

Ab Initio Molecular Simulations with Numeric Atom-Centered Orbitals

Volker Blum^{a,1}, Ralf Gehrke^a, Felix Hanke^a, Paula Havu^a,
Ville Havu^a, Xinguo Ren^a, Karsten Reuter^a,
Matthias Scheffler^{a,1}

^a*Fritz-Haber-Institut der Max-Planck-Gesellschaft, Faradayweg 4-6, D-14195
Berlin, Germany*

Abstract

We describe a complete set of algorithms for *ab initio* molecular simulations based on numerically tabulated atom-centered orbitals (NAOs) to capture a wide range of molecular and materials properties from quantum-mechanical first principles. The full algorithmic framework described here is embodied in the Fritz Haber Institute “*ab initio* molecular simulations” (FHI-aims) computer program package. Its comprehensive description should be relevant to any other first-principles implementation based on NAOs. The focus here is on density-functional theory (DFT) in the local and semilocal (generalized gradient) approximations, but an extension to hybrid functionals, Hartree-Fock theory, and MP2 / *GW* electron self-energies for total energies and excited states is possible within the same underlying algorithms. An all-electron / full-potential treatment that is both computationally efficient and accurate is achieved for periodic and cluster geometries on equal footing, including relaxation and *ab initio* molecular dynamics. We demonstrate the construction of transferable, hierarchical basis sets, allowing the calculation to range from qualitative tight-binding like accuracy to meV-level total energy convergence with the basis set. Since all basis functions are strictly localized, the otherwise computationally dominant grid-based operations scale as $O(N)$ with system size N . Together with a scalar-relativistic treatment, the basis sets thus provide access to all elements from light to heavy. Both low-communication parallelization of all real-space grid based algorithms and a ScaLapack-based, customized handling of the linear algebra for all matrix operations are possible, guaranteeing efficient scaling (CPU time and memory) up to massively parallel computer systems with thousands of CPUs.

PACS: 31.15.-p, 71.15-m, 71.15.Ap, 71.15.Mb

Key words: *Ab initio* molecular simulations; density-functional theory; atom-centered basis functions; Hartree-Fock; MP2; $O(N)$ DFT; *GW* self-energy

1 Introduction

The rapid increase of computer power in the past decades, along with a furious development and refinement of new physical methods and computational algorithms now makes it possible to study systems of thousands of atoms based on their quantum-mechanical electronic structure with parameter-free, essentially “converged” numerical accuracy. Landmark methods in widespread use today include, for example, density-functional theory (DFT) [1,2], quantum-chemical approaches (Hartree-Fock and correlation approaches based on it), or self-energy methods such as *GW* many-body perturbation theory[3–6].

The key to numerical efficiency and accuracy in electronic structure theory is the form of the basis set that defines the Hilbert space of the electrons, $\{|\varphi_i\rangle, i = 1, \dots, N_b\}$. In many respects, the “best” choice of a basis set is a tradeoff between the available numerical algorithms to deal with individual basis functions, and the required basis set size to guarantee accurate convergence of the overall calculation. One particularly advantageous choice is numeric atom-centered orbital (NAO) basis functions of the form:

$$\varphi_i(\mathbf{r}) = \frac{u_i(r)}{r} Y_{lm}(\Omega) \quad . \quad (1)$$

As the name implies, the radial shape $u_i(r)$ is numerically tabulated and therefore fully flexible. This allows the creation of optimized element-dependent basis sets that are as compact as possible while retaining a high and transferable accuracy in production calculations up to meV-level total energy convergence. To obtain real-valued $\varphi_i(\mathbf{r})$, $Y_{lm}(\Omega)$ here denotes the real parts ($m=0, \dots, l$) and imaginary parts ($m=-l, \dots, -1$) of complex spherical harmonics, with l an implicit function of the radial function index i . The utility of NAOs has already been demonstrated in a number of earlier implementations[7–17] of DFT in the local-density approximation (LDA) [2] or generalized gradient approximation (GGA), as well as in a NAO-based Hartree-Fock implementation[18–23]. Specifically, two features set NAOs apart for all-electron DFT: (i) by including the radial functions of occupied free-atom orbitals in the basis, the all-electron orbital shape and nodes near the deep nuclear Z/r potential are close to exact in bonded structures as well; (ii) each radial function $u_i(r)$ can be strictly localized inside a given radius [$u_i(r) = 0$ outside]. Different spatial regions of large systems are thus strictly separated from one another, enabling near- $O(N)$ scaling with system size of the otherwise most expensive grid-based computational steps [15,16,24–26].

