



## UvA-DARE (Digital Academic Repository)

### Transferable basis sets of numerical atomic orbitals

Louwerse, M.J.; Rothenberg, G.

**DOI**

[10.1103/PhysRevB.85.035108](https://doi.org/10.1103/PhysRevB.85.035108)

**Publication date**

2012

**Document Version**

Final published version

**Published in**

Physical Review B

[Link to publication](#)

**Citation for published version (APA):**

Louwerse, M. J., & Rothenberg, G. (2012). Transferable basis sets of numerical atomic orbitals. *Physical Review B*, *85*(3), Article 035108.  
<https://doi.org/10.1103/PhysRevB.85.035108>

**General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

## Transferable basis sets of numerical atomic orbitals

Manuel J. Louwrese\* and Gadi Rothenberg

*Van 't Hoff Institute for Molecular Sciences, University of Amsterdam, Postbus 94720 GS Amsterdam, The Netherlands*

(Received 27 May 2011; revised manuscript received 12 October 2011; published 12 January 2012)

We show that numerical atomic orbital basis sets that are variationally optimized for specific compounds are biased and not fully transferable to other compounds. The situation improves when the basis sets are optimized for several compounds and a compromise is made. We demonstrate this for basis sets for compounds containing H, O, N, Al, Cl, and Cu atoms, using the SIESTA code. Reaction energies are compared to benchmark calculations with the ADF code. Our double-zeta basis sets give a root mean square error of 0.2 eV. With triple-zeta basis sets, we obtain a root mean square error of only 0.1 eV. Finally, for the case of mixed phase systems, we propose that optimizing basis sets for gaseous molecules and then using them also for bulk compounds, is better than the other way around.

DOI: [10.1103/PhysRevB.85.035108](https://doi.org/10.1103/PhysRevB.85.035108)

PACS number(s): 71.15.Ap, 31.15.-p

### I. INTRODUCTION

In quantum chemistry and density functional theory (DFT), basis sets play an essential role. Increasing the number of basis set orbitals in a calculation will always improve the accuracy but also increase the cost. Generally, there are two approaches for constructing basis sets: either use many basis functions of a mathematical shape that is efficient for calculation, or use fewer functions that fit the atomic and molecular orbitals well, but are harder to calculate. Going from one extreme to the other, we mention plane waves, Gaussian-type atomic orbitals,<sup>1</sup> Slater-type atomic orbitals,<sup>2</sup> and numerical atomic orbitals (NAOs).<sup>3</sup> Other methods for increasing the efficiency in DFT codes include frozen core electrons, pseudopotentials,<sup>4,5</sup> projector-augmented waves,<sup>6</sup> and spatial confinement of the basis set orbitals. The well-known SIESTA package<sup>7,8</sup> uses a combination of NAOs, pseudopotentials, a strict spatial confinement, and linear scaling techniques. This gives one of the fastest algorithms to date.

NAOs are extremely flexible. SIESTA offers an adapted split-valence scheme for the basis set expansion, and an automatic method is available for finding parameters for this scheme for all atom types.<sup>9</sup> However, basis sets with parameters that are variationally optimized separately for each atom type perform better.<sup>3,10</sup> In practice, this optimization is usually done on the chemical structure of interest.<sup>10,11</sup>

The problem is that optimizing a basis set for a specific system causes a bias. When an energy difference is calculated between two structures, but the basis set is optimized for only one of them, the energy will be biased towards that one. To solve this problem, we test the transferability of optimized basis sets to other systems. We propose that it is better to optimize the basis sets on several systems and then make a compromise to improve the transferability. A similar approach was followed by Ozaki and Kino, but using a totally different type of basis orbitals.<sup>12</sup> Some other interesting reading on transferability of basis sets can be found in Ref. 13.

Our research is directed at catalysis applications. Specifically, we focus on the interaction of water, copper chloride, and copper nitrate on alumina surfaces. Thus, the transferability of the oxygen basis set between the structures of water, alumina, and nitrates are important for us, as well as the transferability of the copper basis set between different oxidation states. As our

research focuses on catalyst surfaces and gaseous reactants, we prefer optimizing our basis sets on gaseous molecules and not on bulk condensed systems.

For completeness, we include a short section on how basis sets are constructed in SIESTA, thus defining the parameters used in this work.

### II. BASIS SETS IN SIESTA

The basis set orbitals in SIESTA are constructed as follows. First of all, the core electrons are kept frozen and are replaced by pseudopotentials, which are added to the Hamiltonian. The Schrödinger equation with the new pseudo-Hamiltonian is generally referred to as the pseudoatomic problem and its solutions as pseudoatomic orbitals (PAOs). For each angular momentum ( $s$ ,  $p$ ,  $d$ ,  $f$ ), the first basis set orbital is equal to the PAO of that shell, up to a user-defined cutoff radius,  $r_c$ . SIESTA offers two cutoff types, both performed by adding an additional potential to the pseudoatomic problem: the infinite square-well potential of Sankey *et al.*<sup>14</sup> or the soft confinement potential of Junquera *et al.*<sup>3</sup> For energy calculations, as in this work, both methods perform comparably.<sup>3</sup> As the latter method adds two more parameters that need to be optimized for each basis set orbital, we used the default Sankey-type cutoffs.

