [18] code for the comparison calculations using PW bases. For both codes we use the PBE [19] semi-local (GGA) exchange-correlation functional, and the same set of norm-conserving pseudopotentials in Troullier-Martins form [20]. The pseudopotential core radius is 0.66 Å for all angular momentum channels of H and 0.61 Å for all channels of O; we also employ a small non-linear core correction [21] for O. The pseudopotentials are factorized in the separable Kleinman-Bylander form [22], with the same local and non-local components used in both codes [23].

The NAOs making up our basis sets are composed of a freely-varying radial function multiplied by a spherical harmonic; the radial part is defined numerically and is strictly zero beyond a cutoff radius r_c . The SIESTA method for generating NAO bases has been documented in several previous publications [24, 4, 5]. The main ideas are:

- a soft confinement potential of the form $V_0 \exp\{-(r_c - r_i)/(r - r_i)\}/(r_c - r)$ between an inner radius r_i and r_c [4];
- a scheme for obtaining multiple- ζ for each orbital [24], inspired by the split valence method used in quantum chemistry for Gaussian basis sets [25, 26];
- polarization orbitals obtained either by applying a small electric field within perturbation theory [24], or by adding unoccupied atomic shells of higher l with soft confinement [4].

Using these schemes, it is possible to achieve significant optimization of a fixed-size basis by varying the free parameters for each shell: r_c , r_i and V_0 , and the matching radii $\{r_{d\zeta}, r_{t\zeta}, \dots\}$ which define the multiple- ζ orbitals. Increasing the orbital cutoff radius generally provides a better quality basis, but also increases its computational cost. Two schemes have been proposed for achieving an optimal balance between accuracy and cost based on a single parameter: the orbital energy shift in the isolated atom caused by confinement [24], and a fictitious ‘basis enthalpy’ calculated using the orbital volume and a pressure-like variable [5].

For the valence shells, soft confinement has proven to be very satisfactory: not only does it remove the problematic derivative discontinuity introduced by hard-walled confinement, but it also performs better at a given r_c from a variational point of view. This can be attributed to the fact that it has little effect on the shape of the orbital in the core region, resulting in a good agreement with the free atom orbital.

Instead, the choice of polarization orbitals is more problematic, as the free atom orbitals of higher l can become quite extended, or even entirely unbound. The explicit polarization of the pseudoatom by a small electric field provides an elegant and parameter-free solution, as the extent of the polarization orbitals is defined by that of the orbitals they polarize. Practical applications, however, have shown that a better variational estimate is instead usually obtained by including the unoccupied atomic shells of higher l with an aggressive soft confinement, in order to control the position of the maximum of the radial part of the orbital ‡. It is also important to note that the former method only allows for orbitals up to $l_v + 1$ (l_v being the highest

‡ This shows some similarity to the maximum overlap approach commonly used to define the polarization orbitals in Gaussian basis sets [25, 26].

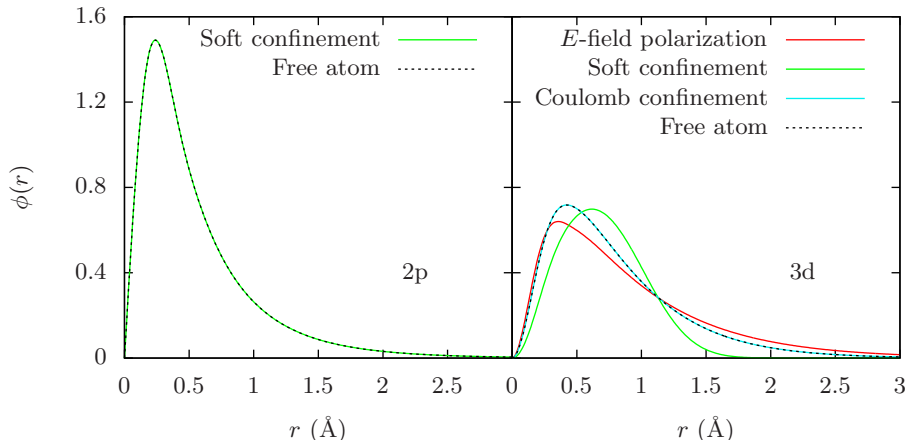


Figure 1. Shape of the 2p (occupied) and 3d (unoccupied) orbitals for O. $r_c = 5.3 \text{ \AA}$ both for the two confinement schemes and the E -field polarization orbital.

angular momentum for the valence shells), while the latter method can include shells of any l .

In this work we propose a new confinement scheme for obtaining short-range polarization orbitals that exhibit a good agreement with the free atom orbitals in the core region. We use a Yukawa-like screened Coulomb potential of the form

$$V(r) = -Q_0 \frac{e^{-\lambda r}}{\sqrt{r^2 + \delta^2}}, \quad (1)$$

for which the main parameter for variational optimization is the strength Q_0 §. δ is introduced to avoid numerical difficulties arising from the singularity at $r = 0$, and after some tests has been fixed to $0.01 a_0$. λ can additionally be used for fine tuning of the orbital tail (the default value is set to 0, thus making Eq. 1 a normal Coulomb potential).

Fig. 1 shows the radial part of the orbitals obtained with our new Coulomb confinement scheme for the 3d shell of O. Its decay is intermediate between those obtained by E -field polarization and soft confinement, and is similar to that of the free atom orbital. Within a double- ζ basis with a single polarization shell ($d\zeta + p$), the Coulomb confinement scheme results in a variationally better basis for the water monomer and dimer respect to soft confinement (by ~ 80 meV/molecule), as well as a smaller counterpoise (CP) correction [27, 28] (by 5 meV). Fig. 2 shows a similar behaviour for the 3d shell of Si; in this case, however, the free atom orbital is unbound. As before, the radial part of the Coulomb-confined d orbitals exhibits a slower decay than the soft-confined one, thereby achieving a greater overlap with that of the p orbitals they are polarizing, which gives the latter more flexibility. In bulk silicon, this results in Coulomb confinement gaining ~ 4 meV/atom respect to soft confinement.

§ The soft confinement potential is included in addition to the Coulomb potential, but now only for removing the discontinuity at r_c rather than for shaping the orbital. Default values of $V_0 = 40 \text{ Ry}\cdot a_0$ and $r_i = 0.9r_c$ can generally be used without further optimization.