In the present paper, we describe a consistent set of numerical algorithms to implement computational quantum mechanics (electronic structure, total en-

¹ aims-coordinators@fhi-berlin.mpg.de

ergy, energy derivatives with respect to nuclear positions, etc.) for molecules, clusters, and periodic systems based on NAOs. We have implemented all algorithms described here in our own code, the “Fritz Haber Institute *ab initio* molecular simulations” (FHI-aims) package [17], but the observations made should be relevant for any NAO-based first-principles development. We use our implementation to demonstrate how the algorithms below fulfill our general goal of an accurate all-electron, full-potential framework for computational quantum mechanics, with (i) efficiency similar to the fastest existing plane-wave pseudopotential schemes; (ii) good scaling with system size N , up to thousands of atoms; (iii) good scaling on modern hardware from individual PCs up to massively parallel computers with thousands of CPUs. The focus here is on the steps required for DFT-LDA and GGA, including energy gradients and some brief comments on structure relaxation and *ab initio* molecular dynamics on the Born-Oppenheimer surface. A scalar relativistic treatment enables calculations for light and heavy elements with high accuracy and efficiency.

From a physical point of view, there remain today some well-known limitations of standard DFT-LDA or GGA exchange-correlation (XC) functionals for specific applications. These include, e.g.: (i) the incorrect cancellation of the Hartree-term self-interaction of individual electrons in explicit functionals of the electron density; (ii) the treatment of dispersion interactions (van der Waals); (iii) the description of electronic excitations, even of simple single quasiparticle energies. For these cases, an explicit wave-function based treatment of exchange and /or the correlated two-electron motion is necessary to verify the accuracy of simplified approaches. Although not the focus of the present paper, we note that the algorithms presented here can be used directly as a base to implement correlated methods beyond DFT-LDA or GGA through the full two-electron Coulomb interaction using NAO basis functions. Thus, it is possible to access directly methods such as: bare or screened single-determinant exchange; hybrid XC functionals [27–33]; quantum-chemical perturbation theory for the Coulomb interaction, e.g., second-order Møller-Plesset (MP2) perturbation theory [34]; bare (MP2) or screened (GW) self-energies for single-electron excitation energies; or the random-phase approximation (RPA) in the adiabatic connection fluctuation dissipation theorem [35–37]. The essential step to make the calculation of the two-electron Coulomb operator feasible is to expand all products of basis functions into a single auxiliary basis set[38]. A significant step in this regard is variational Coulomb fitting, which has a long history in quantum chemistry [39] and solid state physics [40], and as “resolution of the identity” in later developments [41–43]. A detailed account of our implementation will be presented elsewhere [6]; in particular, we note that larger NAO basis sets than for ground-state DFT may be required for MP2 and GW .

Accurate, numerically tabulated radial functions from, e.g., self-consistent

free-atom calculations have long been available at various levels of theory [44–47]. Since their first sustained applications as multicenter basis functions [7–9], many important practical steps have been contributed by a community of theoreticians. Since we aim here for a coherent and (mostly) self-contained description, some of these earlier results – otherwise scattered over many different sources – are restated in one consistent framework with proper references. In the next section (Sect. 2), we repeat briefly the basic physical equations and scope, allowing to place all the subsequent material in this context. In Sect. 3, we address the question of how to choose reliable basis functions, outlining a procedure that can be used to generate appropriate preconstructed hierarchical basis sets for all elements, ranging from fast qualitative up to the highest accuracy. The detailed numerical algorithms used to implement computational quantum-mechanics with these basis functions are addressed in Sect. 4. In Sect. 5, we demonstrate the practical performance and scaling of these algorithms for characteristic cluster-type and periodic systems, as given by our own implementation, **FHI-aims**.

