# Shape and energy consistent pseudopotentials for correlated electron systems

J. R. Trail and , and R. J. Needs

---

**Articles you may be interested in**

---

# Shape and energy consistent pseudopotentials for correlated electron systems

J. R. Trail[a)] and R. J. Needs
*Theory of Condensed Matter Group, Cavendish Laboratory, J J Thomson Avenue,
Cambridge CB3 0HE, United Kingdom*

A method is developed for generating pseudopotentials for use in correlated-electron calculations. The paradigms of shape and energy consistency are combined and defined in terms of correlated-electron wave-functions. The resulting energy consistent correlated electron pseudopotentials (eCEPPs) are constructed for H, Li–F, Sc–Fe, and Cu. Their accuracy is quantified by comparing the relaxed molecular geometries and dissociation energies which they provide with all electron results, with all quantities evaluated using coupled cluster singles, doubles, and triples calculations. Errors inherent in the pseudopotentials are also compared with those arising from a number of approximations commonly used with pseudopotentials. The eCEPPs provide a significant improvement in optimised geometries and dissociation energies for small molecules, with errors for the latter being an order-of-magnitude smaller than for Hartree-Fock-based pseudopotentials available in the literature. Gaussian basis sets are optimised for use with these pseudopotentials. *Published by AIP Publishing.*
[http://dx.doi.org/10.1063/1.4984046]

## I. INTRODUCTION

Electronic structure methods in solid state theory and quantum chemistry provide a powerful hierarchy of *ab initio* tools for the accurate description of interacting atomic systems. There are fundamental limits to the accuracy at each level of theory. Pseudopotentials, or effective core potentials, provide an approximation that replaces the interaction of largely inactive core electrons and valence electrons with a potential, which reduces the number of particles required.[1] Such an approximation is often necessary due to limitations on computational resources, and it also allows the elimination of analytic and numerical problems associated with singularities in electron-nuclear interaction potentials.

Methods for generating pseudopotentials within density functional theory (DFT) are well established. Errors associated with pseudopotentials are often systematically controlled and reliably estimated to be small and secondary to the errors due to approximate exchange correlation functionals.[2] However, the accuracy of DFT is not sufficient for many systems. Methods that involve explicitly correlated wave functions, such as coupled cluster[3] (CC) or quantum Monte Carlo[4–7] (QMC) theory, are often required.

The cost of coupled-cluster calculations with single, double, and perturbative triple excitations [CCSD(T)] scales with the seventh power of the number of electrons, or as $O(n^3N^4)$,[3] with $n$ the number of electrons and $N$ the number of basis functions. The use of pseudopotentials reduces both $n$ and $N$ by removing core electrons and requiring fewer basis functions than the nucleus-electron interaction.

The computational cost of fermion QMC calculations increases more slowly, as the third or fourth power of the number of particles. However, the scaling of the error with atomic number, $Z$, is $Z^5 - Z^{6.5}$.[8,9] The use of pseudopotentials reduces the effective value of $Z$, making QMC calculations feasible for systems with heavy atoms.

At these higher levels of theory, estimates of, and methods to control, pseudopotential errors are not usually available. Furthermore, available pseudopotentials often reproduce atomic properties at the DFT or HF level of theory. This introduces an uncontrolled error due to the inconsistent application of theory to the definition and application of the pseudopotentials.

In this study, we generate pseudopotentials from explicitly correlated atomic wave-functions, with no recourse to reduce these many-body wave-functions to orbitals arising from a self-consistent potential. The influence of core electrons on valence electrons is recreated by consistently combining three distinct representations of the core-valence interaction, namely, core scattering, core polarisation, and differences between ground state energies.

Scattering properties of core electrons are defined in terms of density matrices of atoms or ions with one valence electron. Long-range effects from dynamic core relaxation (arising from electron-electron correlation) can be included in the pseudopotential via a core polarisation potential (CPP).[10]

Pseudopotentials defined to reproduce the one-body core-valence interaction and core polarisation are not unique. Consequently, the accuracy may be further improved by searching over all pseudopotentials that reproduce such properties, and seeking that which most accurately reproduces ionisation and excitation energies, and electron affinities for an isolated atom.

A combined reproduction of core scattering, core polarisation, and atomic excitation energies allows the generation of a new pseudopotential from correlated electron

a)jrt32@cam.ac.uk

calculations, referred to as an energy consistent correlated electron pseudopotential (eCEPP).

Our procedure for constructing eCEPPs is an extension of the correlated electron pseudopotential (CEPP) method described in two previous papers,[11,12] which are summarised here. Only two of the criteria given above are used to generate CEPPs, the preservation of core scattering and the representation of core polarizability. An all-electron (AE) multideterminant wave-function for an ion/atom with one valence electron provides an AE charge density. From this a single-electron charge density is constructed by removing an *ab initio* core electron charge density. Within a "core radius" of the nuclei, the single-electron charge density is replaced by a standard form that preserves the continuity of the value and first four derivatives at the core radius. Direct inversion of a one-electron Schrödinger equation then provides the pseudopotential. The analytic basis for this approach is an extension of norm-conservation[13] to the many-body Hamiltonian, which conserves the scattering properties of core electrons to first order by reproducing the one-body density matrix outside of the core region.[14]

Two subtle issues arise in the process of constructing CEPPs. The multi-determinant wave-function is composed of orbitals that are not unique and without associated energy eigenvalues. A separation of the density matrix into core and valence components is achieved using natural orbitals[15] and defining the core in terms of the natural orbitals with the largest occupation numbers. While core natural orbitals have no direct effect on the pseudo wave-function in the core region, they validate the core radius used and remove a small but significant influence of core electrons at the core radius.

The long-range, one-body part of the CPP naturally emerges in the process of generating a CEPP. This is represented using well-known parameterised CPP forms[16,17] such that the CEPP is the sum of an effective potential and the CPP, and the one-body part of the total potential is *ab initio*. The ion-ion and many-body part of the CPP remains semiempirical, but with parameters that are consistent with the one-body component.

The CEPPs generated for H, Li–F, and Sc–Fe, together with CCSD(T), were shown to reproduce AE results for small molecules. Perhaps the strongest indicator that the CEPP provides a good starting point is that it is a pseudopotential generated from a highly ionic state which is accurate for a neutral molecule. This is in contrast to the well known poor transferability of Kohn-Sham (KS)-DFT pseudopotentials between different oxidation states, which is primarily caused by a pseudopotential definition that takes the self-consistent potential to be fixed and uses de-screening to remove the interaction between valence electrons. Neither of these approximations is used for the CEPPs. Iron provides a good example, with a CEPP generated from a $Fe^{15+}$ ion resulting in properties for molecules containing iron which are as accurate as for molecules with few valence electrons such as $Li_2$.

Before defining eCEPPs, we note two alternative strategies for improving the accuracy of pseudopotentials that have not been investigated. The definition of the CEPPs could be extended to reproduce two-body core scattering properties. This would require the construction of a pseudo-Hamiltonian that reproduces both first- and second-order density matrices outside the core region of an isolated ion/atom with two valence electrons. Such a pseudo-Hamiltonian would contain an extra two-body potential that describes the electron-electron interaction mediated by core electrons. This two-body interaction potential is approximately included in the CPP part of the CEPPs and is small, so it is unlikely that a more accurate description would significantly improve the pseudopotential accuracy.

Another, related, strategy would be to reproduce one-body core scattering properties higher than first order. An extension of norm-conserving KS-DFT pseudopotentials to reproduce higher order scattering properties[18] provides a modest reduction in the transfer error, suggesting that this approach will not reliably provide a more transferable pseudopotential.

These observations suggest that an eCEPP that conserves the CEPP properties while enforcing further "energy consistency" between a number of states may provide more accurate pseudopotentials. Such a combination of extrapolation and interpolation of atomic properties combines the strengths of shape-consistent and energy consistent *ab initio* pseudopotentials. Electron correlation can be included throughout, without the need for an independent electron approximation.

The article is organised as follows. Section II describes the basis and implementation of the combined energy and shape consistent strategy outlined above. The ability of eCEPPs to reproduce AE results in correlated electron calculations is analysed in Sec. III. The performance of the pseudopotentials for titanium is addressed in Sec. III A, and errors in optimised geometries and dissociation energies for a test set of molecules composed of first row and $3d$ transition metal atoms are assessed in Sec. III B. Section III C provides a comparison of errors arising from the eCEPPs with those from other approximations commonly used in *ab initio* calculations. Gaussian basis sets optimised for use with these potentials are described and validated in Sec. III D.

## II. SHAPE AND ENERGY CONSISTENT CORRELATED PSEUDOPOTENTIALS

### A. Theoretical basis

The pseudopotentials are formulated as the sum of a semilocal operator and a many-body potential,

$$\hat{V}^{pp} = \sum_{0}^{l_{max}-1} \sum_{m=-l}^{l} |Y_{lm}\rangle \left[ V_l^{pp\text{-}cpp} - V_{l_{max}}^{pp\text{-}cpp} \right] \langle Y_{lm}|$$
$$+ V_{l_{max}}^{pp\text{-}cpp} + V^{cpp}, \qquad (1)$$

where each $V_l^{pp\text{-}cpp}$ and $V_{l_{max}}^{pp\text{-}cpp}$ act on the spherical harmonic projection to channels $l < l_{max}$ and $l \geq l_{max}$, respectively. The potential $V^{cpp}$ is the many-body CPP.

