Achieving plane wave accuracy in linear-scaling density functional theory applied to periodic systems: A case study on crystalline silicon

Chris-Kriton Skylaris[a)]
School of Chemistry, University of Southampton, Highfield, Southampton SO17 1BJ, United Kingdom

Peter D. Haynes
Thomas Young Centre, Imperial College London, London SW7 2AZ, United Kingdom

(Received 17 July 2007; accepted 17 September 2007; published online 26 October 2007)

Linear-scaling methods for density functional theory promise to revolutionize the scope and scale of first-principles quantum mechanical calculations. Crystalline silicon has been the system of choice for exploratory tests of such methods in the literature, yet attempts at quantitative comparisons under linear-scaling conditions with traditional methods or experimental results have not been forthcoming. A detailed study using the ONETEP code is reported here, demonstrating for the first time that plane wave accuracy can be achieved in linear-scaling calculations on periodic systems. © *2007 American Institute of Physics*. [DOI: 10.1063/1.2796168]

I. INTRODUCTION

In conventional first-principles calculations with Kohn-Sham[1] density functional theory (DFT),[2] the computational effort increases in proportion to the cube of the number of atoms N, limiting their applicability to systems of no more than a few hundred atoms. However, the scope and scale of DFT calculations are now being radically extended by the emergence of new "order-N" $[O(N)]$ methods with linear-scaling cost. While the nearsightedness principle of quantum mechanics[3,4] guarantees that linear scaling is possible, at least for semiconductors and insulators, achieving it in practice has necessitated overcoming a considerable number of significant issues of detail.[5,6] The last decade has seen intensive research worldwide into the development of reliable and generally applicable linear-scaling approaches. As a result, currently, a number of linear-scaling DFT codes[7–12] have emerged that are robust and stable enough to be used to study properties of materials, biomolecules, and nanostructures.

Our ONETEP[10] linear-scaling DFT code has been implemented[13] from the beginning for calculations on parallel computers. For systems with a gap, linear scaling is achieved by exploiting the exponential decay of the single-particle density matrix, which is represented as

$$\rho(\mathbf{r},\mathbf{r}') = \sum_{\alpha,\beta} \phi_\alpha(\mathbf{r}) K^{\alpha\beta} \phi^*_\beta(\mathbf{r}'). \quad (1)$$

This is a quadratic form in a set of *nonorthogonal generalized Wannier functions*[14] (NGWFs) $\{\phi_\alpha(\mathbf{r})\}$, involving the *density kernel* $\mathbf{K}$ which relates the NGWFs to the Kohn-Sham orbitals and their occupancies. We optimize the energy fully self-consistently with respect to both the density kernel and the NGWFs. This *in situ* optimization of the NGWFs makes our method free from basis set superposition error.[15] Linear-scaling calculations with thousands of atoms[10] can be performed with this approach, as the NGWFs are kept localized in spherical regions of radius r_{NGWF} centered on atoms, and the elements of the density kernel that correspond to atoms further than a predefined spatial cutoff threshold $r_{\mathbf{K}}$ are truncated. For cases where a quick calculation with limited accuracy is sufficient, we can keep selected NGWFs fixed (e.g., set to atomic orbitals) and optimize self-consistently only $\mathbf{K}$, resulting in a linear-scaling self-consistent *ab initio* tight binding[16] (SC-AITB) calculation.

In conventional plane wave calculations, the basis set is comprised of plane waves with kinetic energy up to a set cutoff E_{cut}. In ONETEP we use a basis set of periodic sinc[14,17] (psinc) functions which are constructed from plane waves. However, in this case, E_{cut} signifies that the plane waves are selected from a cube of wave vectors in momentum space with the same volume as the sphere of wave vectors defined by E_{cut}.[18]

In what follows, we present an investigation of the factors that determine the accuracy of a linear-scaling method in calculations on periodic solids using crystalline silicon as a case study.