Additional basis set orbitals, used for double- $\zeta$  (DZ) and triple- $\zeta$  (TZ) sets, are constructed using a matching radius,  $r_m$ , shorter than  $r_c$ . A temporary function is defined, that outside  $r_m$  is identical to the PAO, but inside  $r_m$  forms a smooth function toward zero at  $r = 0$ . The multiple- $\zeta$  orbitals are calculated as the difference between this smooth function and the PAO, giving orbitals that are exactly zero outside their  $r_m$ .<sup>3</sup> Thus, an atom with  $s$  and  $p$  electrons and a DZ basis set needs four parameters: one  $r_c$  for the first  $s$  orbital, one  $r_m$  for the second  $s$  orbital, one  $r_c$  for the first  $p$  orbital, and one  $r_m$  for the second  $p$  orbital.

Furthermore, usually polarization orbitals are included, i.e., orbitals with a higher angular momentum than the electrons present in the pseudoatom. For pseudoatoms with  $s$  and  $p$  electrons, we use polarization orbitals with  $d$  symmetry. For pseudoatoms with  $s$  and  $d$  electrons, we use polarization orbitals with  $p$  (polarization of  $s$ ) and  $f$  (polarization of  $d$ ) symmetry. In SIESTA, the polarization orbitals are constructed from the actual polarization of the pseudoatom in the presence

of a small electric field.<sup>9</sup> These polarized pseudoatomic calculations use the same cutoff potential with the same  $r_c$  as the nonpolarized orbitals. When multiple polarization orbitals are used (e.g. in a TZ2P basis set), for the second polarization orbital again an  $r_m$  must be set, and a smooth function is constructed now based on the polarized PAO.

All cutoff and matching radii,  $r_c$  and  $r_m$ , can be set manually or automatically. SIESTA uses two parameters, the energy shift and the split norm, that allow setting all  $r_c$  and  $r_m$  values for all atom types at once. The energy shift sets the  $r_c$  values and is the energy raise experienced by the PAO as result of the cutoff. Small energy shift values give large cutoff radii. The split norm sets the  $r_m$  values and is the normalized integrated value of the PAO outside  $r_m$ .<sup>9</sup> In the following, *default settings* denotes that the energy shift and split norm parameters are used, while *optimized* basis sets means that all cutoff and matching radii are optimized independently.

Lastly, it is possible to calculate the PAOs with a potentially noninteger net charge on the pseudoatom. This contracts or enlarges the resulting orbitals, making them more fit for (partially) charged atoms. However, we expect this to hamper the transferability of the basis sets, so this option was not used here.

### III. METHODS

All calculations were performed with SIESTA\_2.0.2<sup>7,8</sup> using the rPBE functional.<sup>15</sup> We used improved Troullier-Martins pseudopotentials<sup>5</sup> with parameters taken from the Octopus project.<sup>16</sup> Scalar relativistic corrections were applied for all atom types. For aluminum and copper, core corrections<sup>17</sup> were included with pseudocore radii of 0.92 bohr and 0.84 bohr, respectively. For the density mesh a cutoff was used at 200 Ry. Electronic states were calculated by conventional diagonalization. Gaseous compounds were calculated in a cubic box of 20 Å in length. Atomic positions were optimized to a precision of 0.01 eV/Å. All copper compounds were calculated in spin state  $S = 1$ .

We variationally optimized basis sets for H, O, N, Cl, Cu, and Al atoms. The  $r_c$  and  $r_m$  radii were varied with steps of 0.1 bohr with a maximum value for  $r_c$  of 9.0 bohr (for  $\text{Cu}^{2+}$  a minimum of 5.0 bohr was enforced for the  $s$ - $r_c$ ). The pseudoatomic charge was not optimized but set to zero in all cases. We used Sankey-type cutoffs. We constructed both DZP and TZP basis sets. For aluminum we found, though, that a second polarization orbital gave a large improvement. Therefore, for aluminum instead of TZP a TZ2P set was used. The matching radius for this second polarization orbital was set via the split-norm parameter, which was optimized with steps of 0.05.

We optimized the basis sets for in total 19 different gaseous molecules, radicals, and ions:  $\text{H}_2$ ,  $\text{H}^\bullet$ ,  $\text{H}_2\text{O}$ ,  $\text{OH}^-$ ,  $\text{OH}^\bullet$ ,  $\text{NO}_2$ ,  $\text{NO}_3^-$ ,  $\text{HNO}_3$ ,  $\text{N}_2$ ,  $\text{HCl}$ ,  $\text{Cl}^-$ ,  $\text{Cu}^0$ ,  $\text{Cu}^{2+}$ ,  $\text{Cu}_3$ ,  $\text{CuCl}^+$ ,  $\text{CuCl}_2$ ,  $\text{Al}_2\text{O}_3$ ,  $\text{Al}_2\text{O}_4\text{H}_2$ , and  $\text{Al}_4\text{O}_6$ . The resulting basis sets were tested by cross-validation: per atom type, all molecules from the test set were calculated with every optimized basis set for that atom type. Based on the cross-validation results, we manually chose a compromise set of parameters for each atom type and tested these again for all molecules in the test set. For some parameters,

simple averages were taken. For others, values closer to the best performing ones were chosen; see the Discussion section.

Furthermore, we optimized basis sets for two bulk and two surface structures:  $\alpha$ -alumina and metallic aluminum. Both surface structures contained two atomic layers (all atoms being surface atoms) and were calculated with a periodic vacuum spacing of  $\sim 10$  Å. The metallic aluminum was calculated with a k-grid cutoff of 10 Å. For the surface structures and for the  $\alpha$ -alumina bulk structure, the unit cell lengths were not varied during the geometry optimizations but were kept fixed to the experimental values. Note that the bulk- and surface structure-optimized basis set parameters were not taken into account for the compromise parameter selection.