The $V_l^{pp\text{-}cpp}$ terms are represented using the parameterised form

$$\tilde{V}_l^{\text{pp-cpp}} = \sum_{q=1}^{6} A_{ql} r^{n_{ql}} e^{-a_{ql}r^2}$$
$$= \begin{cases} V_l^{\text{pp-cpp}} - V_{l_{max}}^{\text{pp-cpp}}, & l \neq l_{max}, \\ Z_v/r + V_{l_{max}}^{\text{pp-cpp}}, & l = l_{max}. \end{cases} \quad (2)$$

For the local channel ($l = l_{max}$), $n_{ql} = -1$ for $q = 1$, $n_{ql} = +1$ for $q = 6$, and $n_{ql} = 0$ otherwise. For other channels, $n_{ql} = 0$. The pseudo-atomic number is $Z_v = Z - Z_c$, where $Z_c$ is the number of core electrons represented by the pseudopotential. The local channel is $l_{max} = max\{l_{occ}\}$, where $\{l_{occ}\}$ is the set of all $l$ values occurring in the electronic configurations used to generate the pseudopotential.

In terms of the parameterised components of the pseudopotential, $\tilde{V}_l^{\text{pp-cpp}}$, the complete pseudopotential is

$$V_l^{\text{pp}} = \begin{cases} \tilde{V}_l^{\text{pp-cpp}} - Z_v/r + \tilde{V}_{l_{max}}^{\text{pp-cpp}} + V^{\text{cpp}}, & l \neq l_{max}, \\ -Z_v/r + \tilde{V}_{l_{max}}^{\text{pp-cpp}} + V^{\text{cpp}}, & l = l_{max}. \end{cases} \quad (3)$$

The $V^{\text{cpp}}$ potential is composed of the sum of a one-body core-electron potential, a two-body electron-electron interaction, and two further terms involving the positions of both electrons and nuclei which are expressed in the form provided by Shirley and Martin.[16] For a single atom with one valence electron, only the local component of the CPP remains,

$$V^{\text{cpp}} = -\frac{1}{2} f^2(r/r_0) \frac{\alpha}{r^4}. \quad (4)$$

The short range truncation function $f(x) = (1 - e^{-x^2})^2$ removes the non-physical singularity at $r = 0$; $r_0$ is the cutoff radii for this function and $\alpha$ is the dipole polarizability of the core.

The previously available CEPPs are parameterised in the same form. An eCEPP candidate is expressed in terms of scaled CEPP parameters,

$$A_{ql} = \lambda_l^A A_{ql}^{CEPP}, \quad (5)$$

$$a_{ql} = \lambda_{ql}^a a_{ql}^{CEPP}, \quad (6)$$

and $r_0$ is varied without scaling, so that the candidate potentials are equal to the CEPP for scaling parameters ($\{\lambda_l^A\}, \{\lambda_{ql}^a\}$) all equal to unity and $r_0 = r_0^{CEPP}$.

Overcompleteness, and the accompanying emergence of undesirable oscillations in the potential, is prevented by the uniform scaling of $\{A_{ql}^{CEPP}\}$ parameters. The CPP cutoff radii, $r_0$, are taken to be a free parameter, while the core polarizability, $\alpha$ in Eq. (4), is taken from previously published highly accurate values available for all the atoms considered.[16,19]

The pseudopotential is required to be finite and has first and second derivatives equal to zero at $r = 0$. Taking these constraints into account results in $8 + 7l_{max}$ free parameters (for first row and $3d$ transition metal atoms $l_{max} = 2$ and 3, resulting in 22 and 29 parameters, respectively).

In order to seek an eCEPP, we start with all scaling parameters equal to unity. Scaling parameters are then varied to search for new values that preserve the CEPP charge density outside of the core region for single-valence electron states and reproduce AE energy differences between atomic states of different symmetry and charge. The relaxation of the one-electron

pseudo charge density within the core region provides the freedom for a candidate potential to reproduce the AE energy differences while preserving the core-valence interaction that defines the CEPP.

We define a penalty function that is zero for an eCEPP and express it as a sum of two terms, $P = P_1 + P_2$. The first of these, $P_1$, measures the degree to which a candidate potential reproduces the one-electron density outside of the core region. This is given by

$$P_1(\{\lambda_l^A\}, \{\lambda_{ql}^a\}, r_0) = \sum_{l=0}^{l_{max}} \left( \left[ \phi_l(r_l^c) - \rho_l^{\frac{1}{2}}(r_l^c) \right]^2 + 1 \right.$$
$$\left. - \left[ \int_{r_l^c}^{\infty} \phi_l \rho_l^{\frac{1}{2}} dr \right]^2 / \left[ \int_{r_l^c}^{\infty} \phi_l^2 dr \int_{r_l^c}^{\infty} \rho_l dr \right] \right), \quad (7)$$

where the sum is taken over the single-valence-electron $^2L$ states corresponding to $l = L$ for $l \leq l_{max}$, and the core radius for each channel, $r_l^c$, defines the sphere within which the one-electron pseudo charge density is allowed to vary.

In Eq. (7), $\rho_l$ is the AE electron density with the core electrons removed, and $\phi_l$ is the one-body wave-function corresponding to the candidate pseudopotential. For the initial parameter values, $\rho_l = \phi_l^2$ and $P_1 = 0$.

Equation (7) is constructed as the simplest form for which $P_1$ is non-negative, the lowest order variation of $P_1$ about the minimum is second order, and $\rho_l = \phi_l^2$ outside of the core region results in $P_1 = 0$.

A minimum with $P_1 = 0$ also occurs for parameter values that are not equal to those for the CEPP. It is these parameters that we seek, as they correspond to the many pseudopotentials for which the one-electron charge densities are equal to the AE valence charge density outside of the core region only.

The second component of the penalty function, $P_2$, measures the degree to which the candidate potential reproduces atomic excitation energies. A set of $N + 1$ states is considered, which is generally composed of the neutral atom ground state, the anion ground state, and several ground and excited states for neutral atoms and ions. We index the set of states using $j$, with $j = 0$ the neutral ground state, and define the second penalty function as

$$P_2(\{\lambda_l^A\}, \{\lambda_{ql}^a\}, r_0) = \frac{1}{\beta^2} \times \sum_{j=1}^{N} \left[ \left( E_j^{\text{AE}} - E_0^{\text{AE}} \right) - \left( E_j^{\text{pp}} - E_0^{\text{pp}} \right) \right]^2, \quad (8)$$

where $E_j^{\text{AE}}$ and $E_j^{\text{pp}}$ are the total energies of the $j$th atomic state evaluated with all electrons present and the candidate pseudopotential, respectively. The prefactor $1/\beta^2$ sets the importance of the second penalty function relative to the first, with $\beta$ setting the accuracy with which the eCEPP is required to reproduce AE total energy differences. (No prefactor is used in $P_1$ as it is expected to be bounded by 0 and ∼1 for all parameter values.)

Minimising the full penalty function, $P(\{\lambda_l^A\}, \{\lambda_{ql}^a\}, r_0)$, with respect to ($\{\lambda_l^A\}, \{\lambda_{ql}^a\}, r_0$) provides the eCEPP that preserves core scattering properties as well as reproducing the AE energy spectrum of an isolated atom.

## B. Implementation

The evaluation of the two components of the penalty function, $P = P_1 + P_2$, is distinct from each other.

The function $P_1$ includes correlation effects if the target charge density, $\rho_l$, is provided by a correlated-electron calculation. We take the target density outside of the core region to be that provided by the CEPP as, by definition, this is equal to the AE charge density provided by Multi-Configuration Hartree-Fock (MCHF)[20,21] calculations with the core contribution removed. Consequently, the target and candidate charge densities outside of the core are both obtained by the direct solution of the one-body Schrödinger equation. The target charge density is provided using the CEPP, while the candidate charge density is provided by the candidate pseudopotential. Each term in $P_1$ is evaluated using summation and numerical integration.

The evaluation of $P_2$ is performed by summation of the squared differences between CCSD(T) energies, with energies evaluated using Gaussian basis sets and extrapolation to the complete basis set (CBS) limit.

All CCSD(T) energies are evaluated using the MOL-PRO[22] code for both AE and candidate eCEPP atoms to provide the target and candidate energies, respectively. The active space is defined to include all electron excitations, including core electrons when they are present. Uncontracted (relativistic) aug-cc-pV$n$Z(-DK)[23,24] basis sets are used to evaluate (AE) energies, where $n$ is the number of correlating functions present in the basis. Differences between total energies, such as electron affinities and ionisation energies, are provided in the CBS limit using extrapolation[11,12,25] and CPP corrections when required. Extrapolating energies from $n = Q$, 5 basis sets and correcting to include CPP energies from $n = T$ basis sets is referred to as "$n = (TQ5)$ extrapolation" (see the supplementary material).

## C. eCEPPs for H, Li–F, Sc–Fe, and Cu

Energy-consistent correlated electron pseudopotentials are constructed for the atoms H, Li–F, Sc–Fe, and Cu. This is the set of atoms for which CEPPs are available, with Cu added due to its many properties and applications. There is no core for H, for Li–F a [He] core is represented by the pseudopotentials, and for Sc–Fe and Cu the core is [Ne]. The initial CEPP parameter values are taken from previous publications, except for Cu for which a CEPP is generated using the same method. Core radii for each pseudopotential channel are the same as for the CEPPs.[11,12] The total of core natural orbital occupation numbers arising when generating CEPPs is close to the total number of core electrons, deviating by $7$–$1 \times 10^{-3}$ and $6$–$2 \times 10^{-3}$ for the first row and $3d$ transition metal atoms, respectively. Across each row, this small deviation decreases monotonically with $Z$ and is correlated with the CEPP (and eCEPP) core radii.