2 Physical scope

We aim to compute the total energy and other observables (forces, stresses, electrostatic moments, polarizabilities, ...) for any given system of atoms in a cluster-type or periodic geometry within the Born-Oppenheimer approximation. The electronic ground state energy for a fixed nuclear configuration is treated quantum-mechanically, whereas the nuclei themselves are treated classically. In density-functional theory, this requires solution of the electronic Kohn-Sham single-particle equations [2]

$$\hat{h}^{\text{KS}}|\psi_l\rangle = \epsilon_l|\psi_l\rangle \quad (2)$$

for the effective single-particle orbitals $|\psi_l\rangle$ (called Kohn-Sham orbitals below) and eigenvalues ϵ_l , leading to the electron density $n(\mathbf{r}) = \sum_l f_l |\psi_l(\mathbf{r})|^2$. f_l denote occupation numbers, i.e., $f_l=1$ or 0 for occupied or empty orbitals, or fractional occupation numbers in the context of DFT [48]. For clarity, we avoid an explicitly spin-polarized notation for most of this work, but including spin effects through collinear (scalar) spin-DFT is of course straightforward if required. In addition, we plan to extend our own implementation to noncollinear spin in the near future.

The electronic Hamiltonian \hat{h}^{KS} includes the Kohn-Sham effective single-particle kinetic energy \hat{t}_s (possibly with relativistic corrections), the external potential \hat{v}_{ext} (nuclei and a possible embedding into an external field), the electrostatic potential (Hartree potential) of the electron density, \hat{v}_{es} , and the exchange-

correlation potential \hat{v}_{xc} :

$$\hat{h}^{\text{KS}} = \hat{t}_s + \hat{v}_{\text{ext}} + \hat{v}_{\text{es}} + \hat{v}_{\text{xc}} \quad . \quad (3)$$

As is well-known, Eqs. (2) and (3) must be solved *self-consistently*.

In each iteration towards self-consistency, we use the basis functions $\varphi_i(\mathbf{r})$ ($i = 1, \dots, N_{\text{b}}$) and the expansion $\psi_l(\mathbf{r}) = \sum_{i=1}^{N_{\text{b}}} c_{il} \varphi_i(\mathbf{r})$ to discretize Eq. (2) into a generalized eigenvalue problem [49,50]

$$\sum_j h_{ij} c_{jl} = \epsilon_l \sum_j s_{ij} c_{jl} \quad . \quad (4)$$

Since $\varphi_i(\mathbf{r})$ are tabulated numerically, we obtain the Hamilton and overlap matrix elements h_{ij} and s_{ij} by numerical integration (Sect. 4.1)

$$\begin{aligned} h_{ij} &= \int d^3r [\varphi_i(\mathbf{r}) \hat{h}^{\text{KS}} \varphi_j(\mathbf{r})] \quad , \\ s_{ij} &= \int d^3r [\varphi_i(\mathbf{r}) \varphi_j(\mathbf{r})] \quad . \end{aligned} \quad (5)$$

We omit the complex conjugate notation here and elsewhere in the paper, since we use the real-valued real and imaginary parts of complex spherical harmonics in our basis functions φ_i as given in Eq. (1).

In DFT, the Kohn-Sham functional defines the total energy (electrons and nuclei) as

$$E_{\text{tot}} = T_s[n] + V_{\text{ext}}[n] + E_{\text{es}}[n] + E_{\text{xc}}[n] + E_{\text{nuc-nuc}} \quad , \quad (6)$$

i.e., the Kohn-Sham kinetic energy T_s , the (external) potential energy V_{ext} , the classical electrostatic energy of the electron density E_{es} , the exchange-correlation functional E_{xc} , and the internuclear repulsion $E_{\text{nuc-nuc}}$. In practice, Eq. (4) gives us direct access to the single-particle eigenvalues ϵ_l , which are the expectation values of the Hamiltonian Eq. (3). It is therefore conventional to evaluate the total energy as:

$$\begin{aligned} E_{\text{tot}} &= \sum_{l=1}^{N_{\text{states}}} f_l \epsilon_l - \int d^3r [n(\mathbf{r}) v_{\text{xc}}(\mathbf{r})] + E_{\text{xc}}[n] \\ &\quad - \frac{1}{2} \int d^3r [n(\mathbf{r}) v_{\text{es}}(\mathbf{r})] + E_{\text{nuc-nuc}} \quad . \end{aligned} \quad (7)$$

Starting from the sum of single-particle eigenvalues (in principle, up to the computed number of Kohn-Sham states $N_{\text{states}} \leq N_{\text{basis}}$), the exchange-correlation energy functional $E_{\text{xc}}[n]$ (correct for the total energy) replaces the exchange-correlation *potential* energy (which is part of the sum of eigenvalues), and the double-counting of the electrostatic electron-electron interaction is removed.

The same basic expressions Eqs. (6) and (7) can be written to cover Hartree-Fock theory and hybrid functionals, which formally correspond to DFT as “generalized Kohn-Sham schemes” [51]. Once solved, the single-particle orbitals $|\psi_l\rangle$ may be used as a starting point for many-body perturbation theory or MP2, detailed in Ref. [6]. In total, there exist today a wide range of approaches to the exchange-correlation energy, E_{xc} , for example those supported in our own implementation [17]: the most widely used local-density [52–55] and generalized gradient approximations [56–61], as well as hybrid exchange-correlation functionals [28,30–33], Hartree-Fock and MP2 perturbation theory. Long-range van der Waals interactions, missing in DFT-LDA/GGA, may be incorporated on a total-energy level, based on damped interatomic C_6/R^6 contributions parameterized using high-level quantum-chemical benchmark data (e.g., Ref. [62] and references therein). Finally, the localized nature of NAOs allows imposition of explicit constraints on the effective occupation numbers of certain spatial regions and/or spin channels with electrons [63], e.g. allowing approximate enforcement of a certain spin state of a molecule in a given environment.

In addition to the total energy itself, its derivatives can be analytically calculated from splined expressions for the basis functions (see Sect. 3 below). Again, its uses are manifold. In our own work, this includes efficient structure optimizations, nudged-elastic band transition state searches [64–66], Born-Oppenheimer molecular dynamics, and vibrational properties, both harmonic (by finite differences) and anharmonic (using the dipole-dipole correlation function in a sufficiently long molecular dynamics run, see Ref. [67] for a thorough discussion). In particular, we note that, for $T = 0$ K total energy stability analyses, the zero-point motion of the nuclei may constitute an important additional term, which is readily obtained from a calculation of vibrational frequencies, if required.

3 Numeric atom-centered basis functions

3.1 Numerical definition of NAO basis functions

The general shape of basis functions in Eq. (1) includes both analytically and numerically defined functions, i.e., also the analytically defined Gaussian-type or Slater-type orbitals of traditional quantum chemistry. However, the advantage of the flexibility of $u_i(r)$ is due to the fact that numerical solutions of (possibly scalar-relativistic) Schrödinger-like radial equations can be chosen:

$$\left[-\frac{1}{2} \frac{d^2}{dr^2} + \frac{l(l+1)}{r^2} + v_i(r) + v_{\text{cut}}(r) \right] u_i(r) = \epsilon_i u_i(r). \quad (8)$$

Each radial function is thus defined by two parts: a potential $v_i(r)$ which defines the main behaviour of $u_i(r)$, and a steeply increasing *confining* potential $v_{\text{cut}}(r)$, which ensures a smooth decay of each radial function to be strictly zero outside a confining radius r_{cut} (see Sect. 3.2).