For comparison, we performed benchmark calculations with the ADF\_2009 package<sup>18,19</sup> for reactions with all gaseous compounds from the test set and several additional compounds. These calculations were all-electron calculations (no frozen core) with the two largest basis sets available in ADF: the TZ2P and QZ4P basis sets.<sup>2</sup> For proper comparison, again the rPBE functional was used, with scalar relativistic corrections.<sup>20</sup>

### IV. RESULTS

As our research focuses on catalyst surfaces and gaseous reactants, we optimized our basis sets on gaseous molecules and not on bulk condensed systems. In the latter, for which solid-state codes like SIESTA are optimized, the basis sets of neighboring atoms overlap strongly, describing together the electron density. For that reason, relatively small confinement radii can be used without losing accuracy. For gaseous molecules or surfaces, however, the atomically centered basis orbitals must also describe the electron density reaching into vacuum. As a result, choosing the right basis sets is more critical for gases or surfaces than for bulk systems. We propose that optimizing basis sets on gaseous molecules and then using them also for surfaces or bulk systems is better than the other way around.

We examined the transferability of the per molecule optimized basis sets. As an example, Table I gives the cross-validation errors for the nitrogen DZP sets. The values are normalized to the number of nitrogen atoms in the molecule (e.g., the energy of  $\text{NO}_3^-$  computed with the N basis set of  $\text{NO}_2$  [and the O basis set of  $\text{NO}_3^-$ ] is 0.05 eV less negative than the energy of  $\text{NO}_3^-$  with optimal basis sets). Likewise,

TABLE I. Transferability of nitrogen DZP basis sets.<sup>a</sup>

	Error per N atom (eV)			
	$\text{NO}_2$	$\text{NO}_3^-$	$\text{HNO}_3$	$\text{N}_2$
Basis N_ $\text{NO}_2$	0.00	0.05	0.22	0.00
Basis N_ $\text{NO}_3^-$	0.01	0.00	0.22	0.02
Basis N_ $\text{HNO}_3$	0.18	0.09	0.00	0.67
Basis N_ $\text{N}_2$	0.01	0.06	0.18	0.00
Basis N_ <i>compromise</i>	0.03	0.08	0.12	0.02

<sup>a</sup>Nitrogen basis sets optimized for the molecules on the left were tested for the other molecules in the set. For other atom types, the optimized basis set for each test molecule was kept.