For first row atoms, $P_1$ is defined using the $^2S$, $^2P$, $^2D$ single-valence ground-states, while for the $3d$ transition metal atoms the $^2F$ state is also included.

A finite number of states are used to define $P_2$. Selection rules provide $N + 1 = 5$–8 states for the first row atoms and $N + 1 = 8$–9 states for the $3d$ transition metal atoms (see the supplementary material).

Our target accuracy for the eCEPPs is better than chemical accuracy, so we take the parameter $\beta$ in Eq. (8) to be 43 meV. The combination of parameters, numerical integration for $P_1$, and CCSD(T) data extrapolated to the CBS limit for $P_2$ provides the penalty function for a given set of free parameters, $P(\{\lambda_l^A\}, \{\lambda_{ql}^a\}, r_0)$.

Optimisation of the penalty function is carried out using the Broyden−Fletcher−Goldfarb−Shanno (BFGS) method,[26] with the transformed parameters ($\{\lambda_l^A\}, \{\ln \lambda_{ql}^a\}, r_0$) used to improve stability. Numerical derivatives are defined with a finite difference of $\Delta = 10^{-5}$, the smallest value possible given the precision of the CCSD(T) calculations.

Energies and derivatives for the first row atoms are evaluated using the $n = (TQ5)$ extrapolation. However, for the $3d$ transition metal atoms, a slightly different approach was required. Convergence with a basis set is slow and extrapolation using $n = (TQ5)$ is required to achieve the target accuracy. However, although this choice of basis sets provides CCSD(T) energies converged to the required precision, a significant numerical error in the finite difference prevents the optimisation method from succeeding reliably.

Reliable convergence to the required accuracy is achieved by evaluating the penalty function using $n = (TQ5)$ extrapolation, but evaluating the *gradient* of the penalty function using energies provided by the $n = Q$ basis only (for both AE and candidate eCEPP total energies). This stabilises and accelerates the optimisation process.

The optimisation was terminated for $P < N/400$. This criterion provides a mean absolute agreement between eCEPP and AE energy differences of 3 meV, with a maximum disagreement of 11 meV occurring for ionisation to the $^6S$ state of Cr$^+$. This optimisation error is small when compared with the estimated CCSD(T) basis set error for all atoms.

Optimisation provides an eCEPP for which the penalty function is small but non-zero. This corresponds to the charge density of the single valence electron ion relaxing outside of the core region and not exactly reproducing the target MCHF AE density. The degree of relaxation is small and adequately controlled by the penalty function chosen, with the overlap term differing from unity by $< 2 \times 10^{-4}$ for all atoms and states (the maximum occurs for the $^2D$ state of Sc). The mean absolute error for $\phi(r_l^c)$ is 0.6%, with all errors less than 3.5%.

Pseudopotentials for H are unusual in that the CEPP is equivalent to a norm-conserving HF pseudopotential since neither correlation nor exchange occurs and no CPP arises. Furthermore, only a few bound atomic states are available and the CEPP accurately reproduces the energies of these states. Consequently, for H the CEPP and eCEPP are identical and equal to a norm-conserving HF pseudopotential.

## III. RESULTS

## A. Titanium atom and titanium oxide molecule

Two basic requirements of an accurate pseudopotential are to reproduce AE energy differences between atomic states and to reproduce charge densities outside of the core region.
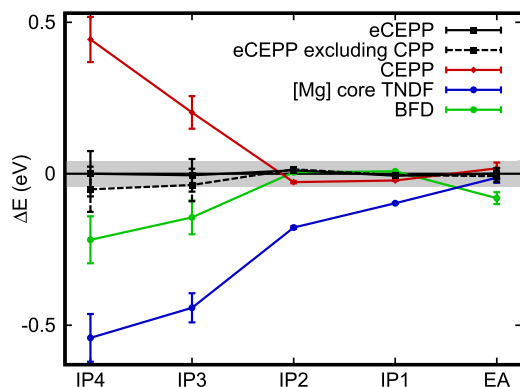
FIG. 1. The difference between atomic Ti ionisation and electron affinities evaluated using pseudopotentials, and those from relativistic AE calculations. The ionisation energies IP4 to IP1 and the electron affinity EA are energy differences between ground state total energies for a total charge of +4 to −1. The grey region indicates the chemical accuracy of 1 kcal mol$^{-1}$ = 43 meV, and the error bars are for the estimated CBS limit. All energies are evaluated using CCSD(T).
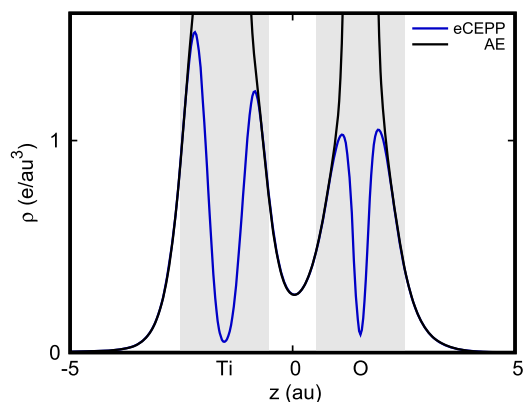


FIG. 2. The charge densities for TiO along the axis containing both nuclear sites. Densities are provided by AE and eCEPP, with CISD. The grey areas contain the core regions for each atom.

Figure 1 shows the deviation of pseudo-atomic energies from AE energies for several Ti pseudopotentials (including the eCEPP). Energy differences shown are ionisation energies and the electron affinity, evaluated as differences between CCSD(T) energies extrapolated to the CBS limit.

The eCEPP energies agree with AE results better than both CBS error and chemical accuracy. This is expected in light of the generation procedure. Invoking the eCEPP with no CPP included introduces a small error due to the low polarizability of the [Ne] core.

The figure also provides a useful comparison of the accuracy of the eCEPP with the Trail-Needs Dirac-Fock[27,28] (TNDF) norm-conserving pseudopotential, the Burkatzki-Filippi-Dolg[29,30] (BFD) energy consistent HF pseudopotential, and the CEPP. For all four pseudopotentials, the CBS error consistently increases with ionisation, as expected given that the aug-cc-pV$n$Z(-DK) basis sets are optimised for neutral atoms. Although the CBS error is largely removed by de-contracting basis sets, the width parameters of the Gaussian basis become increasingly sub-optimal as charge densities contract with increasing total charge.

All three non-eCEPPs show an increase in the pseudopotential error with increasing charge. This trend is caused by the cancellation of errors in energy differences becoming less perfect with increasing energy difference and is particularly apparent for the large ionisation energies where electrons are removed from the $4s$ sub-shell (IP3 and IP4).

Comparing errors arising for all pseudopotential types suggests an interpretation of the errors removed from the eCEPPs. The large errors for the TNDF pseudopotentials suggest that a [Ne] core is necessary for $3d$ transition metal atoms.[31] The large errors for the BFD pseudopotential suggest that energy consistency at the CCSD(T) level of theory is required, and HF is not sufficient. The opposite sign, and similar magnitude, of the errors for the CEPP and BFD pseudopotentials suggests that a combination of CEPP and energy consistency is likely to be successful. The accuracy provided by the eCEPP confirms that it successfully removes these errors.

The eCEPP is not constrained to reproduce the charge density for any system except the single-valence ion, so we compare the AE and pseudo charge densities for a small, neutral molecule. Figure 2 shows such charge densities for the TiO molecule, with an experimental bond length.[32] Densities are evaluated using configuration interaction singles and doubles (CISD) calculations and plotted along the primary symmetry axis. The densities provided by the AE and eCEPP calculations agree well outside of the core region for each atom, demonstrating that good transferability is not limited to atomic ionisation energies.

These results confirm that the eCEPP for titanium accurately reproduces the atomic properties it is designed to recreate and that the pseudopotentials transfer well to the TiO molecule, accurately reproducing AE charge densities outside of the core regions.

## B. Small molecule geometries and dissociation energies

To assess the accuracy of the eCEPPs in reproducing molecular dissociation energies and equilibrium geometries, we consider a moderately large set of small molecules. A test set is constructed starting with the neutral members of the G1 set[33] which contains only the atoms H and Li–F. We then add LiB, LiC, LiF, $H_2$, BH, $Be_2$, $B_2$, $C_2$, and $NO_2$, to give 38 first row molecules (in 39 states). To this set we add the diatomic molecules composed of Sc–Fe and Cu together with H, C, N, O, and F atoms, selecting those for which the ground state term and dominant configuration are not in doubt, as described by Harrison,[32] and for which single-reference CCSD(T) is stable. Titanium is of particular interest,[34,35] so we also include $TiH_4$, $TiO_2$, and three excited states of TiN. The final test set contains 65 molecules in 69 states (see the supplementary material).

Geometry optimisation is performed for each molecule by minimisation of the AE and pseudopotential CCSD(T) energies with respect to bond lengths and bond angles that characterise the geometry of each molecule, $x_i$. All $x_i$ are corrected to include the influences of the CPP, and the basis set error is estimated (see the supplementary material).