In particular, Eq. (8) allows one to include a *minimal basis* that consists of the core and valence functions of spherically symmetric free atoms, by simply setting $v_i(r)$ to the self-consistent free-atom radial potential $v_{\text{at}}^{\text{free}}$. In practice, we use non-spinpolarized free atoms for this task. We have found that the resulting basis sets are transferable also to spin-polarized structures. This minimal basis greatly facilitates the all-electron treatment, because it naturally accounts for wave function oscillations near the nucleus (where the nuclear Z/r potential dominates). Any basis set errors originating from this region are therefore strongly reduced at the outset.

In practice, we solve the radial equations Eq. (8) in one dimension on a dense logarithmic radial grid $r(i) = r_0 \exp[(i - 1)\alpha]$ ($i = 1, \dots, N_{\text{log}}$), using the solver described in Ref. [68]. Smooth radial functions are ensured through cubic spline interpolation [10]. Since radial functions originating from different defining potentials $v_i(r)$ are not necessarily orthonormal to one another even on the same atomic site, we orthonormalize all on-site radial functions explicitly on the logarithmic grid, using a simple Gram-Schmidt process [69]. Note that, time-wise, this step is completely insignificant compared to a (later) self-consistent calculation: the orthonormalization need only be done once for each set of radial functions within the same angular momentum channel of every chemical element in the structure. All radial functions within each angular momentum channel can first be grouped in order of increasing outermost radius (beyond which their tails can be numerically neglected). The resulting orthonormalized radial functions then each retain the minimally possible radial extent, rather than admixing longer-ranged radial functions to (originally) shorter-ranged ones.

3.2 Basis localization: The confining potential

The addition of a confining potential $v_{\text{cut}}(r)$ in Eq. (8) to all our basis functions prevents any extended, slow-decaying analytical or numerical radial function tails. The resulting strict separation of basis functions in distant spatial regions is critical to ensure computational efficiency for large structures, including the $O(N)$ -like scaling of numerical integrations (Sect. 4.1) and the electron density update (Sect. 4.3). Possible functional forms of $v_{\text{cut}}(r)$ have been discussed by several authors, e.g., Refs. [12–14,70–74]. Most importantly, care must be taken to choose a shape $v_{\text{cut}}(r)$ that ensures a smooth decay of all basis functions and their derivatives to zero. For example, a simple “hard-wall” potential

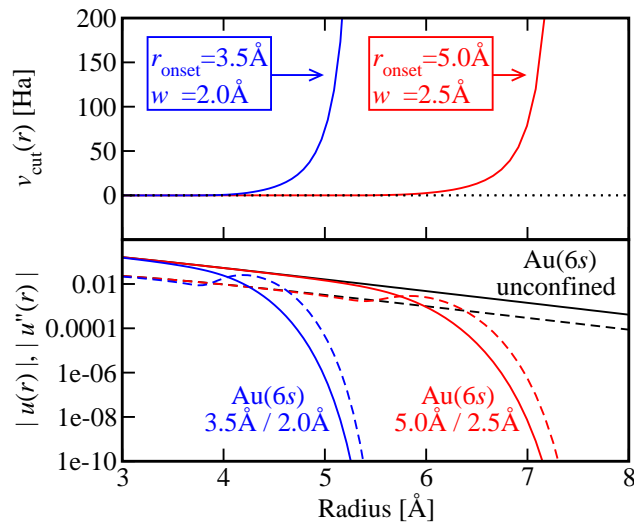


Fig. 1. Upper panel: Basis set confining potentials $v_{\text{cut}}(r)$ as defined in Eq. (9), for two different settings of r_{onset} , w ($s=200$ Hartree in both cases). Lower panel: Influence of $v_{\text{cut}}(r)$ on the tails of the Au(6s) radial function $u(r)$ (solid lines) and its second derivative $u''(r)$ (dashed lines) for both $v_{\text{cut}}(r)$ settings, compared to the unconfined case.