II. METHODOLOGY

We have chosen silicon not only because it is particularly challenging for linear-scaling approaches due to its high atom density and small band gap (especially when the local density approximation[19,20] (LDA) for exchange and correlation is used), but also because it is the prototypical periodic solid and as such it has been the subject of a plethora of studies.[21–23] We have performed our calculations using a simulation cell large enough (1000 atoms) to allow us to test the density matrix truncation techniques that are responsible for the linear-scaling behavior. We compare our results to calculations with the CASTEP[24] cubic-scaling pseudopotential plane wave code which we use as an accuracy benchmark. To "imitate" the ONETEP 1000-atom calculations with

[a)]Electronic mail: cks@soton.ac.uk and URL: http://www.soton.ac.uk/chemistry/research/skylaris/skylaris.html

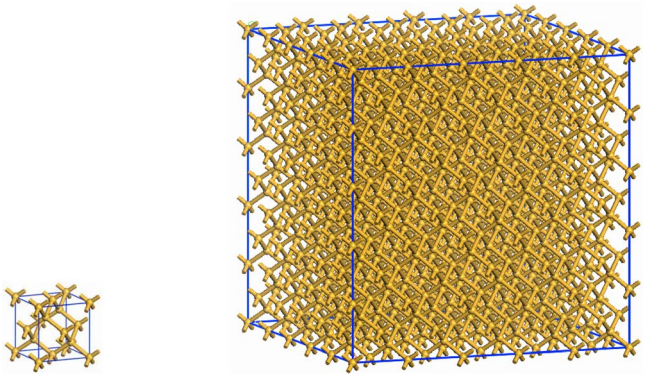

FIG. 1. (Color online) The 8-atom simulation cell used by CASTEP and the 1000-atom cell used by ONETEP.

CASTEP, we have used a conventional 8-atom simulation cell with a grid of $5\times5\times5$ **k** points.

The two simulation cells are compared in Fig. 1. The two codes used the same Si norm-conserving pseudopotential and a plane wave cutoff E_{cut} of 600 eV, which ensured that all calculated quantities were converged with respect to E_{cut}. All calculations were performed with the LDA exchange-correlation functional.

The localization of the NGWFs in spherical regions is a basis set variational[25] procedure and, therefore, the energy is lowered as the sphere radii r_{NGWF} increase. This is confirmed in Fig. 2, which shows a lowering of the energy as a function of r_{NGWF}. In addition, in ONETEP due to the equivalence of psinc functions and plane waves, we can also define values of E_{cut} for which the CASTEP energy will be an upper or lower bound.[18] These bounds are also shown in Fig. 2, where we can confirm that they are indeed obeyed for sufficiently large r_{NGWF}.

A defining property of a crystalline solid is its lattice constant, which we can calculate by varying the shape of the simulation cell until its energy is minimized. However, as this variation changes the lattice vectors, it will also change the reciprocal lattice and hence the plane wave basis set. Therefore, one can either keep fixed the number of plane

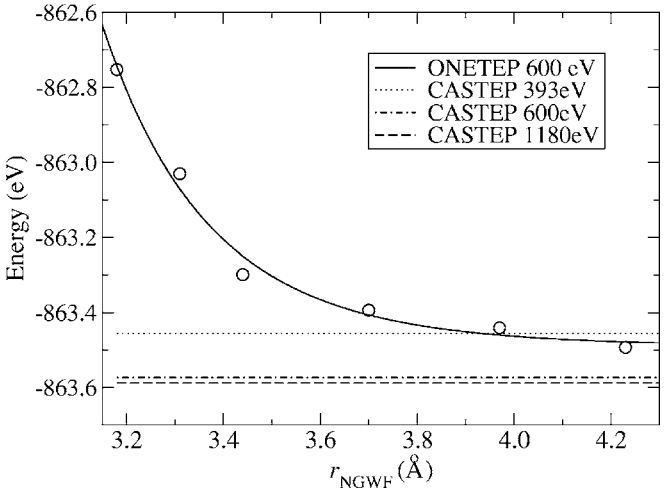

FIG. 2. Energy per 8-atom cell as a function of r_{NGWF} with E_{cut}=600 eV. The energies obtained with CASTEP for E_{cut} set to 393 eV (upper bound), 600 eV (equivalent cutoff), and 1180 eV (lower bound) are also shown.