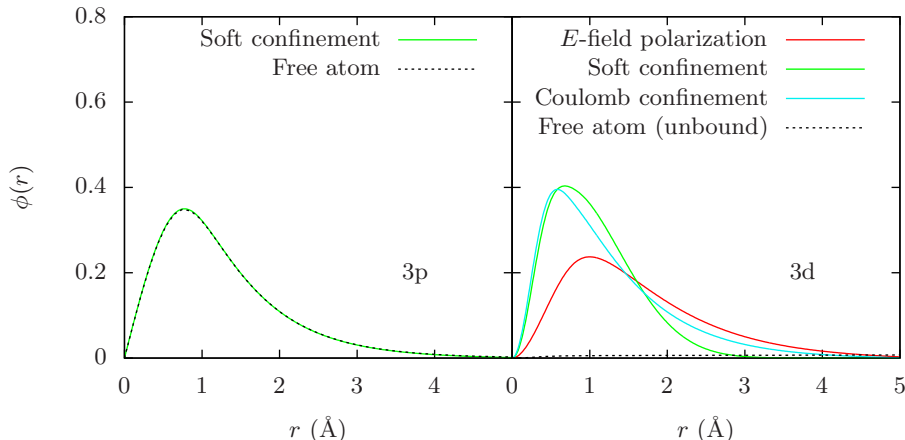


Figure 2. Shape of the 3p (occupied) and 3d (unoccupied) orbitals for Si. $r_c = 5.3$ Å both for the two confinement schemes and the E -field polarization orbital.

Making use of the Coulomb confinement scheme for the polarization orbitals, we have developed a series of 13 basis sets of increasing accuracy for the water molecule, ranging from 23 to 91 basis orbitals/molecule. Following the systematic convergence strategy proposed for correlation-consistent basis sets [29], the maximum size of our bases depends on the number of ζ orbitals used for the valence shells, decreasing by one for each additional polarization shell: $d\zeta + p$, $t\zeta + dp + p'$, $q\zeta + tp + dp' + p''$, and so on. We also provide intermediary bases with various subsets of the full number of polarization orbitals indicated by this scheme. We limit $l \leq 3$, and so the largest basis we consider is $q\zeta + tp + dp'$. For cases in which the number of ζ orbitals in the polarization shells differs between O and H, we adopt the following naming convention: the highest level of ζ is used, followed in brackets by the element for which it applies; the other element in such cases has one fewer shell. For example, $t\zeta + d_{(H)}p$ includes a double- ζ polarization shell for H, and a single- ζ polarization shell for O.

We do not here consider the optimization of r_c , but simply fix it to 4.5 Å for the double- and triple- ζ bases, and to 5.3 Å for the quadruple- ζ bases. We test the effect of varying the cutoff radius for the $q\zeta + dp$ basis, by decreasing r_c to 4.5 Å (the shorter basis is denoted as $^{(s)}q\zeta + dp$); the difference in energy is minimal (Table 1). Other parameters, including Q_0 and λ , are optimized variationally for the water monomer (see Supplementary Material for full specifications of all basis sets used). We also include in our results two previously proposed bases [15], obtained using the basis enthalpy optimization procedure [5] with a confining pressure of 0.2 GPa; for these bases (which we denote as $^{(P)}d\zeta + p$ and $^{(P)}t\zeta + p$) the value of r_c varies between shells, up to a maximum of 3.3 Å for $^{(P)}d\zeta + p$ and 3.7 Å for $^{(P)}t\zeta + p$.

2.1. General guidelines for generating a basis set

The many different options for defining the basis orbitals discussed and reviewed above are at the disposal of the user for generating good basis sets. It must be remembered, however, that there is no unique way of defining a basis, and that, as

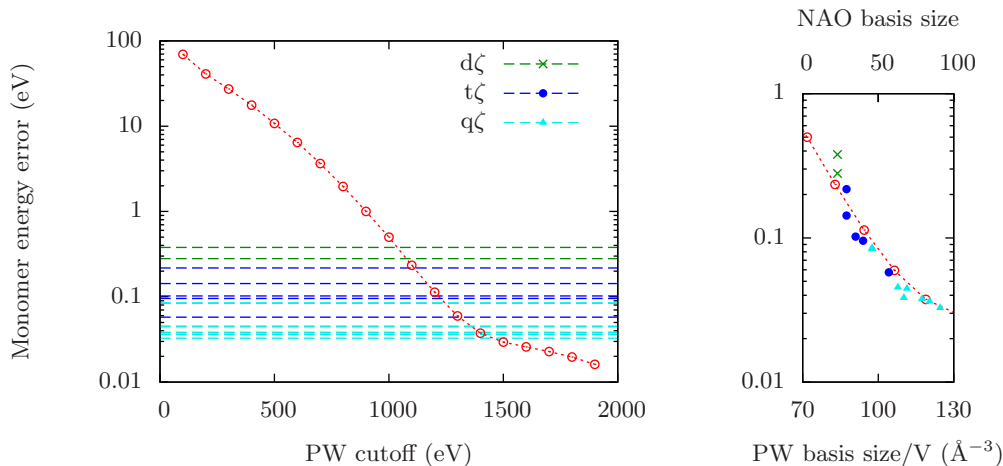


Figure 3. Convergence of the total energy for the monomer. The points indicate ABINIT PW calculations at different cutoffs, and the dashed horizontal lines show the values obtained with all SIESTA NAO basis sets. On the right we show the same results in a smaller range, with the SIESTA results now plotted against basis size, and the ABINIT results plotted against basis size per unit volume. All energy values are referenced to a PW calculation with extremely high cutoff (5000 eV).

for pseudopotentials, the responsibility of the choice of basis lies with the person performing the calculations. Methods like SIESTA are limited to using NAOs of finite range, but there is absolute freedom in the radial shapes, their range, the number of orbitals, the angular momentum values, and even the location (which is not restricted to be centred on an atom).