The differences between molecular geometry spatial parameters arising from the eCEPP and AE Hamiltonians are shown in Fig. 3. The figure also shows the geometry parameters
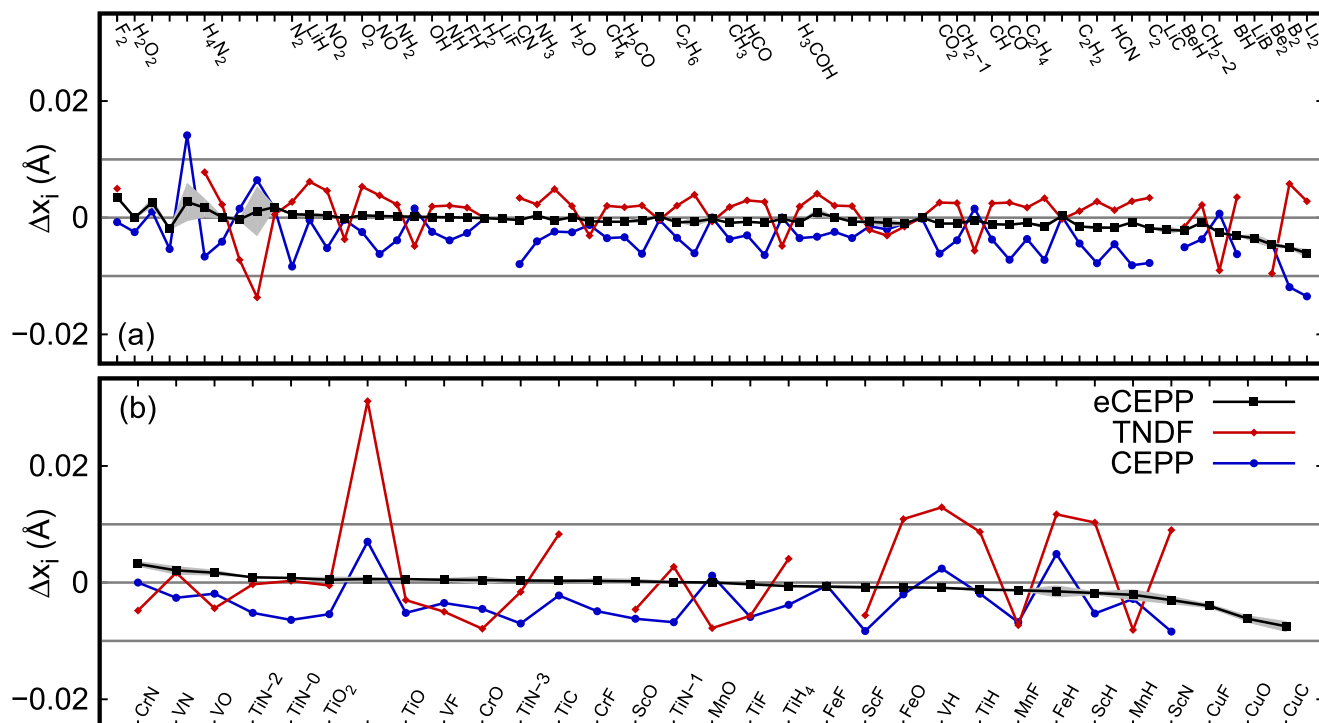
FIG. 3. The difference between optimised molecular geometry parameters, $x_i$, from pseudopotentials and AE CCSD(T) calculations. The top bar shows differences for molecules containing 1st row atoms only, and the bottom bar shows differences for molecules containing $3d$ transition metal atoms. Data for the CEPP and TNDF pseudopotentials are from Refs. 11 and 12. Bond and dihedral angle parameters are expressed as the arc-length on a circle of radius 1.0 Å. Grey lines at ±0.01 Å contain all data points that reproduce AE geometries to within chemical accuracy. Labels absent from the horizontal axes mark further geometry parameters of the molecule labelled immediately to the left.

previously reported for the CEPP and TNDF pseudopotentials when available.

All eCEPP geometry parameters fall within the chemical accuracy (of 0.01 Å) of the relativistic AE values, unlike the results for both the TNDF and CEPP pseudopotentials.

The eCEPP geometry parameters for the mean absolute deviation (MAD) from AE results are ~1/4× that for either the CEPPs or TNDF pseudopotentials. The overall bias is small for the eCEPPs, with an average deviation of $6 \times 10^{-4}$ Å.

For all molecules considered, the estimated basis set error is negligible, except for $H_2O_2$ and $H_4N_2$ for which a particularly shallow Born-Oppenheimer surface magnifies this error for certain angular degrees of freedom. There is a weak correlation between errors for the TNDF pseudopotentials and CEPPs, but no correlation remains in the eCEPP errors.

For molecules containing Li, Be, and B, the bond lengths for the eCEPPs are distinctly underestimated, although by less than chemical accuracy. Given the spatially large and highly polarizable [He] cores of these atoms, it seems likely that the dominant source of this error is the incomplete description of core relaxation provided by the CPP.

The negligible error for $H_2$ (of $1(3) \times 10^{-4}$ Å) confirms that most of the error for other molecules is due to the difficulty of representing the core-valence interaction with a pseudopotential rather than limiting the reproduction of scattering properties to linear order (all three pseudopotentials are equivalent for H).

The improvement in optimised geometries provided by the transition metal eCEPPs is greater than that for the first row atoms. This is particularly apparent for $TiO_2$ for which

the notable error in the bond angle present for the other two pseudopotentials is completely removed.

Of all the transition metal molecules, the three containing Cu show the largest underestimation of bond lengths, which is probably due to the eventual emergence of an expected increase in the pseudopotential error with the increasing number of valence electrons.

The differences between molecular dissociation energies evaluated using CCSD(T) for the eCEPP and AE Hamiltonians are shown in Fig. 4. The figure also shows the dissociation energies previously reported for the CEPP and TNDF pseudopotentials when available.

All eCEPP dissociation energies fall within the chemical accuracy (of 43 meV) of the relativistic AE values, unlike the results for the TNDF and CEPP pseudopotentials. This improvement in accuracy is greater than for optimised geometries, with the MAD for an eCEPP ~1/10× that for TNDF pseudopotentials. For all molecules considered, including $H_2O_2$ and $H_4N_2$, the estimated basis set error is negligible. Unlike the optimised geometries, errors in dissociation energies for molecules containing Li, Be, or B are not unusually large and the CPP contribution to energies is accurate. The overall bias is small for the eCEPPs when compared with other pseudopotentials, with an average deviation of 2 meV.

There is only a weak correlation between the errors for the TNDF and CEPP potentials, and no correlation is apparent with the eCEPP errors. No consistent trend is apparent for the eCEPPs except for a slow increase in error with the number of valence electrons. As for optimised geometries, the absence of the core-valence electron interaction for $H_2$
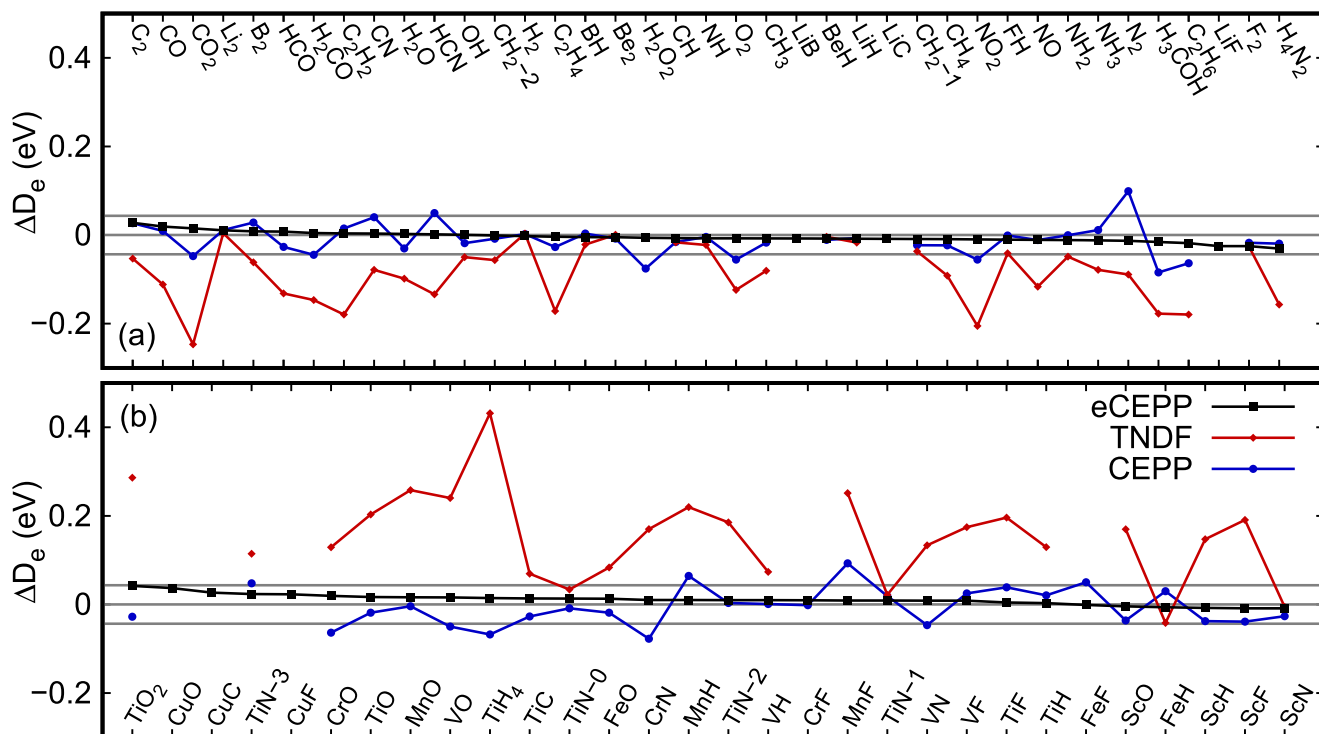
FIG. 4. The difference between molecular dissociation energies, $D_e$, from pseudopotential and AE CCSD(T) calculations. The top bar shows differences for molecules containing 1st row atoms only, and the bottom bar shows differences for molecules containing $3d$ transition metal atoms. Data for the CEPP and TNDF pseudopotentials are from Refs. 11 and 12. The grey lines at ±43 meV contain all data points that reproduce AE energies to within chemical accuracy.

results in a negligible error [of 0.043(4) meV] for all three pseudopotentials.