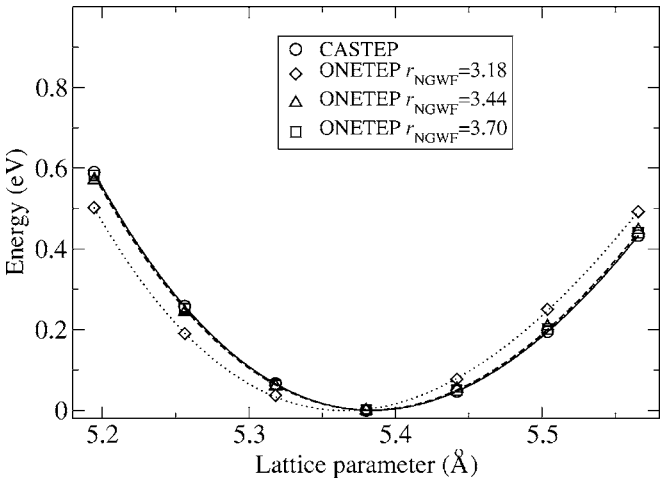

FIG. 3. Energy per 8-atom cell for different values of r_{NGWF} (in Å) and $r_{\mathbf{K}}=\infty$. The curve obtained with CASTEP is also shown.

waves and vary E_{cut} or fix E_{cut} and vary the number of plane waves. We have chosen the latter as it is physically more meaningful and, in addition, it enables the centers of our psinc basis functions to remain stationary. This restricts the variation of the lattice parameter to discrete steps equal to the psinc spacing (0.31 Å for our E_{cut} of 600 eV). It removes, however, the need to scale r_{NGWF} with the lattice parameter to ensure a constant number of psinc functions per NGWF localization sphere, which is required to obtain smooth curves when the number of plane waves is kept fixed.[18] Finally, we need to consider how many NGWFs to use per Si atom. In previous work,[25] we have shown that having as many NGWFs as the number of valence atomic orbitals on each atom, "single zeta" (SZ), was sufficient for achieving plane wave accuracy in organic molecules and biomolecules with first and second row elements.[18] For crystalline silicon, however, the level of accuracy achieved with four NGWFs per Si atom is considerably lower.[25] To ensure high accuracy we have, therefore, used here nine NGWFs per Si atom which correspond to a "single zeta plus polarization" (SZP) choice. Even though the NGWFs have arbitrary angular dependence, it appears that the flexibility afforded by the extra functions is needed to describe correctly the top of the silicon valence band. Similar trends are well known in atomic orbital approaches where the description of the band gap of silicon becomes qualitatively wrong (direct instead of indirect) when not including either a shell of *d* functions per atom[26] or a shell of *s* functions at each tetrahedral interstitial site.[27]

III. RESULTS AND DISCUSSION

In order to study how the localization we impose on the NGWFs affects the calculated properties, we have performed a series of calculations with NGWFs of increasing radii r_{NGWF} while keeping $r_{\mathbf{K}}$ set to infinity (no density kernel truncation). Plots of the energy as a function of the lattice parameter are shown in Fig. 3 which depicts the points obtained from the calculations and the curves obtained by fitting to the Birch-Murnaghan[28] equation of state. To facilitate comparisons, the curves have been shifted so that their mini-

TABLE I. The lattice constant a and bulk modulus B of crystalline silicon for different values of NGWF and kernel radii (both given in Å). Also shown are the number of self-consistent cycles performed in each case as well as the convergence tolerance for the energy per atom.