Here we present some guidelines for producing reasonable basis sets:

- Use as many basis functions as required for the accuracy needed, following the canonical hierarchy described above starting from the minimal single- ζ basis: $s\zeta$, $d\zeta + p$, $t\zeta + dp + p'$, $q\zeta + tp + dp' + p''$, and so on.
- Avoid too large or too small orbital ranges. Some procedures, like the one based on an energy shift [24], or blind variational optimization based on a single reference [5] can produce very short orbitals for light elements, inner orbitals, or cations. This severely limits the transferability of the basis. Never use orbitals with radii smaller than 3 Å. On the other hand, some procedures generate extremely large cutoff radii (e.g., in loosely bound states, as the 4s shell in transition metals). If they are used for bulk calculations, functions with radii larger than 5 Å can be quite useless (there are sufficient other functions in neighbouring atoms) but make the calculations substantially less efficient. SIESTA allows the user to establish both a minimum and a maximum cutoff radius globally.
- Use the soft confinement potential for smoothing the radial function close to r_c only, with $r_i \approx 0.9r_c$.
- For generating a polarization shell, use the Coulomb confinement potential described above (if not using the perturbative E -field polarization option). Q_0

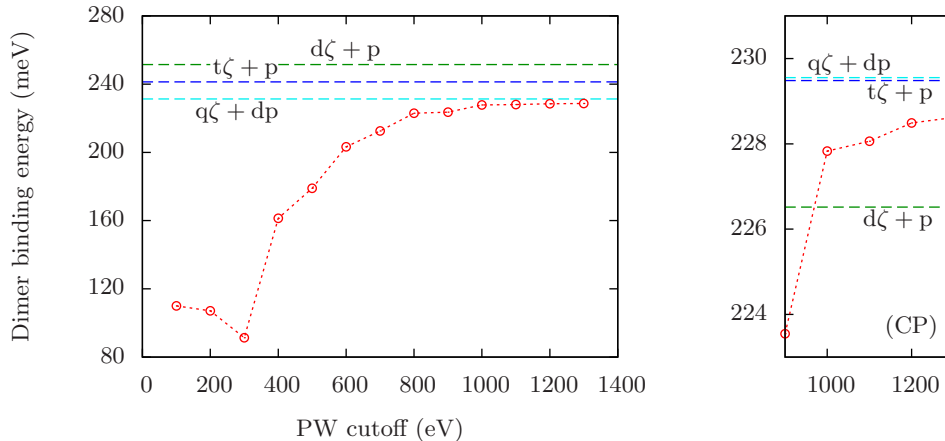


Figure 4. Convergence of the dimer binding energy. The points indicate ABINIT PW calculations at different cutoffs, and the dashed horizontal lines show the values obtained with three SIESTA NAO basis sets. On the right we show the same results in a smaller range, with the SIESTA results now including the counterpoise correction.

(and, optionally, λ) can be defined variationally on a set of representative simple systems, or by maximizing the overlap of the radial part of the polarization orbitals $\phi_{l'}(r)$ with the one of the shell to be polarized $\phi_l(r)$: $\int_0^{r_c} r^2 \phi_{l'}(r) \phi_l(r) dr$.

- Finally, the matching radii for the multiple- ζ orbitals can be defined variationally.

3. Results

3.1. Water monomer and dimer

We first show the convergence of the total energy for a single water molecule in a large box ($10 \times 10 \times 10 \text{ \AA}^3$). We perform a series of PW calculations of increasing kinetic energy cutoff, up to a maximum of 5000 eV (by which point the total energy is converged to < 1 meV). In Fig 3, we show how our NAO bases compare with the PW bases. Since the same pseudopotential is used in both sets of calculations, the total energies can be compared directly. The NAO bases show a smooth convergence with basis size, similar to that of PWs; it is tempting therefore to equate each NAO basis with the PW cutoff giving the same total energy, despite the fact that the Hilbert space spanned by them is very different. Nevertheless, the NAO bases undoubtedly achieve a good level of variational convergence, with the $q\zeta + tp + dp'$ basis reducing the error in the total energy to ~ 30 meV, as much as a 1460 eV PW cutoff. Even our smallest basis, $d\zeta + p$, is reasonably well converged: the total energy error is ~ 380 meV, equivalent to a 1080 eV cutoff. The error for all NAO basis sets is listed in Table 1.

There is a fairly large energy difference between the $^{(P)}d\zeta + p$ and $^{(P)}t\zeta + p$ bases and our newly parametrized ones of the same size (denoted simply $d\zeta + p$ and $t\zeta + p$). This we attribute both to the use of Coulomb confinement for the first polarization shell and the larger cutoff radii. In general, the convergence with basis size is almost

Table 1. Summary of results for the monomer and dimer. Listed are the basis sets tested, the basis size (number of orbitals/molecule), the total energy of the monomer (E , referenced to a PW calculation with 5000 eV cutoff), the binding energy of the dimer, first without and then with the counterpoise correction (E_b and E_b^{CP} , respectively), and the value of the correction. The PW binding energy listed is calculated with 1300 eV cutoff.

Basis set	Basis size	E (meV)	E_b (meV)	E_b^{CP} (meV)	CP corr. (meV)
$^{(P)}\text{d}\zeta + \text{p}$	23	378.61	266.38	241.16	-25.22
$\text{d}\zeta + \text{p}$	23	280.02	251.51	226.52	-24.99
$^{(P)}\text{t}\zeta + \text{p}$	29	217.98	249.42	239.89	-9.53
$\text{t}\zeta + \text{p}$	29	142.84	241.34	229.48	-11.86
$\text{t}\zeta + \text{d}_{(\text{H})}\text{p}$	35	102.14	232.01	226.80	-5.21
$\text{t}\zeta + \text{dp}$	40	95.59	230.68	225.96	-4.72
$\text{t}\zeta + \text{dp} + \text{p}'$	57	57.64	228.68	223.65	-5.03
$^{(s)}\text{q}\zeta + \text{dp}$	46	84.97	231.50	229.70	-1.80
$\text{q}\zeta + \text{dp}$	46	83.98	231.28	229.56	-1.72
$\text{q}\zeta + \text{dp} + \text{p}'$	63	45.37	227.99	226.25	-1.74
$\text{q}\zeta + \text{dp} + \text{p}' + \text{s}_{(\text{O})}\text{d}$	67	38.43	229.26	227.16	-2.10
$\text{q}\zeta + \text{t}_{(\text{H})}\text{p} + \text{p}'$	69	44.43	227.68	226.58	-1.10
$\text{q}\zeta + \text{t}_{(\text{H})}\text{p} + \text{d}_{(\text{H})}\text{p}'$	79	37.71	228.01	226.54	-1.47
$\text{q}\zeta + \text{tp} + \text{d}_{(\text{H})}\text{p}'$	84	36.17	228.10	226.68	-1.42
$\text{q}\zeta + \text{tp} + \text{dp}'$	91	32.64	228.63	227.12	-1.51
PW	-	0.00	228.62	-	-

monotonic; the only exception is the $\text{q}\zeta + \text{dp} + \text{p}' + \text{s}_{(\text{O})}\text{d}$ basis, for which we add a set of diffuse orbitals of s and p symmetry onto the O ion in order to capture the swelling due to its anionic character. This basis adds four new orbitals onto $\text{q}\zeta + \text{dp} + \text{p}'$, yet gains 6 meV more energy than $\text{q}\zeta + \text{t}_{(\text{H})}\text{p} + \text{p}'$, which adds six. Such small energy differences, however, do not significantly alter the overall convergence behaviour.