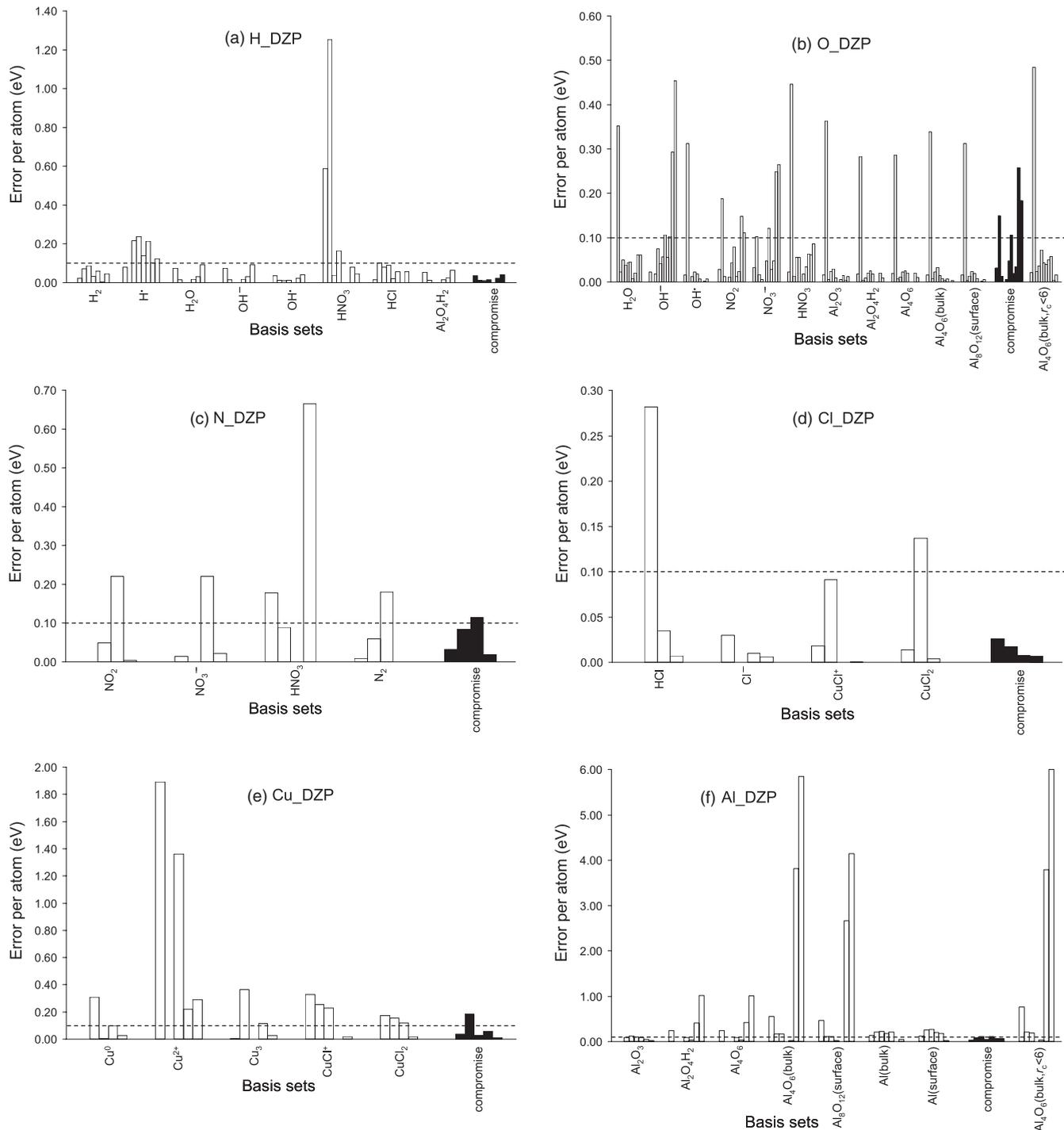


FIG. 1. Transferability of DZP basis sets for H, O, N, Cl, Cu, and Al. Shorter bars mean better transferability. The basis sets were optimized for the molecules on the  $x$ -axes and tested on the same sets of molecules. The solid black bars show our compromise basis sets.

the energy of N<sub>2</sub> computed with the N basis set of HNO<sub>3</sub> is 1.33 eV less negative than the energy of N<sub>2</sub> with an optimal basis set. Divided by the number of nitrogen atoms in N<sub>2</sub>, the error per atom is 0.67 eV.

Figures 1 and 2 show the error per atom in the cross-validations for all atom types. For easy comparison, a line is drawn at 0.1 eV in all graphs. We see that most optimized basis sets perform reasonably well for most molecules, but

there are also many combinations that give rather poor results. This shows that indeed the transferability of these basis sets is limited, and comparing basis sets optimized for several molecules is needed. Indeed, the basis sets with compromise parameters, based on such a comparison, perform significantly better.

Note that also between bulk structures the cross-tests show errors comparable with the errors between gaseous compounds

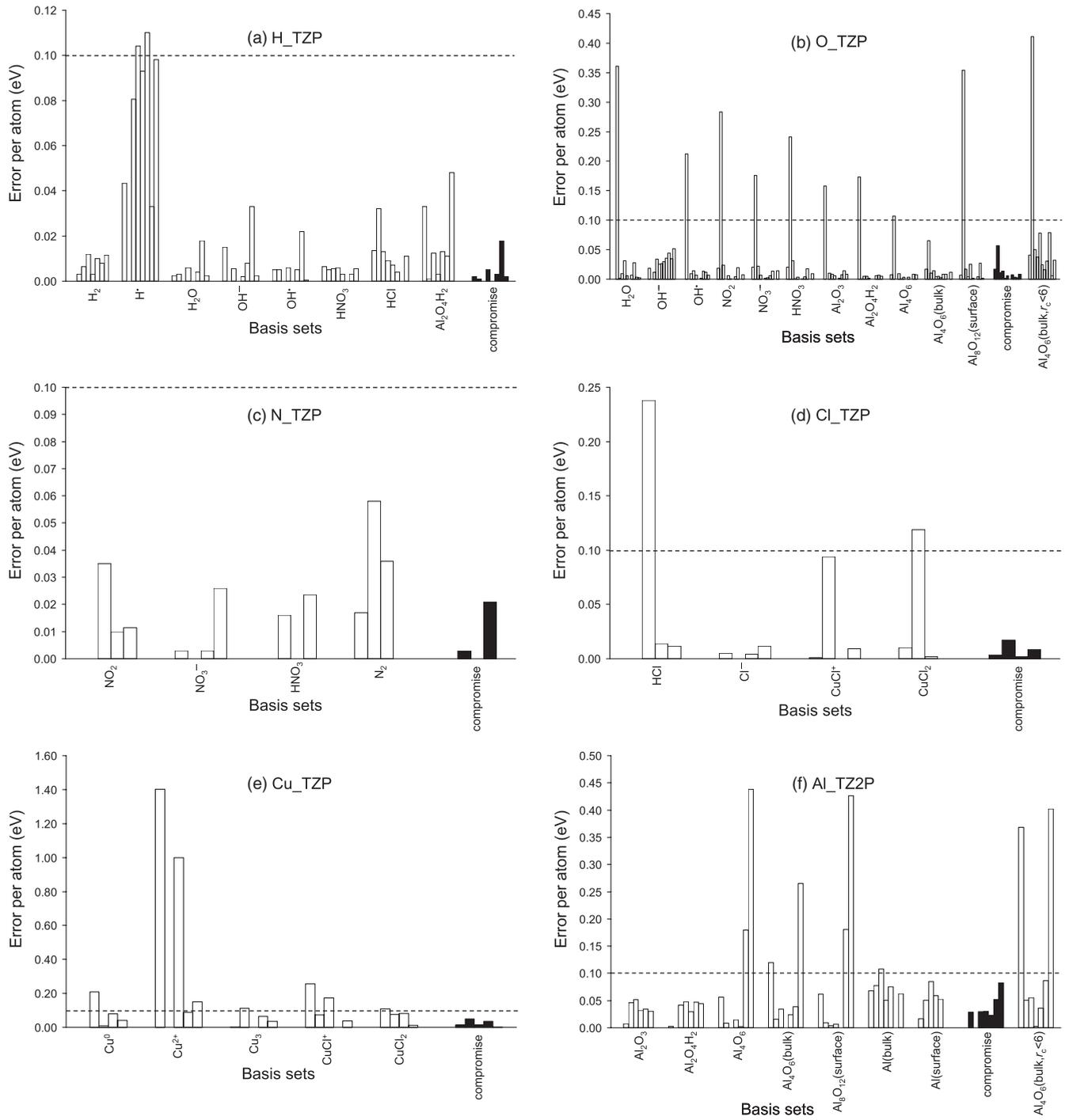


FIG. 2. Transferability of TZP basis sets for H, O, N, Cl, and Cu, and TZ2P basis sets for Al. Shorter bars mean better transferability. The basis sets were optimized for the molecules on the *x*-axes and tested on the same sets of molecules. The solid black bars show our compromise basis sets.