Method	r_{NGWF}	$r_{\mathbf{K}}$	a (Å)	B (GPa)	N_{it}	tol (eV/atom)
CASTEP	N/A	N/A	5.384	96.3	18	5×10^{-7}
ONETEP	3.70	∞	5.382	96.5	19	3×10^{-8}
ONETEP	3.44	∞	5.380	96.2	18	3×10^{-8}
ONETEP	3.18	∞	5.365	96.8	13	3×10^{-7}
ONETEP (fireballs)	3.70	∞	5.482	84.4	6	5×10^{-6}
ONETEP (STO-3G*)	3.70	∞	5.290	134.2	6	2×10^{-5}
ONETEP (scaled **K**)	3.70	10.58	5.385	97.6	11	1×10^{-5}
ONETEP	3.70	10.58	5.385	97.8	11	1×10^{-5}
ONETEP	3.70	7.94	5.378	99.5	11	3×10^{-5}
ONETEP	3.70	5.29	5.412	122.7	11	3×10^{-5}

mum is at the zero of energy. Physical properties depend on energy differences and are thus not affected by this adjustment.

We observe that for $r_{\mathrm{NGWF}}\geqslant 3.44$ Å, the ONETEP curves practically coincide with CASTEP. This remarkable agreement between the two plane wave approaches is also confirmed by the lattice constants and bulk moduli that we obtain from the curves, which are presented in Table I.

These calculations are not strictly linear scaling because while the Hamiltonian and overlap matrices are sparse and built with $O(N)$ cost, **K** is a full square matrix. Fully linear-scaling calculations are included in Fig. 4 which plots the energy as a function of the lattice parameters for different values of $r_{\mathbf{K}}$ with r_{NGWF} set to 3.70 Å. The resulting lattice constants and bulk moduli are again shown in Table I.

The excellent agreement between the two plane wave approaches is evident from the table. For example, in the case of no kernel truncation, the error we obtain for the $r_{\mathrm{NGWF}}=3.70$ Å case is 0.04% for the lattice constant and 0.2% for the bulk modulus. The SIESTA developers have reported[8] agreement with a plane wave calculation of 0.02% for the lattice constant and 0.5% for the bulk modulus when using a TZTPF (34 atomic orbitals per Si atom) pseudoatomic orbital basis set. These authors do not, however, report the atomic orbital localization radii they employed for these results and, furthermore, they imply that their calculation was performed by diagonalization in a small simulation cell so that their Wannier function-based linear-scaling energy optimization technique was not tested. In contrast, no diagonalization was employed to optimize the energy in ONETEP, which did so directly in the parameter space of 1000 atoms using density matrix search techniques.[29,30] Haynes and Payne[30] using the penalty functional linear-scaling approach reported agreements of 0.7% for the lattice parameter and 2.6% for the bulk modulus with a spherical wave basis set. To our knowledge, no values for the lattice constant and bulk modulus of crystalline silicon calculated in a fully self-consistent fashion with the CONQUEST code using the blip basis set have been reported.

In Table I we also show a SC-AITB calculation we run with a SZP set of pseudoatomic orbitals preoptimized for an isolated Si atom within a spherical well with radius of 3.7 Å in the "fireballs" approach of Sankey and Niklewski.[31] The results obtained are poor by *ab initio* standards with errors of 1.8% for the lattice constant and 12.4% for the bulk modulus; they are, however, comparable to those reported for linear-scaling approaches of a tight-binding nature with various levels of sophistication.[21,29,32–35] Results of similar quality are obtained by performing SC-AITB calculations with the STO-3G* Gaussian basis set,[36] which is also of SZP type.