Fig. 4 shows the PW convergence and values for three selected NAO bases of the binding energy E_b of the water dimer in a fixed geometry (from Supplementary Material of Ref. [30]). The results for all NAO bases are given in Table 1. In order to ensure that periodic image interactions are negligible, we use a $25 \times 25 \times 25 \text{ \AA}$ box. Here, the difference between NAO and PW bases is clear, as the former converges from above and the latter from below. All our quadruple- ζ bases give an error in E_b within 2–3 meV of the converged PW result. For the double- and triple- ζ bases, basis set superposition error (BSSE) accounts for most of the total error. The addition of a CP correction term to E_b brings all bases to within the same level of precision of the quadruple- ζ ones; only $^{(P)}\text{d}\zeta + \text{p}$ and $^{(P)}\text{t}\zeta + \text{p}$ are somewhat less precise (with errors of 10–13 meV), despite having a smaller CP correction due to their shorter range.

3.2. Ice phases: I_c and VIII

We now turn to the first test of our bases in bulk water systems, by focussing on two of the ordered phases of ice: cubic ice I_c and high-density tetragonal ice VIII. The study of ice by first principles DFT methods has progressed considerably in recent years [31, 32, 33, 34, 35], but understanding in detail the relative energetic contributions that give rise to its phase diagram and the properties of the various phases remains a challenging problem of current interest. Typically for such studies,

Table 2. Summary of results for ice. Listed are total energies/molecule (meV) for the two phases considered and the energy difference/molecule (meV) between these phases for seven different geometries, obtained by relaxing with six NAO basis sets, and a PW relaxation with 4500 eV cutoff. For each geometry we give the energy calculated with SIESTA using the same NAO basis for which the relaxation was performed, and with ABINIT at 4500 eV cutoff (independently of geometry). Note that the total energies of the two phases are given referenced to their respective PW results.

Phase	Basis	Geometry						PW
		^(P) d ζ + p	d ζ + p	^(P) t ζ + p	t ζ + p	q ζ + dp	q ζ + tp + dp'	
I _c	NAO	295.93	197.35	148.34	107.68	68.70	25.29	-
	PW	0.78	2.73	2.11	0.99	0.55	0.11	0.00
VIII	NAO	259.76	166.58	161.32	111.92	73.89	26.86	-
	PW	1.49	2.88	1.95	1.04	0.53	0.17	0.00
Diff.	NAO	-142.58	-147.99	-191.73	-182.99	-183.95	-180.32	-
	PW	-179.46	-178.90	-178.59	-178.80	-178.73	-178.82	-178.75
	Ref. [35]	-	-	-	-	-	-	-178

it is desirable to calculate energy differences between configurations to within a few meV/molecule. To this end, we test the accuracy of the NAO bases both in terms of the relaxation of the ionic positions and the final energy difference between the two phases under consideration.

For the calculations, we fix the volume/molecule to 30.51 Å³ for ice I_c and 20.45 Å³ for ice VIII, using the equilibrium volumes reported by Murray and Galli [35] for PBE at ambient pressure, and the c/a ratio to 1.44 for ice VIII. We use a 4 × 4 × 4 Monkhorst-Pack (MP) k-point grid [36] for both unit cells. First, we perform PW cutoff convergence tests for the two ice phases up to a 5000 eV cutoff; we find that 4500 eV gives extremely accurate results (within 0.1 meV for total energies, 0.5 meV/Å for ionic forces, and 0.05 kbar for pressures), and use this value for all PW results given in this section.

We choose four representative NAO basis sets for testing from those presented in the previous section; additionally, the two old parametrizations for double- and triple- ζ are also tested. The ice unit cells are relaxed with respect to the ionic positions for these six bases, and independently for PWs, using a maximum force tolerance of 10 meV/Å. Finally, we recalculate the energy of the system with PWs using the six geometries obtained from the NAO bases. All results are given in Table 2.

The PW results for the different geometries show that all NAO bases give sufficiently accurate ionic relaxations. The errors in the PW energy difference/molecule between the two phases using the NAO-relaxed geometries, as compared with the PW-relaxed ones, are less than a meV in all cases, ranging from 0.71 meV for the ^(P)d ζ + p geometries to only 0.02 meV for the q ζ + dp ones. The errors in total energies are similarly small, less than 3 meV for all geometries. The accuracy of the relaxations is also confirmed by examining the bond lengths, with the largest errors found for the ^(P)d ζ + p geometries (up to 21 mÅ), and the smallest for the q ζ + tp + dp' ones (up to 3 mÅ).

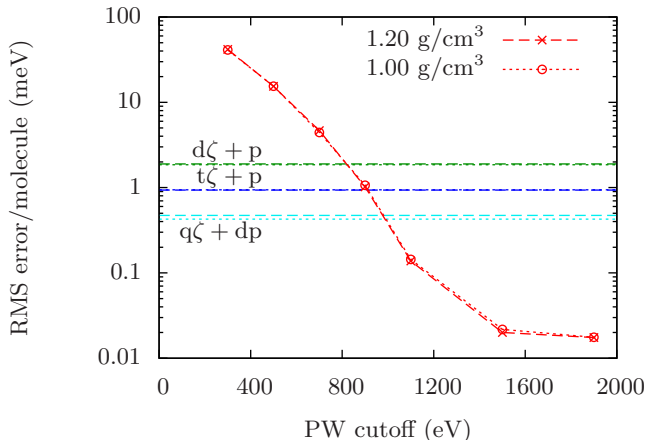


Figure 5. RMS error in energy differences/molecule for 100 snapshots at each density. The points indicate PW calculations at different cutoffs, and the dashed horizontal lines show the value for three NAO basis sets (long dashes for 1.20 g/cm³ density and short dashes for 1.00 g/cm³ density).