[aluminum bulk calculated with an alumina bulk DZP basis set is even completely off, see Fig. 1(f)]. This shows that the transferability issue raised here is not limited to some gaseous compounds but also applies for bulk structures. We maintain that it is a very general issue that basis sets that are optimized for one structure cannot be trusted for other

structures. Because usually one wants to calculate the effect of changes in a structure, this is a serious danger that needs to be addressed.

In Figs. 1 and 2, we also give results for bulk alumina optimized basis sets with *r<sub>c</sub>* parameters that are constrained below 6.0 bohr. Interestingly, the bulk alumina basis sets with

$r_c < 9.0$  perform rather well for the gaseous molecules, but the  $r_c < 6.0$  basis sets perform poorly for gaseous compounds. For the bulk structure itself and for the alumina surface structure, only small errors ( $< 0.01$  eV/atom) are found with the confined basis sets. This means that by only testing on bulk and surface structures the  $r_c < 6.0$  basis sets would have been considered a good choice, while for the gaseous species they are not, with errors up to 0.4 eV/atom. We conclude that indeed bulk structure-optimized basis sets are not necessarily good basis sets for nonbulk systems. Our compromise basis sets based on optimized sets for gaseous molecules, on the other hand, perform well for all structures, including the bulk and surface ones. The final parameters for our manually chosen sets are given in Table II.

### A. The effect of large cutoff radii

In this work, we allow cutoff radii up to 9.0 bohr, which is significantly larger than commonly used for bulk systems.<sup>10,21</sup> This choice was made because for gaseous molecules and surfaces, longer orbital tails are needed for reaching into vacuum. However, larger cutoff radii also mean more overlap with other orbitals and therefore an increase in computational cost. Figure 3 shows the improvement of quality of longer basis orbitals, compared to the increase in computational cost. For this analysis, our compromise basis sets (Table II) were used, with cutoff radii larger than the desired maximum radius being replaced by the desired value. Figure 3(a) gives the errors per atom, averaged over all gaseous compounds (19), all surface structures (2), and all bulk structures (2) used in the current work. Figure 3(b) shows the average time to converge the self-consistent field (SCF) of the first geometry step, again normalized to the amount of atoms in each structure. Clearly, for bulk solid structures larger cutoff radii hardly give any improvement, accompanied by a substantially increase in computational cost (this is exactly why smaller radii are commonly used for bulk systems). For gases, however, basis sets with larger cutoff radii do perform significantly better than confined basis sets, and the increase in computational cost is reasonable. For surfaces, the situation is somewhere

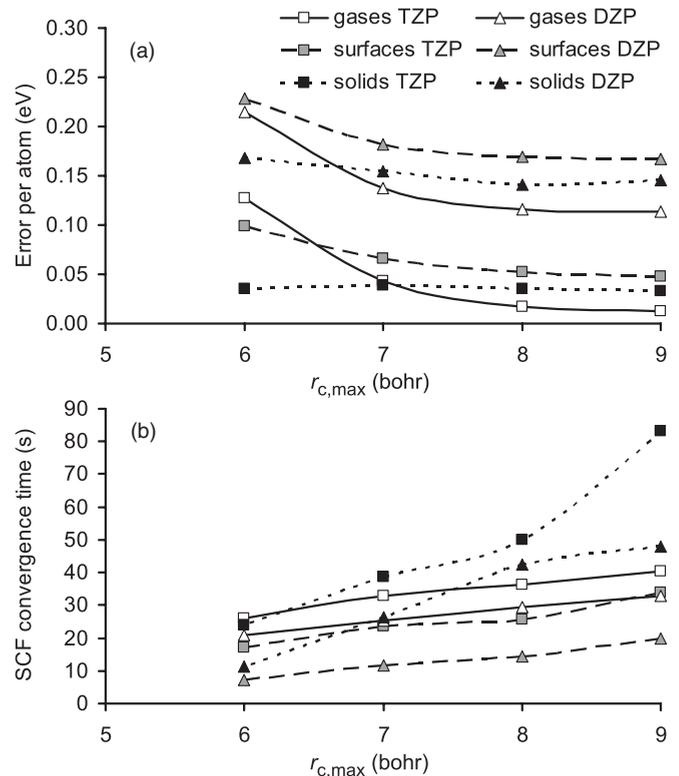


FIG. 3. Error per atom (a) and SCF convergence time per atom (b) for compromise basis sets with varying maximum cutoff radii. The errors per atom are calculated as the difference with optimized TZP basis sets.

in between. This means that one should choose the maximum cutoff radius depending on the degree of accuracy needed.

### B. Reaction energies

In practice, reaction energies are more important to predict correctly than absolute energies. In Table III, we compare our compromise basis sets with top-level ADF basis sets for reaction energies including reactions with complexes that were not used for the optimizations. First, SIESTA default basis sets

TABLE II. Final parameters for the basis sets optimized in this work.