Mean, absolute mean, and maximum deviations from AE results for the geometry parameters and dissociation energies evaluated with the three pseudopotential types are shown in Table I.

In order to assess the energetics absent from CCSD(T), the available monoxide molecules are considered in more detail. Table II shows the equilibrium bond lengths and dissociation energies evaluated with all electrons present, and with core electrons represented by the eCEPPs. We compare this with the experimental data selected by Miliordos *et al.* for ScO–FeO[36] and Harrison for CuO.[32] Results are also quoted for TiO[37] and CrO,[38] calculated using a composite method that corrects accurate CCSD(T) energies to include spin-orbit coupling and higher-order correlation.

TABLE I. Summary of pseudopotential errors for TNDF pseudopotentials, CEPPs, and eCEPPs. Values are for the mean, absolute mean, and maximum error taken over the test set, denoted as $\overline{\Delta x_i}$, $\overline{|\Delta x_i|}$, and Max[$\Delta x_i$] for geometry parameters, and $\overline{\Delta D_e}$, $\overline{|\Delta D_e|}$, and Max[$\Delta D_e$] for dissociation energies. The test set of CEPPs does not include LiB, LiC, LiF, CuC, CuO, or CuF as these molecules were not included in the previous work. The test set for the TNDF pseudopotentials excludes the same molecules, as well as $H_2O_2$, CrF, and FeF due to lack of convergence.

| Pseudopotential | ($\times 10^{-3}$ Å) | | | (meV) | | |
| --- | --- | --- | --- | --- | --- | --- |
| | $\overline{\Delta x_i}$ | $\overline{|\Delta x_i|}$ | Max[$\Delta x_i$] | $\overline{\Delta D_e}$ | $\overline{|\Delta D_e|}$ | Max[$\Delta D_e$] |
| TNDF | 1.2 | 4.3 | 31.1 | 13.65 | 117.05 | 431.70 |
| CEPP | −3.3 | 4.1 | 14.1 | −8.89 | 30.72 | 99.11 |
| eCEPP | −0.6 | 1.2 | −7.5 | 1.67 | 11.04 | 42.02 |

The disagreement between AE and eCEPP results is consistently negligible compared with the significant disagreement with experimental results, for both geometries and energies.

Bond lengths agree with experiment to within 0.01 Å for all oxides except CrO and MnO, with the largest underestimate of −0.016 Å occurring for CrO. Chan *et al.* provide a CCSD(T) CrO bond length closer to experiment, but this agreement is due to the small cc-pwCVTZ basis set and disappears with a more complete basis. This suggests that when an error in bond length occurs for these oxides, it is primarily due to missing energetics rather than the pseudopotential error.

Only the ScO dissociation energy (for AE or eCEPP) agrees with experiment to within 43 meV. The maximum error, of −0.44 eV, occurs for FeO, with CCSD(T) overestimating $D_e$ for ScO–VO and underestimating for CrO–CuO. The composite dissociation energies for TiO and CrO are in much better agreement with experiment, with most of the improvement provided by spin-orbit corrections. Higher-order correlation corrections are of secondary importance for these two molecules.

So far we have not considered the accuracy with which the pseudopotentials reproduce molecular potential energy surfaces beyond the minima, that is, energy differences considerably smaller than dissociation energies. Hydrogen peroxide is of particular interest due to the relatively large error in equilibrium spatial variables apparent in Fig. 3. This is primarily due to the shallow variation of the total energy with the rotation of O–$H_2$ groups around the O–O axis, an internal rotation that plays an important role in conformational analysis.[39]

TABLE II. Bond lengths ($r_e$) and dissociation energies ($D_e$) for monoxides of Sc–Fe and Cu. Results are shown evaluated using CCSD(T) with all electrons present and with eCEPPs. Experimental values from the literature are shown for all oxides, and previously published composite results for TiO and CrO are evaluated by correcting CCSD(T) energies by including spin-orbit coupling and higher-order correlation effects [with geometries optimised using CCSD(T) only].

| Method | ScO | TiO | VO | CrO | MnO | FeO | CuO |
|---|---|---|---|---|---|---|---|
| | | | $r_e$ (Å) | | | | |
| eCEPP | 1.669(6) | 1.620(3) | 1.584(5) | 1.606(3) | 1.633(5) | 1.616(2) | 1.730(2) |
| AE | 1.669(7) | 1.619(3) | 1.582(6) | 1.605(5) | 1.633(5) | 1.617(2) | 1.736(3) |
| Composite | . . . | 1.619[a] | . . . | 1.6185[b] | . . . | . . . | . . . |
| Expt.[c] | 1.6656 | 1.6203 | 1.589 | 1.6213 | 1.6477 | 1.6194 | 1.724 |
| | | | $D_e$ (eV) | | | | |
| eCEPP | 7.019(8) | 7.097(10) | 6.765(10) | 4.540(5) | 3.712(9) | 3.804(3) | 2.813(2) |
| AE | 7.024(11) | 7.080(11) | 6.749(10) | 4.521(4) | 3.696(12) | 3.791(5) | 2.776(3) |
| Composite | . . . | 6.916[a] | . . . | 4.660[b] | . . . | . . . | . . . |
| Expt.[c] | 6.981(9) | 6.93(7) | 6.55(9) | 4.63(7) | 3.88(8) | 4.235 | 2.928 |
| | | | | 4.80(9) | 3.8(2) | | |

[a]From Ref. 37. Bond lengths are optimised using CCSD(T), and dissociation energies are CCSD(T) corrected to include spin-orbit coupling and higher-order correlation.
[b]From Ref. 38. As Ref. 37.
[c]Sum of experimental atomisation energies and zero-point vibrational energies. Data as cited by Ref. 36 for Sc–Fe and Ref. 32 for Cu.

Geometries for $H_2O_2$ are optimised at the CCSD(T) level of theory using the uncontracted aug-cc-pVTZ(-DK) basis sets (for AE calculations), with the dihedral angle between O–$H_2$ groups, $\phi$, held constant. Relaxed geometries are evaluated for several values of $\phi$, followed by an extrapolation of the total energy to the CBS limit using $n = (TQ5)$. Energy profiles evaluated with AEs present and using eCEPP pseudopotentials are shown in Fig. 5. Agreement is good, with a maximum pseudopotential and basis set error of 2.5 and 2.7 meV, respectively, both occurring for $\phi = 0$.

Except for TiO and CrO, no attempt has been made to estimate the error due to correlation neglected at the CCSD(T) level of theory. While this error has been quantified as small when compared with basis set errors for first-row molecules,[40] it is expected to be significant for some molecules containing $3d$ transition metal atoms.[41] However, this "missing



FIG. 5. Energy profiles for the internal rotation of O–$H_2$ relative to each other in $H_2O_2$. Results are shown from CCSD(T) calculations with all electrons present and with core electrons represented by the eCEPPs. The grey region indicates the error due to extrapolation to the CBS limit for both AE and eCEPP calculations.

correlation" is usually found to be between valence electrons in molecules, rather than isolated atoms, suggesting that the accuracy of the eCEPPs will be preserved for calculations including correlation beyond that present in CCSD(T).

## C. Comparison of errors for a theory hierarchy

An important application of the eCEPPs is expected to be QMC methods for bulk systems, particularly Diffusion Monte Carlo (DMC). DMC requires a fermionic nodal surface to be supplied, and for bulk systems this is normally approximated as the nodal surface of a determinant constructed from orbitals resulting from a KS-DFT calculation with a plane-wave basis.

In light of this, we begin with an AE CCSD(T), then quantify and compare the error introduced by each approximation involved in an eCEPP KS-DFT calculation with a plane-wave basis set. The purpose of this analysis is to compare the relative size of the various components of the error in the final calculation.

If the error due to the eCEPP alone is small when compared with other errors, then it may be expected that the contribution of the eCEPP to the error in the approximate nodal surface is also small. While this measure of the accuracy of the nodal surface is not rigorous (an exact exchange-correlation functional would not provide an exact nodal surface, and, in principle, an inaccurate pseudopotential or exchange-correlation functional could provide a more accurate nodal surface), it is physically reasonable that less accurate KS-DFT molecular properties will correspond to a less accurate nodal surface.