We now consider not only the geometries, but also the energies calculated with the NAO bases. As should be expected, total energies/molecule of the ice phases have errors on the order of those reported in Table 1 for the total energy of the monomer; however, it is interesting to note that they are systematically smaller, resulting from the fact that we are now considering a condensed phase, thus enabling the NAOs to represent the charge density everywhere in the unit cell.

The most important results are for the energy differences/molecule between phases given by the NAO calculations. $d\zeta + p$ gives a fairly substantial error (~ 30 meV), while both $t\zeta + p$ and $q\zeta + dp$ are in good agreement with the converged PW result, with errors of ~ 5 meV. The most accurate basis, $q\zeta + tp + dp'$, succeeds in reducing the error to less than 2 meV. As with the dimer binding energy, we find our new bases to give a substantial improvement with respect to the old parametrizations, especially in this case for $t\zeta + p$.

3.3. Liquid water

Assessing the accuracy of our bases for water in its liquid state is more challenging than for the solid state, as the calculation of any quantity of interest will involve a statistical average over a MD trajectory. Performing a full AIMD simulation using a tightly converged PW cutoff would be computationally very expensive, and, furthermore, would not allow for a detailed comparison with the results obtained for the same simulation using one of the NAO basis sets, as the trajectories would quickly decorrelate.

Here we employ a simple alternative method for testing basis sets for liquid systems, which can be used routinely to obtain accurate estimates of the errors in quantities such as total energies, energy differences, ionic forces and cell pressures, as well as for direct parametrization of the basis set using the same procedures proposed for fitting classical force fields from *ab initio* data [37]. We first employ the

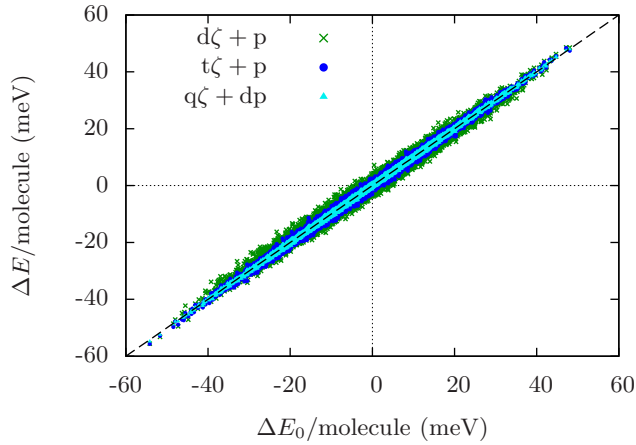


Figure 6. Scatter plot comparing energy differences/molecule between all snapshots obtained with ABINIT PW calculations at 2700 eV cutoff (ΔE_0) and SIESTA calculations with three different NAO basis sets.

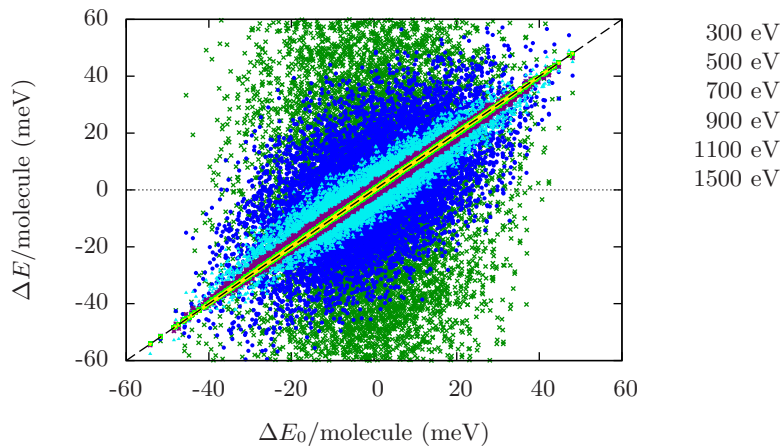


Figure 7. Scatter plot comparing energy differences/molecule between all snapshots obtained with ABINIT PW calculations at 2700 eV cutoff (ΔE_0) and lower cutoffs.

GROMACS [38] code to perform two long 1 ns MD runs of a box of 32 water molecules using the TIP4P force field [39], at equilibrium density (1.00 g/cm^3) and high density (1.20 g/cm^3). We select 100 random snapshots from each run, which we use as our testing set; the long MD run ensures that they are uncorrelated. We include the high density run to ensure that we are sampling sufficient configurations with occupied interstitial anti-tetrahedral sites between the first and second coordination shells, as the correct description of these configurations is important for reproducing several key properties of the liquid [15]. Finally, we perform single-point DFT calculations of the 200 snapshots, using the various NAO bases in SIESTA and PWs at different

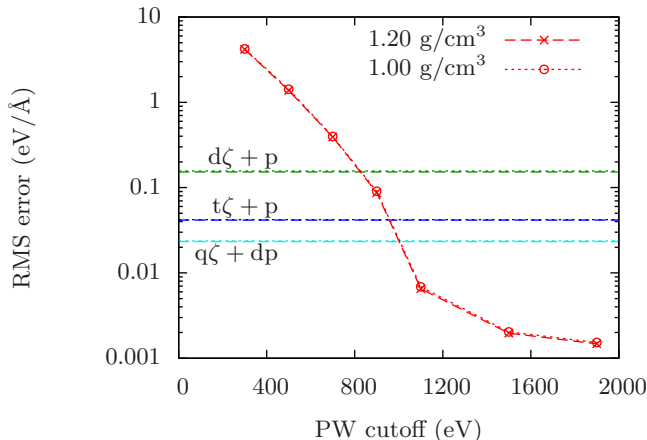


Figure 8. RMS error in the magnitude of the forces on all ions for 100 snapshots at each density. The points indicate PW calculations at different cutoffs, and the dashed horizontal lines show the value for three NAO basis sets.

cutoffs in ABINIT. In this case we choose 2700 eV as our ‘converged’ PW cutoff value for comparing to all other bases. Cutoff convergence tests on a few snapshots up to 5000 eV show this value to give total energies to within 1 meV, ionic forces to within 0.6 meV/Å, and pressures to within 0.5 kbar.

We perform our tests on five NAO bases, $^{(P)}d\zeta + p$, $d\zeta + p$, $^{(P)}t\zeta + p$, $t\zeta + p$, and $q\zeta + dp$; all results are given in Table 3. Firstly, we examine the error in total energies/molecule for all the snapshots. As expected, calculating the average shift in total energy for all snapshots gives a similar convergence to that discussed previously for the water monomer and the two ice phases. One point of interest is that the average shifts calculated for the two densities become closer to each other as the quality of the basis increases, anticipating the fact that pressure values also become more precise.