A very important observation can be made from Fig. 4 and Table I concerning the calculations where density kernel truncation has been applied. It is evident that in ONETEP both truncations (r_{NGWF} and $r_{\mathbf{K}}$) that are used to achieve linear scaling did not prevent us from achieving plane wave accuracy. For example, setting $r_{\mathbf{K}}$ equal to 10.58 Å results in errors of only 0.02% for the lattice constant and 1.6% for the bulk modulus with respect to the "exact" CASTEP values. In Table I we also show results for "scaled **K**," which refers to scaling $r_{\mathbf{K}}$ with the lattice parameter. We have used this approach in the past,[18] but here we show that it makes virtually no difference to the results, confirming that ONETEP can be viewed as a "black box" technique that can be used by nonexperts to predict properties of real materials using only a minimal set of calculation parameters (E_{cut}, r_{NGWF}, and $r_{\mathbf{K}}$) whose specification can be automated.

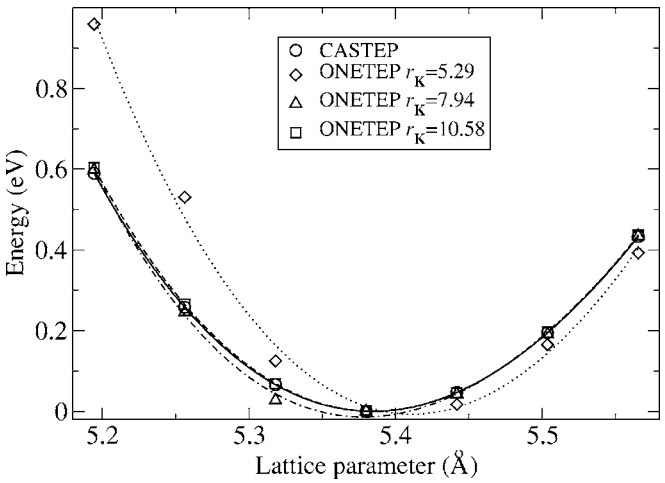

FIG. 4. Energy per 8-atom unit cell as a function of the lattice parameter for different values of $r_{\mathbf{K}}$ (in Å) and $r_{\mathrm{NGWF}}=3.70$ Å. The curve obtained by CASTEP is also shown.

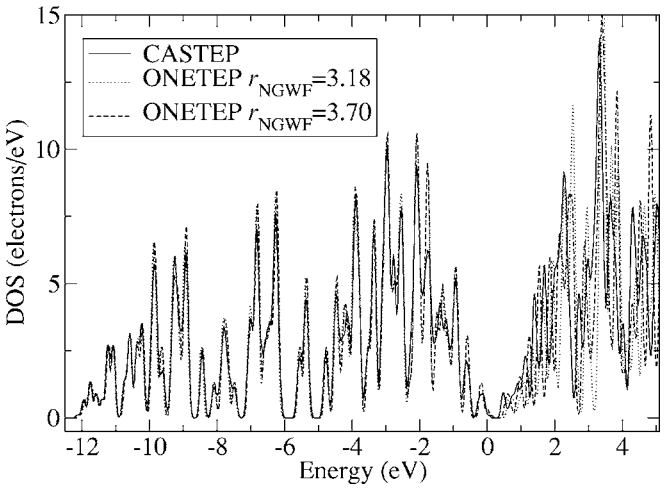

FIG. 5. Density of states (DOS) per 8-atom cell for different values of r_{NGWF} (in Å) and $r_{\mathbf{K}}=\infty$. The DOS obtained with CASTEP is also shown.

To examine the description of individual states by our linear-scaling formalism, we have calculated densities of states (DOS), with a very narrow Gaussian smearing width (0.05 eV) for a detailed comparison of band structures. In Fig. 5 we show plots for the cases with infinite $r_{\mathbf{K}}$ and varying r_{NGWF}. The agreement with CASTEP is already remarkable from r_{NGWF}=3.18 Å, if we limit our attention only to the occupied states (the band gap occurs at around 0.2 eV). For the conduction states that lie about 0.5 eV higher than the band gap, the agreement between ONETEP and CASTEP is lost and does not improve with increasing r_{NGWF}. This demonstrates that the NGWFs are only capable of describing the valence space, as expected, since our optimization for the total energy focuses only on the valence states. Therefore, the NGWFs should not be used for predicting properties that depend on the accurate description of the conduction bands, except perhaps only the lowest-lying ones. In Fig. 6 we plot the DOS obtained with r_{NGWF} fixed to 3.70 Å and various values of $r_{\mathbf{K}}$. Only the rather aggressive $r_{\mathbf{K}}$ value of 5.29 Å causes significant distortion in the DOS. With a value of 10.58 Å the DOS already coincides with what is obtained with no density kernel truncation.