would produce a discontinuous second derivative, and thus a possible discontinuity in the total energy landscape. In our own work, we employ by default the following smooth analytical shape:

$$v_{\text{cut}}(r) = \begin{cases} 0 & r \leq r_{\text{onset}} , \\ s \cdot \exp\left(\frac{w}{r-r_{\text{onset}}}\right) \cdot \frac{1}{(r-r_{\text{cut}})^2} & r_{\text{onset}} < r < r_{\text{cut}} , \\ \infty & r \geq r_{\text{cut}} . \end{cases} \quad (9)$$

At r_{onset} , $v_{\text{cut}}(r)$ begins with an infinitely smooth onset, and then increases gradually to infinity at r_{cut} over a width $w = (r_{\text{cut}} - r_{\text{onset}})$. s is a global scaling parameter ($s=200$ Ha and $w=2.0$ Å are safe default values). Eq. (9) is a modification of the proposal of Ref. [73], using $\frac{1}{(r-r_{\text{cut}})^2}$ instead of $\frac{1}{(r-r_{\text{cut}})}$ to spread the damping of the radial function more evenly across the width of the confining potential.

The confinement potential v_{cut} and its impact on the tails of an atomic Au 6s radial function and its second derivative are shown in Fig. 1 for two different settings. On a logarithmic scale, the numerical importance of cutting off the basis function tails is obvious even for a wide confining potential. It is also apparent that the increased curvature in the cutoff region leads to a change in the second derivative of $u_i(r)$ (affecting the kinetic energy) that becomes sharper with decreasing cutoff width w . For our standard width ($w=2.0$ Å), this is not problematic, but for significantly smaller widths, denser integration grids might be required to maintain the same accuracy.

Some authors [12,14,70,73,74] have advocated the use of $v_{\text{cut}}(r)$ as an explicit

basis-shaping parameter, to gain some extra system-specific basis flexibility at fixed basis size (e.g., when only using the minimal basis). In contrast, we presently prefer to employ $v_{\text{cut}}(r)$ as a purely *technical* quantity that should be chosen conservatively, so as to not impact any physical results. We stress that r_{onset} can and should be employed as an explicit convergence parameter for a given basis set, verifying, e.g., the numerical development of a given energy difference with increasing r_{onset} . In our experience, $r_{\text{onset}}=3.5 \text{ \AA}$ typically yields well-converged energy differences for light-element structures (for example, meV-level convergence of the benchmark 203-atom molecular energy difference discussed in Sec. 5 below), and only slightly larger values ($r_{\text{onset}}=4.0 \text{ \AA}$) suffice for transition metals.

For numerical efficiency, it is additionally important to note that different radial functions $u_i(r)$ will decay towards large r with different speed. For example, radial functions generated for high angular momenta (f, g) may in practice be peaked more sharply and closer to the nucleus than lower angular momenta (s, p). It is thus advantageous to introduce a cutoff criterion that accounts for the actual extent of each individual radial function $u_i(r)$. Starting with a version of $u_i(r)$ calculated *without* a tight confinement, we define a new onset radius $r_{\text{onset},i}$ for $v_{\text{cut}}(r)$ through

$$\int_{r_{\text{onset},i}}^{\infty} dr |u_i(r)|^2 = \eta_{\text{cut}}. \quad (10)$$

Here, η_{cut} is a small threshold, typically $\eta_{\text{cut}} \approx 10^{-4}$. After $r_{\text{onset},i}$ is thus determined, $u_i(r)$ is recalculated *with* that new confinement onset, and the tail of that radial function is now pushed to exactly zero over the width of $v_{\text{cut}}(r)$. In addition, a maximum value for r_{onset} that *no* radial function may exceed should still be set.

3.3 Iterative construction of accurate, transferable basis sets

In order to enable both fast production calculations and basis set convergence studies down to meV-level total energy accuracy, the availability of reliable, preconstructed basis sets is a necessity. Our strategy to construct basis sets, described below, is guided by the following goals. The basis sets should be accurate both on a numerical level and converge systematically towards the converged basis set limit as the basis size increases. For a given accuracy level, the basis sets should be as small as possible, while remaining transferable with the same accuracy between different atomic structures. Any approximation to the converged basis set limit is thus well documented and controlled. We have used the strategy defined below to obtain preconstructed basis sets from hydrogen to fermium (elements 1-100); further details will be part of a separate publication [75].