The best quantitative estimator of overall accuracy for a fixed-volume NVE AIMD simulation is the energy difference between snapshots. For this, we calculate the root mean square (RMS) error in our test set; the results are shown in Fig. 5, compared to PW calculations of increasing cutoff. There is a negligible difference between the two densities, with the $d\zeta + p$, $t\zeta + p$, and $q\zeta + dp$ bases giving RMS errors equivalent to PW cutoffs of 820 eV, 910 eV, and 980 eV, respectively. This is similar to the errors in the dimer binding energy before applying the CP correction.

The scatter plots in Figs. 6 and 7 show these results in more detail, by comparing either a NAO or a lower-cutoff PW basis with the converged PW basis, and plotting the energy difference between each pair of snapshots; both densities are included in the plot. Low-accuracy PW calculations give a large scatter of results, while all the NAO bases considered give a tight clustering around the diagonal, as do the higher PW cutoffs. Indeed, the NAO bases show no spurious outliers at all, and, equally importantly, no obvious systematic trend away from the diagonal that might cause a significant difference in the region of configuration space explored during a MD run.

Fig. 8 shows the RMS error in ionic forces. The results are very similar to those already discussed for energy differences, with the three new NAO bases giving errors

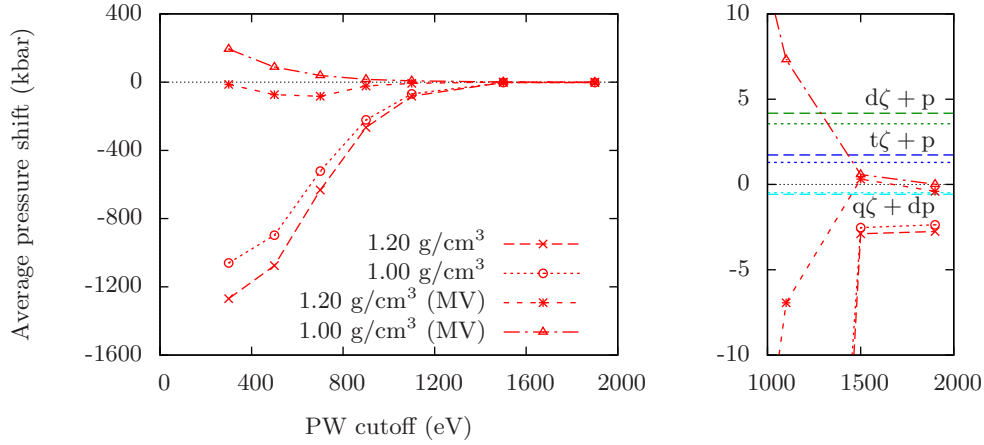


Figure 9. The average shift in pressure for 100 snapshots at each density. The points indicate PW calculations at different cutoffs (both with and without the MV correction for finite PW cutoff), and the dashed horizontal lines show the value for three NAO basis sets with no correction applied.

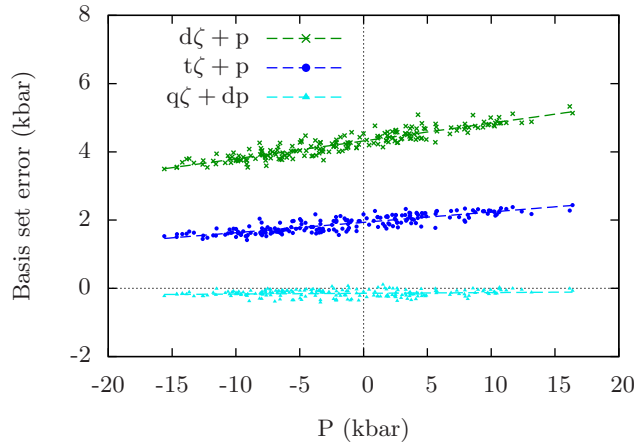


Figure 10. Error in NAO basis set pressures with respect to converged PW pressures for all 200 snapshots. The dashed lines show the linear fits used for the basis set correction in Fig. 11 and Table 3.

equivalent to PW cutoffs of 830 eV, 960 eV, and 1000 eV, respectively. Differences between O and H ions are negligible (see Table 3). The table also reports results for the RMS angle between the forces obtained with the NAO basis and the converged PW one; these range from $\sim 14^\circ$ for the $d\zeta + p$ basis to $\sim 1^\circ$ for the $q\zeta + dp$ one. It is noteworthy that the $(^P)d\zeta + p$ basis performs better than the newly parametrized version in terms of force magnitudes and angles; this explains the relatively accurate relaxed geometries obtained for ice (see Table 2).

Finally, we examine the accuracy of our bases for calculating cell pressures. It is

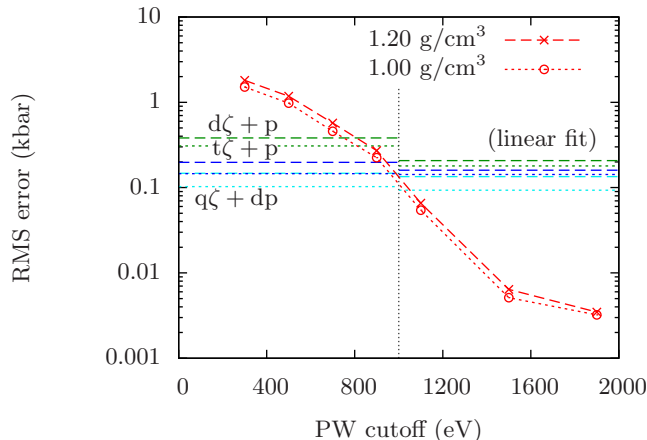


Figure 11. RMS error in differences in pressures between 100 snapshots at each density. The points indicate PW calculations at different cutoffs, and the dashed horizontal lines show the value for three NAO basis sets (uncorrected, on the left side of the plot, and with the correction explained in the text on the right side).

well-known that PW calculations suffer in this respect from the fact that the kinetic energy cutoff effectively changes with an infinitesimal change in volume (since the calculation of stresses is implicitly performed at constant PW number), leading to a spurious tensile stress. In order to eliminate this, the correction by Meade and Vanderbilt [40] (here referred to as MV) is generally applied. The MV correction is given by

$$\frac{2}{3V} \frac{\partial E(E^{\text{cut}}, V)}{\partial \ln E^{\text{cut}}}, \quad (2)$$

where E is the total energy, V the cell volume, and E^{cut} the kinetic energy cutoff. In contrast, the only systematic error introduced by NAO basis sets in the calculation of stresses is that from BSSE.