A separation of errors naturally arises by considering the application of several levels of approximation. The most accurate description considered is AE CCSD(T) with a Gaussian basis. The next level of approximation is to replace atomic cores with the full eCEPP. We then increase the level of
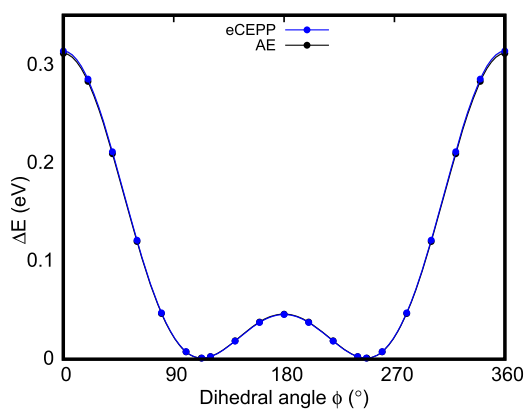
approximation progressively by removing the CPP component of the eCEPP, followed by replacing CCSD(T) with KS-DFT, and finally by replacing the full eCEPP with a Kleinman-Bylander[42] representation (of the eCEPP) together with a plane-wave basis. Both KS-DFT calculations are performed using the Perdew-Burke-Ernzerhof (PBE) exchange-correlation functional.[43]

The KB pseudopotential representation of the eCEPP is given by

$$\hat{V} = V^{pp}_{l_{max}} + \sum_l \sum_{ij} |\delta V_l \phi_{li}\rangle \, B_{l,ij} \, \langle \phi_{lj} \delta V_l|, \qquad (9)$$

with $\delta V_l = V^{pp}_l - V^{pp}_{l_{max}}$. The $\phi_{li}$ functions used for each projector are supplied for each channel of each pseudopotential, and the matrices for each channel are given by $B_{l,ij} = \langle \phi_{li}|\delta V_l|\phi_{lj}\rangle^{-1}$. This separable form is less accurate than the semi-local spherical harmonic projector representation used to define the eCEPPs. However, the KB representation provides a considerable reduction in computational cost, and the additional error introduced is often small.

Three projection orbitals are used for first row atoms, and six projection orbitals for the $3d$ transition metal atoms. For all atoms, the local channel is taken to be that with the largest $l$ value, $l_{max}$. In each case, orbitals are taken from the neutral atom and supplemented by orbitals from the lowest-energy excited states that provide projectors for all pseudopotential channels (see the supplementary material). For example, for Ti the ground state provides projectors from the $3s$, $4s$, $3p$, and $3d$ orbitals, to which we add $4p$ and $4f$ projectors from the $[Ar]3d^2 4s 4p$ and $[Ar]3d^2 4f$ states, respectively.

Note that projectors for $l = l_{max}$ are evaluated but unused. These are provided to allow flexibility in the choice of the local channel of the KB representation, should it be required. All projectors are generated using the PBE functional and the atomic KS-DFT code contained in the QUANTUM ESPRESSO package.[44]

Errors are assessed for the subset of diatomic molecules present in the full test set, excluding the excited TiN states. Bond lengths are relaxed and dissociation energies are evaluated for each of these molecules and at each level of theory.

All electron and eCEPP results with the CPP are identical to those included in Sec. III B. Dissociation energies with the CPP removed are evaluated with the same data but excluding CPP corrections.

Results for KS-DFT without the KB representation are generated using the uncontracted aug-pp-pV5Z Gaussian basis sets. Extrapolation is not required as exponential convergence results from correlation being contained in the exchange-correlation functional, rather than the wave function itself. Estimated basis set error is <8 meV, from differences between dissociation energies evaluated using the $n = (Q5)$ basis sets.

For KS-DFT with a plane-wave basis set, the isolated molecule is modelled by a single molecule in a $(15 \, \text{Å})^3$ periodic unit cell, and using the CASTEP[45] plane-wave pseudopotential code (with a small modification to allow for more than one projector per pseudopotential channel). Calculations for all atoms and molecules are performed using the plane-wave basis set defined by an energy cutoff of 150 ha, resulting in a basis set error of <9 meV. The maximum error occurs for CuF due to the number of valence electrons present and the depth of the Cu and F potentials. The same accuracy could be achieved with considerably smaller basis sets for most of the molecules in our test set.

The total error arising from all approximations taken together is

$$\Delta = D_e^{\text{pp-cpp,DFT+KB}} - D_e^{\text{AE,CCSD(T)}}$$
$$= \Delta_{\text{pp}} + \Delta_{\text{cpp}} + \Delta_{\text{DFT}} + \Delta_{\text{KB}}. \qquad (10)$$

Components of the total error due to pseudopotentials, the absent CPP, density functional theory (with PBE), and the KB representation are given by
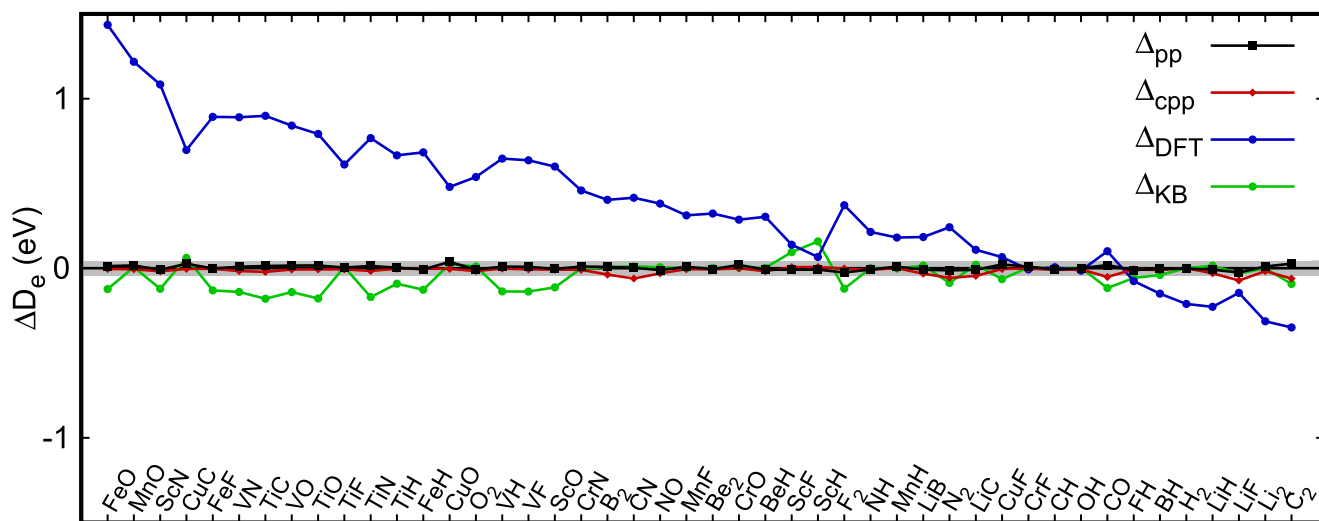


FIG. 6. A breakdown of the errors in dissociation energies for diatomic molecules. Errors are shown arising from the eCEPPs themselves ($\Delta_{pp}$), from excluding core polarisation potentials ($\Delta_{cpp}$), from employing KS-DFT (with PBE exchange-correlation) ($\Delta_{DFT}$), and from using a Kleinman-Bylander representation for the eCEPPs ($\Delta_{KB}$). Differences within the grey region are less than 43 meV.

$$\Delta_{pp} = D_e^{pp,CCSD(T)} - D_e^{AE,CCSD(T)},$$

$$\Delta_{cpp} = D_e^{pp\text{-}cpp,CCSD(T)} - D_e^{pp,CCSD(T)},$$

$$\Delta_{DFT} = D_e^{pp\text{-}cpp,DFT} - D_e^{pp\text{-}cpp,CCSD(T)},$$

$$\Delta_{KB} = D_e^{pp\text{-}cpp,DFT+KB} - D_e^{pp\text{-}cpp,DFT}, \qquad (11)$$

respectively. The superscripts for each dissociation energy in Eq. (11) describe the level of approximation. Superscripts "AE," "pp," and "pp-cpp" denote calculations including all electrons, describing core electrons using an eCEPP, and describing core electrons using an eCEPP with core polarisation removed. Superscripts "CCSD(T)," "DFT," and "DFT + KB" denote the use of CCSD(T), KS-DFT, and KS-DFT with a KB representation of the eCEPPs.

The error breakdown is shown in Fig. 6. The overall trend is for KS-DFT errors to be dominant when compared with the three errors associated with the pseudopotential. The negligible error that arises for the CrF, CH, and OH molecules is unlikely to be more than fortuitous. The largest KS-DFT errors arise for molecules containing transition metal atoms, which is unsurprising given the strong correlation and highly inhomogeneous charge densities for these molecules.

Errors due to the absence of a CPP are almost negligible compared to the KS-DFT error. Molecules containing Li, with the most polarizable core, show the largest contribution from the CPP, although the net polarizability stands out as particularly large for CN, $N_2$, CO, and $C_2$.

The KB representation results in a relatively small error. No overall pattern is apparent except that the KB representation is significantly less accurate for molecules containing transition metal atoms, and the error is the greatest in TiC. This is as expected, since the non-local and local channels differ the most for these atoms, with the eCEPPs composed of deep $d$ channels and relatively shallow $s$, $p$, and $f$ channels.

Physically unrealistic "ghost" eigenstates can arise when the KB representation is used.[46] Agreement between KS-DFT energy differences evaluated with and without the KB representation (see Fig. 6) provides a useful check that such ghost states do not occur for eCEPPs with the local channel chosen as $l = l_{max}$, and this is further confirmed by good agreement between total energies (not shown). This absence of ghost states for the $3d$ transition metals is very unlike the TNDF pseudopotentials[47] and can be ascribed to the presence of a physically realistic eCEPP channel for $l \geq 3$.

The breakdown of errors in bond lengths (not shown) shows a similar behaviour, except that the error due to the absence of a CPP is comparable to the KS-DFT error for molecules containing Li, Be, and B, and the KB error is dominant for ScH and ScF. Unlike the dissociation energies, bond lengths for CrF, CH, and OH are not notably accurate.