	Orbital	$r_c$ (bohr)	$r_{m1}$ (bohr)	$r_{m2}$ (bohr)	Orbital	$r_c$ (bohr)	$r_{m1}$ (bohr)	$r_{m2}$ (bohr)	
H, DZP	1s:	8.4	2.2	–	–	–	–	–	
O, DZP	2s:	8.0	2.8	–	2p:	9.0	3.5	–	
N, DZP	2s:	9.0	2.7	–	2p:	5.4	2.7	–	
Cl, DZP	3s:	9.0	3.1	–	3p:	9.0	5.0	–	
Cu, DZP	4s:	8.0	4.3	–	3d:	8.0	2.8	–	
Al, DZP	3s:	9.0	3.3	–	3p:	6.5	3.9	–	
H, TZP	1s:	8.0	1.8	4.1	–	–	–	–	
O, TZP	2s:	8.5	2.5	5.0	2p:	9.0	2.4	5.4	
N, TZP	2s:	9.0	2.5	4.3	2p:	9.0	2.4	3.8	
Cl, TZP	3s:	8.7	1.4	4.2	3p:	9.0	3.4	6.0	
Cu, TZP	4s:	8.3	3.7	5.6	3d:	8.9	2.2	3.7	
Al, TZ2P	3s:	9.0	3.1	5.0	3p:	6.2	3.2	4.6	Split norm = 0.85

TABLE III. Gas phase reaction energies in electron volts (eV) calculated with top-level ADF basis sets and several qualities of SIESTA basis sets.

	ADF (TZ2P)	ADF (QZ4P)	Siesta default (DZP, Eshift = 0.01 eV) <sup>a</sup>		DZP compromise Parameters <sup>a</sup>		TZ(2)P compromise Parameters <sup>a,b</sup>	
2 H <sup>•</sup> → H <sub>2</sub>	-4.57	-4.58	-4.29	(0.29)	-4.45	(0.13)	-4.57	(0.01)
OH <sup>-</sup> + H <sup>+</sup> → H <sub>2</sub> O	-17.55	-17.30	-17.55	(-0.25)	-17.27	(0.03)	-17.12	(0.18)
OH <sup>•</sup> + H <sup>•</sup> → H <sub>2</sub> O	-5.19	-5.22	-4.96	(0.26)	-5.02	(0.20)	-5.07	(0.15)
NO <sub>2</sub> + OH <sup>•</sup> → HNO <sub>3</sub>	-2.22	-2.16	-2.09	(0.07)	-2.07	(0.09)	-2.03	(0.13)
H <sup>+</sup> + NO <sub>3</sub> <sup>-</sup> → HNO <sub>3</sub>	-14.50	-14.49	-14.41	(0.08)	-14.42	(0.07)	-14.39	(0.10)
2 Al <sub>2</sub> O <sub>3</sub> → Al <sub>4</sub> O <sub>6</sub>	-11.50	-11.53	-10.48	(1.05)	-10.93	(0.60)	-11.40	(0.13)
Al <sub>2</sub> O <sub>3</sub> + H <sub>2</sub> O → Al <sub>2</sub> O <sub>4</sub> H <sub>2</sub>	-4.52	-4.53	-4.05	(0.48)	-4.34	(0.19)	-4.51	(0.02)
Al <sub>4</sub> O <sub>6</sub> + H <sub>2</sub> O → Al <sub>4</sub> O <sub>7</sub> H <sub>2</sub>	-4.21	-4.16	-4.18	(-0.02)	-4.39	(-0.23)	-4.32	(-0.16)
Al <sub>4</sub> O <sub>6</sub> + NO <sub>3</sub> <sup>-</sup> → Al <sub>4</sub> O <sub>6</sub> -NO <sub>3</sub> <sup>-</sup>	-3.94	-3.84	-4.39	(-0.55)	-4.01	(-0.17)	-3.91	(-0.07)
Al <sub>4</sub> O <sub>6</sub> + OH <sup>-</sup> → Al <sub>4</sub> O <sub>6</sub> -OH <sup>-</sup>	-5.58	-5.31	-6.00	(-0.69)	-5.46	(-0.15)	-5.25	(0.06)
Cu <sup>2+</sup> + NO <sub>3</sub> <sup>-</sup> → CuNO <sub>3</sub> <sup>+</sup>	-19.38	-19.36	-20.20	(-0.84)	-19.64	(-0.28)	-19.30	(0.06)
CuNO <sub>3</sub> <sup>+</sup> + NO <sub>3</sub> <sup>-</sup> → Cu(NO <sub>3</sub> ) <sub>2</sub>	-8.48	-8.40	-8.59	(-0.19)	-8.43	(-0.03)	-8.35	(0.05)
3 Cu → Cu <sub>3</sub>	-3.15	-3.20	-3.02	(0.18)	-2.99	(0.21)	-3.01	(0.19)
Al <sub>4</sub> O <sub>6</sub> + Cu → Al <sub>4</sub> O <sub>6</sub> -Cu	-1.33	-1.35	-1.43	(-0.08)	-1.38	(-0.03)	-1.31	(0.04)
Al <sub>4</sub> O <sub>6</sub> -Cu + Cu → Al <sub>4</sub> O <sub>6</sub> -Cu <sub>2</sub>	-2.64	-2.65	-2.62	(0.03)	-2.62	(0.03)	-2.61	(0.04)
H <sup>+</sup> + Cl <sup>-</sup> → HCl	-14.84	-14.70	-14.85	(-0.15)	-14.62	(0.08)	-14.56	(0.14)
Cu <sup>2+</sup> + Cl <sup>-</sup> → CuCl <sup>+</sup>	-19.48	-19.28	-20.36	(-1.08)	-19.60	(-0.32)	-19.26	(0.02)
CuCl <sup>+</sup> + Cl <sup>-</sup> → CuCl <sub>2</sub>	-9.27	-9.07	-9.37	(-0.30)	-9.05	(0.02)	-8.96	(0.11)
Al <sub>4</sub> O <sub>6</sub> + Cl <sup>-</sup> → Al <sub>4</sub> O <sub>6</sub> -Cl <sup>-</sup>	-3.52	-3.35	-3.95	(-0.60)	-3.53	(-0.18)	-3.39	(-0.04)
Al <sub>4</sub> O <sub>7</sub> H <sub>2</sub> + Cl <sup>-</sup> → Al <sub>4</sub> O <sub>7</sub> H <sub>2</sub> -Cl <sup>-</sup>	-2.82	-2.65	-3.23	(-0.58)	-2.73	(-0.08)	-2.65	(0.00)
RMS deviation:				0.50		0.21		0.10
Average deviation:				0.14		0.01		0.06

<sup>a</sup>The differences with the ADF-QZ4P results are given in brackets.