Fig. 9 shows the average shift in pressure with respect to the converged calculations. Even after applying the MV correction term, the PW calculations give very large shifts, on the order of 10–100 kbar. The NAO shifts, instead, are of no more than 4 kbar (for the $d\zeta + p$ basis). The PW convergence also varies significantly with density, which is not the case for the NAOs.

We can use our results to define a simple and effective correction for cell pressures calculated with NAOs. This is illustrated in Fig. 10, showing a scatter plot of NAO pressures with respect to the converged PW result (note that we are plotting the error in the NAO value against the absolute value from PWs). Two trends can be clearly observed: a constant shift to higher pressures (i.e., the same result as shown in Fig. 9), and an increase in the stiffness of the system; both of these reduce with basis size, and are eliminated almost completely for the $q\zeta + dp$ basis. Therefore, by performing a linear fit to the data points for a given basis we obtain an expression for correcting pressure values in a large range. It is important to note that the figure includes data points at both densities, which show the same trend; the correction can therefore also be used independently of density, at least within the range considered.

Table 3. Summary of results for liquid water. Listed are the average shift/RMS error for all snapshots at each density for various quantities: total energies (E), total pressures (P), energy differences (ΔE), pressure differences (Δ), magnitudes of ionic forces ($|\mathbf{F}|$), and angles between the SIESTA NAO force (\mathbf{F}) and the converged PW force (\mathbf{F}_0). In all cases the reference values are taken from ABINIT PW calculations at 2700 eV cutoff.

Average shift		ρ (g/cm ³)	Basis set				
			^(P) d ζ + p	d ζ + p	^(P) t ζ + p	t ζ + p	q ζ + dp
E /molecule (eV)		1.20	0.28	0.19	0.15	0.12	0.08
		1.00	0.29	0.21	0.16	0.12	0.08
P (kbar)	uncorr.	1.20	0.06	4.18	2.48	1.73	-0.57
		1.00	-0.89	3.56	1.99	1.30	-0.48
	corr.	1.20	-0.27	-0.55	-0.43	-0.37	-0.43
		1.00	-0.48	-0.65	-0.53	-0.50	-0.32
RMS error							
ΔE /molecule (meV)		1.20	1.86	1.90	1.32	0.93	0.47
		1.00	1.54	1.85	1.44	0.94	0.43
ΔP (kbar)	uncorr.	1.20	0.41	0.38	0.27	0.20	0.15
		1.00	0.39	0.31	0.22	0.15	0.10
	corr.	1.20	0.33	0.21	0.14	0.16	0.13
		1.00	0.24	0.18	0.14	0.14	0.09
$ \mathbf{F} $ (eV/Å)	O	1.20	0.11	0.17	0.07	0.05	0.03
		1.00	0.11	0.17	0.07	0.05	0.03
	H	1.20	0.07	0.15	0.09	0.04	0.02
		1.00	0.07	0.15	0.09	0.04	0.02
$\cos^{-1} \left\{ \frac{\mathbf{F} \cdot \mathbf{F}_0}{ \mathbf{F} \mathbf{F}_0 } \right\}$ (°)	O	1.20	8.27	13.83	5.41	4.33	0.98
		1.00	7.56	13.40	5.00	4.33	1.04
	H	1.20	7.57	15.11	8.91	3.88	1.77
		1.00	7.34	15.01	8.77	4.07	1.67

The effect of our correction for NAO calculations is given in Table 3 for average pressure shifts and RMS errors in pressure differences. The correction reduces the average shift to very small values (~ 0.5 kbar) for all bases, the remaining error being due to the small discrepancy between the uncorrected shifts at the two different densities shown in Fig. 9. Pressure differences between snapshots are generally of less interest for MD simulations than absolute values; nevertheless, they can be used as an additional measure of the quality of the basis. The comparison with PWs is shown in Fig. 11, both for the uncorrected and corrected NAO results (the MV correction for PWs has no effect on pressure differences). The three uncorrected NAO bases give errors equivalent to PW cutoffs of approximately 810 eV, 950 eV, and 1000 eV, respectively, similarly to the values reported previously for other quantities of interest. After the correction is applied, the error for all bases is reduced to ~ 0.2 kbar. Once again, we note the peculiarity of the ^(P)d ζ + p basis (used in Ref. [41]), which features a very small average pressure shift without the need for a correction, while the RMS error in pressure differences follows the trend of the other bases.

4. Conclusions

In this paper, we have described the development and testing of finite-range NAO basis sets for water-based systems. We have discussed the general strategy employed for creating basis sets of increasing size and accuracy, and have proposed the use of a screened Coulomb confinement potential for shaping the polarization orbitals, in order to achieve a good agreement with the higher angular momentum shells of the free atom without needing to extend the confinement radius beyond what is physically useful for condensed matter systems.

We have presented 13 new bases for the water molecule, ranging from $d\zeta + p$ to $q\zeta + tp + dp'$. The full list of parameters needed to recreate and use them are given in the Supplementary Material (we provide these instead of the numerical radial functions themselves so as to minimize the dependence on specific pseudopotentials).

In order to perform rigorous tests of the accuracy of our bases, we use auxiliary PW calculations at different kinetic energy cutoffs. We can therefore compare the results obtained with NAOs with the PW ones at a very large (essentially fully converged) cutoff, and also equate the accuracy of the various NAO bases with PW bases at specific cutoffs. This is done using two different DFT codes (SIESTA for NAOs, ABINIT for PWs) with the same pseudopotentials and Kleinman-Bylander factorization; even so, small algorithmic differences between the codes will affect the comparison with the NAO bases of highest accuracy, slightly underestimating their performance with respect to PWs for quantities such as energy differences.