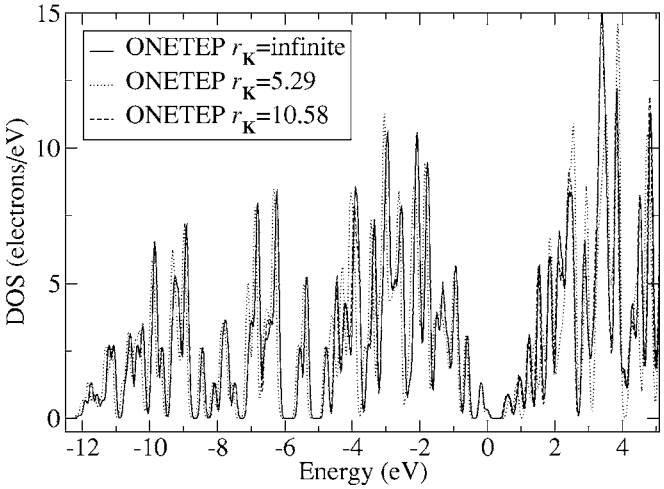

FIG. 6. Density of states (DOS) per 8-atom cell for different values of $r_{\mathbf{K}}$ (in Å) and r_{NGWF}=3.44 Å.

In Table I we also give the number of self-consistent cycles performed for each type of calculation and the levels of convergence achieved for the total energy per atom. For the ONETEP calculations, this refers to the NGWF optimization cycles when NGWF optimization is applied and to the density kernel cycles when the NGWFs are held fixed (the fireballs and "STO-3G*" cases). We observe that when no density kernel truncation is applied, the speed and convergence level achieved are equal to or better than the conventional plane wave approach. With density kernel truncation the convergence in the energy becomes less tight, as dictated by the applied level of truncation.[37] The small number of iterations in all cases is a consequence of our preconditioning scheme.[17]

IV. CONCLUSIONS

In conclusion, crystalline solids present a tough challenge for linear-scaling approaches due to their high atom density, and crystalline silicon which is the prototypical periodic system is one of the most difficult in this respect. We have reported the first detailed quantitative study of structural and electronic properties of silicon based on calculations performed on a sufficiently large simulation cell (1000 atoms) to allow the use of all the density matrix truncation techniques that are employed to obtain linear-scaling behavior. The results show that ONETEP is able to achieve plane wave accuracy with the density matrix localization settings that it employs in routine linear-scaling calculations.

ACKNOWLEDGMENTS

The authors would like to thank the Royal Society for University Research Fellowships and the Southampton University Information Systems Services (particularly Dr. Oz Parchment, Dr. David Baker, and Dr. Ivan Wolton) for support with supercomputing facilities.