<sup>b</sup>TZ(2)P stands for TZ2P for aluminum and TZP for all other atom types.

with Energyshift = 0.01 eV<sup>22</sup> are compared. Second, results for our optimized DZP and TZP (and TZ2P for aluminum) basis sets are shown. The default settings give errors of up to 1 eV with a root mean square (RMS) error of 0.5 eV. Such errors are simply too large compared to chemically relevant energy differences, and especially compared to the subtle differences usually found in catalysis applications. With our DZP basis sets with optimized and combined parameters, the RMS error drops to 0.2 eV, with still a few outliers. The TZP sets (and TZ2P for Al) show an RMS deviation of only 0.10 eV with maximum errors of 0.2 eV. This is in the order of general DFT precision.

Note that all the reactions in Table III are written in the direction of bond formations. This means that basis set superposition errors (BSSE) would make the reaction energies more negative in all cases (compare the ADF-TZ2P and ADF-

QZ4P columns). For the SIESTA basis sets, we also give the average deviation from the ADF-QZ4P results. Improved basis sets yield a higher (more positive) average deviation, reflecting decreasing BSSE errors. For our TZ(2)P basis sets, the average deviation even becomes significantly positive, suggesting that these basis sets outperform the ADF-QZ4P ones.

## V. DISCUSSION

As NAOs can have any shape, in principle only one such basis orbital is needed to fit the correct electron density perfectly. But, orbitals differ in different surroundings, and additional basis orbitals are needed to describe the actual chemistry (orbitals do not only overlap but also change shape as a result of changing interactions). The features needed for these additional basis orbitals depend on the specific

TABLE IV. TZP parameters for Cu as optimized for the different species used in our work.

	Orbital	$r_c$ (bohr)	$r_{m1}$ (bohr)	$r_{m2}$ (bohr)	Orbital	$r_c$ (bohr)	$r_{m1}$ (bohr)	$r_{m2}$ (bohr)
Cu <sup>o</sup>	4s:	9.0	4.1	6.1	3d:	9.0	2.3	3.1
Cu <sup>2+</sup>	4s:	5.0	0.8	4.1	3d:	5.4	2.3	3.7
Cu <sub>3</sub>	4s:	9.0	2.7	5.1	3d:	9.0	1.9	3.4
CuCl <sup>+</sup>	4s:	6.3	4.0	5.6	3d:	7.3	2.7	4.7
CuCl <sub>2</sub>	4s:	7.0	4.1	6.1	3d:	8.8	2.2	4.3

chemical structure. When one optimizes a basis set for a specific structure, it is easy to find orbitals that exactly meet the specific needs for this structure (remember: NAOs are very flexible). However, chances are that the same atom type placed in a different chemical structure will have differently shaped orbitals and needs a different basis set for an optimal fit. As a result, basis sets optimized for one structure may be suboptimal for the same atoms in a different chemical structure. A good transferable basis set offers all features needed for any structure containing that atom type. Automatic generic methods to make basis sets for all atom types (e.g., Ref. 9) are very elegant and useful, but they do not give optimal basis sets (see Table III). Basis sets need to be optimized for a range of specific compounds, and then a compromise must be made to combine the features needed for all compounds into one basis set.

As an example, we give the optimized parameters for all Cu TZP basis sets in Table IV. Here, we see how much the parameters can vary in different surroundings [Fig. 2(e) shows the effect in cross-tests]. We see for instance that positively charged species tend to prefer smaller cutoff radii than negatively charged ones. Also the preferred radii for the additional basis orbitals differ substantially among different compounds. This explains why basis sets optimized for one compound indeed can be very suboptimal for another compound.

Another issue is how to find a good compromise. Simply taking the average of the different optimized parameters would be biased by the chosen set of compounds. Instead, we examined the cross-testing results very closely, checking what parameter values tend to give good results for other compounds and what values clearly give a bad transferability. One interesting observation is that parameters optimized for negatively charged species tend to have a better transferability to positively charged species than the other way around. From such examinations of cross-tests, it follows that for some parameters simple averages can be expected to work well, and for some parameters, for instance, the value from the negatively charged species should be selected (or something in between this value and the average of the other values), leading to the handpicked parameters shown in Table II. A more automatic method could perhaps be to average the optimized parameters weighted by 1 over the errors obtained with each optimized set. Note that while the parameters we vary are method-specific, the general point of improving transferability with a compromise based on cross-tests is not.

Finally, we note that better compromises are found with TZP basis sets than with DZP basis sets. This fits with our explanation that different orbital features are needed in

different compounds, and with more basis sets these features can be mixed more independently.