### D. Optimised basis sets

Contracted correlation consistent Gaussian basis sets are well established to be computationally efficient for correlated electron calculations, providing a controlled convergence behaviour that allows accurate extrapolation to the CBS limit.

The eCEPP pseudo atom wave-functions are very different from AE wave-functions in the core region; hence such a basis set must be reconstructed for each eCEPP. We name basis sets optimised for the eCEPPs as "aug-cc-pV$n$Z-eCEPP." As for the AE case, these are constructed by optimising a Gaussian basis within HF, systematically introducing correlation consistent components of the basis at higher angular momentum quantum numbers, and augmenting the basis with further Gaussians optimised to accurately represent the more diffuse anionic state.

The basis sets are optimised using the procedure developed for first row atoms by Xu $et$ $al.$,[48] and the $3d$ transition metal approach of Balabanov and Peterson[24] adapted to cores described by pseudopotentials (see the supplementary material).

In the presence of pseudopotentials, valence states become smoother and core states are absent; hence fewer Gaussians and contractions are required than for the AE case. This results in a significant improvement in computational efficiency, and the average CCSD(T) run-time is ~1/5× that for uncontracted standard basis sets. This speedup should increase with the molecular size.

Figure 7 shows the errors in eCEPP CCSD(T) dissociation energies evaluated using both the $uncontracted$ standard aug-cc-pV$n$Z basis set and the new $contracted$ aug-cc-pV$n$Z-eCEPP basis set. The error is quantified as the deviation from the baseline of AE CCSD(T) energies evaluated using the $uncontracted$ standard aug-cc-pV$n$Z-DK basis set.

For molecules containing first row atoms, the new contracted basis sets reproduce the uncontracted aug-cc-pV$n$Z basis set errors accurately. Molecules containing $3d$ transition metal atoms show a less consistent agreement. However, the additional error is small and marginally greater than chemical accuracy only for the three molecules containing copper. For most cases, the error due to the contracted aug-cc-pV$n$Z-eCEPP basis is similar or less than that due to the pseudopotential itself.

In Fig. 7 we also show the error introduced to AE CCSD(T) by using $contracted$ standard aug-cc-pV$n$Z-DK basis sets. This is quantified as the difference between dissociation energies evaluated using the $contracted$ and $uncontracted$ aug-cc-pV$n$Z-DK basis sets, with the latter being the more complete basis, which provides a useful scale for assessing the accuracy of the aug-cc-pV$n$Z-eCEPP basis sets.

The basis set error introduced by contraction is significantly greater for the AE calculations than for the eCEPP calculations (except for $H_2$ for which both are negligible), especially for the molecules containing $3d$ transition metal atoms. We conclude that the construction of aug-cc-pV$n$Z-eCEPP basis sets and application with the eCEPPs are successful, with an impressively low error for the test set considered.

Finally, we note that the relatively large error in the relativistic AE dissociation energy calculated with the contracted aug-cc-pV$n$Z-DK basis sets is not unexpected and is not a deficiency of these basis sets $per$ $se$. These standard contractions are constructed to represent correlations between electrons outside of a [He] core for Li–F and an [Ar] core for Sc–Cu, so they do not describe correlation between these cores and valence electrons well, despite the CCSD(T) active space including excitation of all electrons.
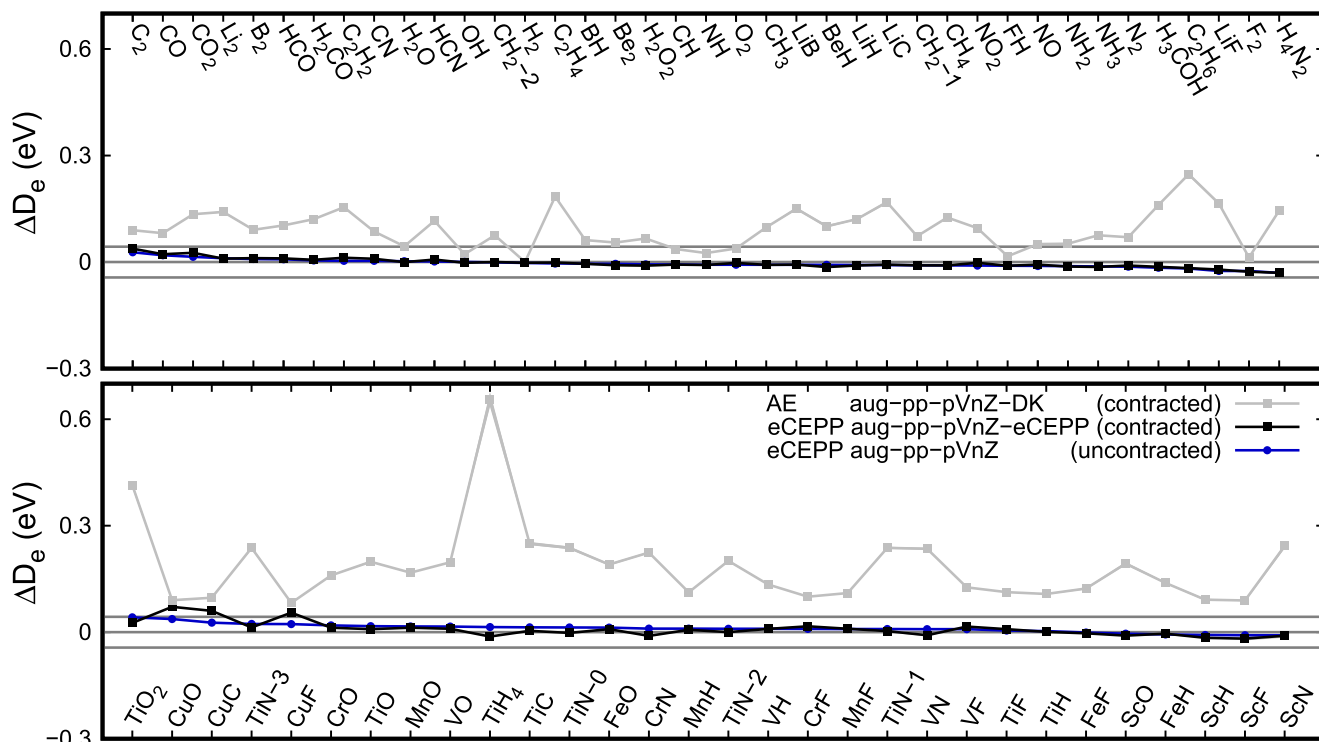
FIG. 7. Comparison of errors in CCSD(T) dissociation energies for different basis sets, evaluated using both AE and eCEPPs. Errors are estimated as the deviation from the AE $D_e$ provided by CCSD(T) with uncontracted aug-pp-pV$n$Z-DK basis sets. Blue line shows the errors in eCEPP energies evaluated using the uncontracted aug-pp-pV$n$Z basis sets. Black line shows the errors in eCEPP energies evaluated using the contracted basis set optimised for eCEPPs, aug-pp-pV$n$Z-eCEPP. Grey line shows the error for AE dissociation energies evaluated using the contracted aug-pp-pV$n$Z-DK basis sets. Grey lines at ±43 meV contain all data points that reproduce AE energies to within chemical accuracy.

This explanation is consistent with the very small error achieved for $H_2$, with the larger errors for the $3d$ transition metal molecules, and the accuracy of the aug-cc-pwCV$n$Z basis sets[24] for which contractions include core-valence correlation effects.

## IV. CONCLUSIONS

The paradigms of shape and energy consistency in the definition of electronic pseudopotentials are successfully and consistently combined to generate accurate effective potentials. The resulting eCEPPs reproduce core electron scattering properties to first order, reproduce atomic excitation and ionisation energies, and partially reproduce core relaxation effects via a dipole polarisation potential. This approach is an extension of the CEPPs (that reproduce only core electron scattering and core polarisation). Electron correlation is included throughout, no recourse is made to independent electron orbitals, and the generation process is *ab initio*.

Comparing AE and eCEPPs results for relaxed geometries and dissociation energies, and for a test set of 65 molecules (in 69 states) at the CCSD(T) level of theory, demonstrates a reduction in error by ~1/4× for geometry parameters and ~1/10× in dissociation energies when compared with Hartree-Fock pseudopotentials.

When compared with the previously published CEPPs, the eCEPPs introduce an interpolation of the effect of the core on several atomic states in addition to the extrapolation of

core scattering. This extension is significant, reducing errors in geometry parameters and dissociation energies by a factor of ~1/4× and ~2/5×.

The eCEPPs will primarily be useful for correlated wave-function methods, particularly for a CCSD(T) description of molecular systems, and for the application of QMC methods to extended systems. For both cases, pseudopotentials are often unavoidable and the eCEPPs provided here are validated by small errors for a reasonably large test set.

The improved accuracy is particularly important when $3d$ transition metal atoms are present as the eCEPP error is sufficiently small to resolve the effect of spin-orbit coupling and aspects of electron-electron correlation often absent in *ab initio* many-body methods.

Further work naturally suggests itself. It would be useful and desirable to extend the eCEPP generation process to more atomic types. Such an extension is however limited by the size and stability of MCHF and CCSD(T) calculations and the absence of correlation consistent basis sets for heavier atoms. However, completing the H–Kr set of atoms can be expected to be straightforward.

It is perhaps more interesting to consider the accuracy of the eCEPPs and how much further improvement is possible. In principle, the accuracy could be improved by extending the reproduction of core scattering properties to higher order and many valence electrons. As discussed in the Introduction, implementing this is non-trivial as it will result in a many-body effective potential.