[1] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).
[2] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).
[3] E. Prodan and W. Kohn, Proc. Natl. Acad. Sci. U.S.A. **102**, 11635 (2005).
[4] W. Kohn, Phys. Rev. Lett. **76**, 3168 (1996).
[5] S. Goedecker, Rev. Mod. Phys. **71**, 1085 (1999).
[6] G. Galli, Curr. Opin. Solid State Mater. Sci. **1**, 864 (1996).
[7] D. R. Bowler, R. Choudhury, M. J. Gillan, and T. Miyazaki, Phys. Status Solidi B **243**, 989 (2006).
[8] J. M. Soler, E. Artacho, J. D. Gale, A. García, J. Junquera, P. Ordejón, and D. Sánchez-Portal, J. Phys.: Condens. Matter **14**, 2745 (2002).
[9] J. L. Fattebert and J. Bernholc, Phys. Rev. B **62**, 1713 (2000).
[10] C.-K. Skylaris, P. D. Haynes, A. A. Mostofi, and M. C. Payne, J. Chem. Phys. **122**, 084119 (2005).
[11] M. Challacombe, J. Chem. Phys. **110**, 2332 (1999).
[12] J. VandeVondele, M. Krack, M. Fawzi, M. Parrinello, T. Chassaing, and J. Hutter, Comput. Phys. Commun. **167**, 103 (2005).
[13] C.-K. Skylaris, P. D. Haynes, A. A. Mostofi, and M. C. Payne, Phys. Status Solidi B **243**, 973 (2006).
[14] C.-K. Skylaris, A. A. Mostofi, P. D. Haynes, O. Diéguez, and M. C. Payne, Phys. Rev. B **66**, 035119 (2002).
[15] P. D. Haynes, C.-K. Skylaris, A. A. Mostofi, and M. C. Payne, Chem. Phys. Lett. **422**, 345 (2006).
[16] A. P. Horsfield and A. M. Bratkovsky, J. Phys.: Condens. Matter **12**, R1 (2000).

[17] A. A. Mostofi, P. D. Haynes, C.-K. Skylaris, and M. C. Payne, J. Chem. Phys. **119**, 8842 (2003).
[18] C.-K. Skylaris, P. D. Haynes, A. A. Mostofi, and M. C. Payne, J. Phys.: Condens. Matter **17**, 5757 (2005).
[19] D. M. Ceperley and B. J. Alder, Phys. Rev. Lett. **45**, 566 (1980).
[20] J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).
[21] P. Ordejón, D. A. Drabold, R. M. Martin, and M. P. Grumbach, Phys. Rev. B **51**, 1456 (1995).
[22] J.-L. Fattebert and F. Gygi, Comput. Phys. Commun. **162**, 24 (2004).
[23] F. J. H. Ehlers, A. P. Horsfield, and D. R. Bowler, Phys. Rev. B **73**, 165207 (2006).
[24] S. J. Clark, M. D. Segall, C. J. Pickard, P. J. Hasnip, M. I. J. Probert, K. Refson, and M. C. Payne, Z. Kristallogr. **220**, 567 (2005).
[25] C.-K. Skylaris, O. Diéguez, P. D. Haynes, and M. C. Payne, Phys. Rev. B **66**, 073103 (2002).
[26] P. Deák and L. C. Snyder, Phys. Rev. B **36**, 9619 (1987).
[27] W. R. Lambrecht and O. K. Andersen, Phys. Rev. B **34**, 2439 (1986).
[28] F. D. Murnaghan, Proc. Natl. Acad. Sci. U.S.A. **30**, 244 (1944).
[29] X.-P. Li, R. W. Nunes, and D. Vanderbilt, Phys. Rev. B **47**, 10891 (1993).
[30] P. D. Haynes and M. C. Payne, Phys. Rev. B **59**, 12173 (1999).
[31] O. F. Sankey and D. J. Niklewski, Phys. Rev. B **40**, 3979 (1989).
[32] F. Mauri, G. Galli, and R. Car, Phys. Rev. B **47**, 9973 (1993).
[33] P. Ordejón, D. A. Drabold, M. P. Grumbach, and R. M. Martin, Phys. Rev. B **48**, 14646 (1993).
[34] P. Ordejón, E. Artacho, and J. M. Soler, Phys. Rev. B **53**, R10441 (1996).
[35] J. Kim, J. W. Wilkins, F. S. Khan, and A. Canning, Phys. Rev. B **55**, 16186 (1997).
[36] W. J. Hehre, R. Ditchfield, R. F. Stewart, and J. A. Pople, J. Chem. Phys. **52**, 2769 (1970).
[37] A. H. R. Palser and D. E. Manolopoulos, Phys. Rev. B **58**, 12704 (1998).