Other work concerning the modeling of surfaces with localized atomic orbitals has been published by García-Gil *et al.*<sup>23</sup> They show that the electron density decay into vacuum is best described by adding ghost basis orbitals in the vacuum region. However, they mention that for modeling adsorption, this solution is not optimal. As best alternative, they advise the use of an additional diffuse orbital at the surface atoms on top of the standard DZP basis set. This is essentially the same as switching from DZP to TZP (with large  $r_c$ ) at the surface atoms. Combined with our finding that good transferability is easier obtained with TZP basis sets than with DZP basis sets, we conclude that for surfaces and gaseous compounds, TZP basis sets with large  $r_c$  should be preferred over DZP ones. For systems with many bulk atoms, for which this would be too costly, one can always choose different basis sets for bulk atoms than for surface atoms.

## VI. CONCLUSION

We showed that the transferability of basis sets that are optimized for specific molecules is limited. Most basis sets perform reasonably well for most molecules in the test set, but there are too many outliers to trust such basis sets. We have improved the transferability by taking a compromise from several optimized sets of parameters. Indeed, these compromise basis sets perform significantly better. Comparing with top-level ADF calculations of reaction energies, we have obtained a RMS error of 0.2 eV, with a few outliers, for our DZP basis sets. With our TZP basis sets the RMS error is only 0.1 eV, which is in the order of general DFT precision. Remarkably, the average deviation suggests that our TZP basis sets outperform the ADF-QZ4P basis sets. We conclude that good practice is to optimize basis sets to several compounds relevant to the system of interest and make a compromise of these basis sets. Without optimizing, unnecessary large errors are obtained. By taking a compromise from several sets, we can avoid a biased basis set. We have seen that this risk of biased basis sets is also present in the case of bulk solids.

Furthermore, we showed that strongly confined basis orbitals, which are fine for condensed bulk systems, do not necessarily perform well for gaseous molecules. Thus, optimizing basis sets for gaseous molecules and using these also for bulk compounds is better than vice versa.

## ACKNOWLEDGMENTS

We thank the Dutch National Research School Combination Catalysis (NRSC-C) for funding.

\*m.j.louwerse@uva.nl

<sup>1</sup>W. J. Hehre, R. F. Stewart, and J. A. Pople, *J. Chem. Phys.* **51**, 2657 (1969).

<sup>2</sup>E. van Lenthe and E. J. Baerends, *J. Comput. Chem.* **24**, 1142 (2003).

<sup>3</sup>J. Junquera, O. Paz, D. Sanchez-Portal, and E. Artacho, *Phys. Rev. B* **64**, 235111 (2001).

<sup>4</sup>D. R. Hamann, M. Schlüter, and C. Chiang, *Phys. Rev. Lett.* **43**, 1494 (1979).

<sup>5</sup>N. Troullier and J. L. Martins, *Phys. Rev. B* **43**, 1993 (1991).

<sup>6</sup>P. E. Blöchl, *Phys. Rev. B* **50**, 17953 (1994).

<sup>7</sup>P. Ordejon, E. Artacho, and J. M. Soler, *Phys. Rev. B* **53**, R10441 (1996).

- <sup>8</sup>J. M. Soler, E. Artacho, J. D. Gale, A. Garcia, J. Junquera, P. Ordejon, and D. Sanchez-Portal, *J. Phys. Condens. Matter* **14**, 2745 (2002).
- <sup>9</sup>E. Artacho, D. Sanchez-Portal, P. Ordejon, A. Garcia, and J. M. Soler, *Phys. Status Solidi B* **215**, 809 (1999).
- <sup>10</sup>G. Poulet, P. Sautet, and E. Artacho, *Phys. Rev. B* **68**, 075118 (2003).
- <sup>11</sup>G. Paglia, A. L. Rohl, C. E. Buckley, and J. D. Gale, *Phys. Rev. B* **71**, 224115 (2005).
- <sup>12</sup>T. Ozaki and H. Kino, *J. Chem. Phys.* **121**, 10879 (2004).
- <sup>13</sup>A. D. Boese, J. M. L. Martin, and N. C. Handy, *J. Chem. Phys.* **119**, 3005 (2003).
- <sup>14</sup>O. F. Sankey and D. J. Niklewski, *Phys. Rev. B* **40**, 3979 (1989).
- <sup>15</sup>B. Hammer, L. B. Hansen, and J. K. Norskov, *Phys. Rev. B* **59**, 7413 (1999).
- <sup>16</sup>Octopus Pseudopotential Generator [<http://www.tddft.org/programs/octopus/pseudo.php>].
- <sup>17</sup>S. G. Louie, S. Froyen, and M. L. Cohen, *Phys. Rev. B* **26**, 1738 (1982).
- <sup>18</sup>C. Fonseca Guerra, J. G. Snijders, G. te Velde, and E. J. Baerends, *Theor. Chim. Acta.* **99**, 391 (1998).
- <sup>19</sup>G. te Velde, F. M. Bickelhaupt, S. J. A. van Gisbergen, C. Fonseca Guerra, E. J. Baerends, J. G. Snijders, and T. Ziegler, *J. Comput. Chem.* **22**, 931 (2001).
- <sup>20</sup>E. Van Lenthe, E. J. Baerends, and J. G. Snijders, *J. Chem. Phys.* **101**, 9783 (1994).
- <sup>21</sup>E. Anglada, J. M. Soler, J. Junquera, and E. Artacho, *Phys. Rev. B* **66**, 205101 (2002).
- <sup>22</sup>Note that due to a mistake in the units made by the SIESTA authors, the true default setting in SIESTA is Energyshift = 0.27 eV. However, this gives truly bad results, so we use a value of 0.01 eV as the default setting.
- <sup>23</sup>S. García-Gil, A. García, N. Lorente, and P. Ordejón, *Phys. Rev. B* **79**, 075441 (2009).