The results for our tests on a variety of molecular and condensed systems show the transferability and accuracy of the bases. In particular, there is a good level of consistency both between different systems and different properties of the same system when comparing to the performance of PW calculations at finite cutoff: errors in total energies for the $d\zeta + p$, $t\zeta + p$, and $q\zeta + dp$ bases are on the order of those for cutoffs of 1100 eV, 1200 eV, and 1300 eV, respectively, while errors in energy differences, ionic forces, and pressure differences are on the order of those for cutoffs of 800 eV, 900 eV, and 1000 eV. However, it is important to remember that the two types of bases are not at all equivalent: this is clearly demonstrated by calculations of absolute pressure, which show NAOs to naturally give much smaller errors for this quantity than PWs. We have also proposed a simple correction for further reducing errors in absolute pressures and pressure differences for NAO calculations, based on a linear fit to the data obtained from 200 liquid water snapshots at two densities; this can be used, e.g., to correct average pressures obtained from NVE AIMD simulations.

Acknowledgments

This work was partly funded by grants FIS2009-12721 and FIS2012-37549 from the Spanish Ministry of Science. MVFS acknowledges a DOE Early Career Award No. de-sc0003871. We thank Javier Junquera for support with the translation of pseudopotentials between SIESTA and ABINIT. The calculations were performed on the following HPC clusters: *kroketa* (CIC nanoGUNE, Spain), *arina* (Universidad del País Vasco/Euskal Herriko Unibertsitatea, Spain), *magerit* (CeSViMa, Universidad Politécnica de Madrid, Spain). We thank the RES-Red Española de Supercomputación for access to *magerit*. SGIker (UPV/EHU, MICINN, GV/EJ, ERDF and ESF) support is gratefully acknowledged.

References

- [1] Bowler D R and Miyazaki T 2012 *Rep. Prog. Phys.* **75** 036503
- [2] Sankey O F and Niklewski D J 1989 *Phys. Rev. B* **40** 3979
- [3] Kenny S D, Horsfield A P and Fujitani H 2000 *Phys. Rev. B* **62** 4899
- [4] Junquera J, Paz O, Sánchez-Portal D and Artacho E 2001 *Phys. Rev. B* **64** 235111
- [5] Anglada E, Soler J M, Junquera J and Artacho E 2002 *Phys. Rev. B* **66** 205101
- [6] Ozaki T 2003 *Phys. Rev. B* **67** 155108
- [7] Ozaki T and Kino H 2004 *Phys. Rev. B* **69** 195113
- [8] García-Gil S, García A, Lorente N and Ordejón P 2009 *Phys. Rev. B* **79** 075441
- [9] Payne M C, Arias T A and Joannopoulos J D 1992 *Rev. Mod. Phys.* **64** 1045
- [10] Chelikowsky J R, Troullier N and Saad Y 1994 *Phys. Rev. Lett.* **72** 1240
- [11] Briggs E L, Sullivan D J and Bernholc J 1995 *Phys. Rev. B* **52** 5471
- [12] Chen J Y and Yoo C S 2011 *Proc. Natl. Acad. Sci. USA* **108** 7685
- [13] Robinson G W, Zhu S B, Singh S and Evans M W 1996 *Water in Biology, Chemistry and Physics: Experimental Overviews and Computational Methodologies* (Singapore: World Scientific)
- [14] Errington J R and Debenedetti P G 2001 *Nature* **409** 318
- [15] Wang J, Román-Pérez G, Soler J M, Artacho E and Fernández-Serra M V 2011 *J. Chem. Phys.* **134** 024516
- [16] Zhang C, Donadio D, Gygi F and Galli G 2011 *J. Chem. Theory Comput.* **7** 1443
- [17] Soler J M, Artacho E, Gale J D, García A, Junquera J, Ordejón P and Sánchez-Portal D 2002 *J. Phys.: Condens. Matter* **14** 2745
- [18] Gonze X *et al* 2009 *Comput. Phys. Commun.* **180** 2582
- [19] Perdew J P, Burke K and Ernzerhof M 1996 *Phys. Rev. Lett.* **77** 3865
- [20] Troullier N and Martins J L 1991 *Phys. Rev. B* **43** 1993
- [21] Louie S G, Froyen S and Cohen M L 1982 *Phys. Rev. B* **26** 1738
- [22] Kleinman L and Bylander D M 1982 *Phys. Rev. Lett.* **48** 1425
- [23] Verstraete M J, Junquera J, Gonze X and García A *in preparation*
- [24] Artacho E, Sánchez-Portal D, Ordejón P, García A and Soler J M 1999 *Phys. Status Solidi (b)* **215** 809–817
- [25] Huzinaga S, Andzelm J, Klobukowski M, Radzio-Andzelm E, Sakai Y and Tatewaki H 1984 *Gaussian Basis Sets for Molecular Calculations* (Amsterdam/New York: Elsevier)
- [26] Poirier R, Kari R and Csizmadia I G 1985 *Handbook of Gaussian Basis Sets: A Compendium for Ab-Initio Molecular Orbital Calculations* (Amsterdam/New York: Elsevier)
- [27] Jansen H B and Ros P 1969 *Chem. Phys. Lett.* **3** 140
- [28] Boys S F and Bernardi F 1970 *Mol. Phys.* **19** 553
- [29] Dunning T H 1989 *J. Chem. Phys.* **90** 1007
- [30] Santra B, Michaelides A and Scheffler M 2007 *J. Chem. Phys.* **127** 184104
- [31] Militzer B and Wilson H F 2010 *Phys. Rev. Lett.* **105** 195701
- [32] Hamada I 2010 *J. Chem. Phys.* **133** 214503
- [33] Hermann A and Schwerdtfeger P 2011 *Phys. Rev. Lett.* **106** 187403
- [34] Santra B, Klimeš J, Alfè D, Tkatchenko A, Slater B, Michaelides A, Car R and Scheffler M 2011 *Phys. Rev. Lett.* **107** 185701
- [35] Murray E D and Galli G 2012 *Phys. Rev. Lett.* **108** 105502
- [36] Monkhorst H J and Pack J D 1976 *Phys. Rev. B* **13** 5188
- [37] Tangney P and Scandolo S 2002 *J. Chem. Phys.* **117** 8898
- [38] Berendsen H J C, van der Spoel D and van Drunen R 1995 *Comput. Phys. Commun.* **91** 43
- [39] Jorgensen W L, Chandrasekhar J, Madura J D, Impey R W and Klein M L 1983 *J. Chem. Phys.* **79** 926
- [40] Meade R D and Vanderbilt D 1989 *Mat. Res. Soc. Symp. Proc.* **141** 451
- [41] Corsetti F, Artacho E, Soler J M, Alexandre S S and Fernández-Serra M V 2013 [arXiv:1307.1645](https://arxiv.org/abs/1307.1645) [physics.chem-ph]