Another possible strategy would be to extend the reproduction of atomic ionisation and excitation energies to valence charge densities outside of the core region. Given the success of the eCEPP in reproducing these densities for TiO, such an extension appears to be unnecessary.

The most direct approach to improve the eCEPP accuracy lies in the choice of excited/ionic states for which energy consistency is enforced. Here we have considered the atomic states with the lowest energy. A more detailed consideration of the molecular bonds each atom is prone to form should, in principle, allow a selection of atomic states which are more appropriate for each atom. This could also be extended to molecular geometries and dissociation energies. The primary drawback of this approach (and why it was not used here) is that the pseudopotential generation process would become more biased.

A simple, and possibly effective, approach to reduce pseudopotential error would be a more careful choice of core radii. These radii parameterise a balance between errors due to defining the eCEPP with a finite core charge density outside of the core region and errors due to replacing the valence charge density with a pseudo density within the core region. Optimised core radii, different from those used in this paper, may well decrease errors significantly. This is likely to be most influential for the lightest atoms with the most diffuse cores.

The most important issue that has not been addressed is the accuracy of the eCEPPs when used in QMC calculations. CCSD(T) results are accurate, and DMC is known to provide a similar or better accuracy than CCSD(T). However, DMC involves a further "pseudopotential localisation" approximation that expresses the semi-local one-body pseudopotential(s) as a local many-body potential via a trial wave-function. The resulting error is second order in the accuracy of the trial wave-function.[5,6] A measure of this error would be desirable, and the molecules considered here are useful candidates for such tests.

## SUPPLEMENTARY MATERIAL

See supplementary material for details of CCSD(T) calculations, atomic and molecular states, KB projectors, eCEPP basis set optimisation, and the eCEPPs themselves.

## ACKNOWLEDGMENTS

[1]M. Dolg and X. Cao, Chem. Rev. **112**, 403 (2012).

[2]V. Milman, B. Winkler, J. A. White, C. J. Pickard, M. C. Payne, E. V. Akhmatskaya, and R. H. Nobes, Int. J. Quantum Chem. **77**, 895 (2000).

[3]I. Shavitt and R. J. Bartlett, *Many-Body Methods in Chemistry and Physics* (Cambridge University Press, Cambridge, UK, 2009).

[4]D. M. Ceperley and B. J. Alder, Phys. Rev. Lett. **45**, 566 (1980).

[5]W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, Rev. Mod. Phys. **73**, 33 (2001).

[6]R. J. Needs, M. D. Towler, N. D. Drummond, and P. López Ríos, J. Phys.: Condens. Matter **22**, 023201 (2010).

[7]B. M. Austin, D. Yu. Zubarev, and W. A. Lester, Jr., Chem. Rev. **112**, 263 (2012).

[8]D. M. Ceperley, J. Stat. Phys. **43**, 815 (1986).

[9]A. Ma, N. D. Drummond, M. D. Towler, and R. J. Needs, Phys. Rev. E **71**, 066704 (2005).

[10]J. Mitroy, M. S. Safronova, and C. W. Clark, J. Phys. B: At., Mol. Opt. Phys. **43**, 202001 (2010).

[11]J. R. Trail and R. J. Needs, J. Chem. Phys. **139**, 014101 (2013).

[12]J. R. Trail and R. J. Needs, J. Chem. Phys. **142**, 064110 (2015).

[13]D. R. Hamann, M. Schlüter, and C. Chiang, Phys. Rev. Lett. **43**, 1494 (1979).

[14]P. H. Acioli and D. M. Ceperley, J. Chem. Phys. **100**, 8169 (1994).

[15]E. R. Davidson, Rev. Mod. Phys. **44**, 451 (1972).

[16]E. L. Shirley and R. M. Martin, Phys. Rev. B **47**, 15413 (1993).

[17]W. Müller, J. Flesch, and W. Meyer, J. Chem. Phys. **80**, 3297 (1984); W. Müller and W. Meyer, *ibid.* **80**, 3311 (1984).

[18]E. L. Shirley, D. C. Allan, R. M. Martin, and J. D. Joannopoulos, Phys. Rev. B **40**, 3652 (1989).

[19]S. H. Patil, J. Chem. Phys. **83**, 5764 (1985).

[20]C. Froese Fischer, T. Brage, and P. Jönsson, *Computational Atomic Structure: An MCHF Approach* (IOP Publishing Ltd., 1997).

[21]C. Froese Fischer, G. Tachiev, G. Gaigalas, and M. Godefroid, Comput. Phys. Commun. **176**, 559 (2007); A. Borgoo, O. Scharf, G. Gaigalas, and M. Godefroid, *ibid.* **181**, 426 (2010).

[22]H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, and M. Schtz, Wiley Interdiscip. Rev.: Comput. Mol. Sci. **2**, 242 (2012); H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, P. Celani, T. Korona, R. Lindh, A. Mitrushenkov, G. Rauhut, K. R. Shamasundar, T. B. Adler, R. D. Amos, A. Bernhardsson, A. Berning, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, E. Goll, C. Hampel, A. Hesselmann, G. Hetzer, T. Hrenar, G.Jansen, C. Köppl, Y. Liu, A. W. Lloyd, R. A. Mata, A. J. May, S. J. McNicholas, W. Meyer, M. E. Mura, A. Nicklass, D. P. O'Neill, P. Palmieri, D. Peng, K. Pflüger, R. Pitzer, M. Reiher, T. Shiozaki, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson, and M. Wang, MOLPRO, version 2012.1, a package of *ab initio* programs, 2012, see http://www.molpro.net.

[23]T. H. Dunning, Jr., J. Chem. Phys. **90**, 1007 (1989); R. K. Kendall, T. H. Dunning, and R. J. Harrison, *ibid.* **96**, 6796 (1992); K. A. Peterson and T. H. Dunning, Jr., *ibid.* **117**, 10548 (2002).

[24]N. B. Balabanov and K. A. Peterson, J. Chem. Phys. **123**, 064107 (2005).

[25]D. Feller, K. A. Peterson, and J. Grant Hill, J. Chem. Phys. **133**, 184102 (2010).

[26]C. T. Kelley, *Iterative Methods for Optimization* (Society for Industrial and Applied Mathematics, 1999).

[27]J. R. Trail and R. J. Needs, J. Chem. Phys. **122**, 174109 (2005).

[28]See http://www.tcm.phy.cam.ac.uk/~mdt26/casino2_pseudopotentials.html for the full set of TNDF pseudopotentials, including those generated with medium sized cores.

[29]M. Burkatzki, C. Filippi, and M. Dolg, J. Chem. Phys. **126**, 234105 (2007); **129**, 164115 (2008).

[30]See http://www.burkatzki.com/pseudos/index.2.html for the full set of BFD pseudopotentials and basis sets.

[31]L. F. Pacios and P. G. Calzada, Int. J. Quantum Chem. **34**, 267 (1988).

[32]J. F. Harrison, Chem. Rev. **100**, 679 (2000).

[33]J. A. Pople, M. Head-Gorden, D. J. Fox, K. Raghavachari, and L. A. Curtiss, J. Chem. Phys. **90**, 5622 (1989).

[34]T. Jones and T. A. Egerton, *Titanium Compounds, Inorganic. Kirk-Othmer Encyclopedia of Chemical Technology* (John Wiley & Sons, Inc., 2012).

[35]J. R. Trail, B. Monserrat, P. L. Ríos, R. Maezono, and R. J. Needs, Phys. Rev. B **95**, 121108 (2017).

[36]E. Miliordos and A. Mavridis, J. Phys. Chem. A **111**, 1953 (2007); **114**, 8536 (2010); C. N. Sakellaris, E. Miliordos, and A. Mavridisa, J. Chem. Phys. **134**, 234308 (2011).

[37]Y. Pan, Z. Luo, Y.-C. Chang, K.-C. Lau, and C. Y. Ng, J. Phys. Chem. A **121**, 669 (2017).

[38]B. Chan, A. Karton, K. Raghavachari, and L. Radom, J. Chem. Theory Comput. **8**, 3159 (2012).

[39]L. Sing, M. Liu, W. Wu, Q. Zhang, and Y. Mo, J. Chem. Theory Comput. **1**, 394 (2005).

[40]D. Feller and K. A. Peterson, J. Chem. Phys. **126**, 114105 (2007).

[41]W. Jiang, W. J. DeYonker, and A. K. Wilson, J. Chem. Theory Comput. **8**, 460 (2012).

[42]L. Kleinman and D. M. Bylander, Phys. Rev. Lett. **48**, 1425 (1982).

[43]J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996); **78**, 1396 (1997).

[44]P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. D. Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari, and R. M. Wentzcovitch, J. Phys.: Condens. Matter **21**, 395502 (2009), http://www.quantum-espresso.org.

[45]S. J. Clark, M. D. Segall, C. J. Pickard, P. J. Hasnip, M. J. Probert, K. Refson, and M. C. Payne, Z. Kristallogr. **220**, 567 (2005).

[46]X. Gonze, R. Stumpf, and M. Scheffler, Phys. Rev. B **44**, 8503 (1991).

[47]N. D. Drummond, J. R. Trail, and R. J. Needs, Phys. Rev. B **94**, 165170 (2016).

[48]J. Xu, M. J. Deible, K. A. Peterson, and K. D. Jordan, J. Chem. Theory Comput. **9**, 2170 (2013).