In order to reach the stated goals, we construct basis sets guided by the following principles:

- (1) Basis sets should be available starting from the minimal basis on up to meV-level total energy convergence.
- (2) Successive basis sets for the same element should be *hierarchical*, i.e., a larger basis set should contain all smaller ones. This ensures a strictly variational total energy convergence.
- (3) The basis construction procedure should be defined as objectively as possible, i.e., with as little human bias as possible. In particular, any necessary human intuition should be subject to the same rigorous verification as any automated parts of the process.

In practice, we ensure the validity of principles (1-3) by assuming the following iterative basis set construction strategy:

- We define a large pool of “candidate” radial function shapes $\{u_i(r)\}$ (e.g., hydrogen-like, cation-like, or atom-like functions with a variable confinement potential; for details, see below). Initially, this pool is kept as open as possible to avoid any *a priori* limitations.
- Starting from a given basis set (initially, the *minimal* free-atom basis), we run through the entire pool of candidate functions, adding each function to the given basis set in turn. The radial function that gives the single largest improvement of a target total energy is added to the original basis set.
- The procedure is repeated until no further significant total energy improvements result.

Formally, this construction scheme could be applied to generate separate basis sets for different XC functionals, and indeed, we vary the minimal free-atom basis according to the underlying LDA / GGA functional. In practice, we have found that the additional basis functions as generated, e.g., in LDA, show interchangeable performance in different GGAs (and also hybrid functionals and Hartree-Fock), and we thus employ the same sets of additional basis functions for different functionals alike. However, we also note that more or special care to ensure basis set convergence is required for methods that rely explicitly on unoccupied orbitals, such as MP2, RPA, or *GW* (a separate discussion will be given in Ref. [6]).

Two important choices remain to specify our basis set construction strategy: the pool of candidate radial functions, and a target for the total energy optimization that is both fast to compute and generic enough to ensure transferable basis sets.

Many suggestions for additional NAO basis functions beyond the minimal basis exist in the literature. Historically the first were simple confined free-atom-

like excited state radial functions [7]. Other suggestions include the radial functions of free ions or of Hydrogen-like atoms for a variable nuclear potential z/r [8–10], or (for pseudopotentials) simple radial polynomials in a given region [15,76]. Exact Hydrogen-like orbitals are analytically known and (by linear combination) similar to Slater-type orbitals [77], but in their numerical form, an explicit localizing confining potential can be applied. In addition, the numerical form allows us to explicitly orthonormalize each additional basis function against all previous ones at the same atom, i.e., in practice, only orthonormal on-site basis sets arise.

For Gaussian-type orbitals and correlated methods, the path to the converged basis set (CBS) limit has been systematically investigated by various groups (e.g., Dunning and coworkers in Ref. [78] and references therein). In contrast, studies that systematically explore the all-electron CBS limit using NAOs are comparatively scarce. For two specific systems (H_2 and N_2), Delley [10] has demonstrated how the CBS limit can be reached, noting that ion-like radial functions always serve very well as the first (“double-numeric”) radial functions to augment the atomic valence shells beyond the minimal basis. Hydrogen-like functions are used in addition, especially as *polarisation functions* for angular momenta beyond those present in the free atom itself.

Regarding the basis function “pool”, we have tested valence and excited-state radial function shapes of single atoms with different confining potentials, (confined) radial functions of cations, and hydrogen-like radial functions. Contrary to the intuitive expectation that confined atomic excited states might provide a systematic “ladder” of basis functions that is also efficient, we find that such functions are almost always outperformed by ion-like or hydrogen-like functions. In agreement with Ref. [10], the valence functions of cations always perform very well as initial additions to the minimal basis; however, even then, hydrogen-like functions that perform almost equally well or slightly better can usually be found.

Our eventual radial function pool for production basis sets is thus restricted to (i) the radial functions of doubly positive charged free ions, and (ii) hydrogen-like radial functions for a potential z/r with z in the range $0.1 \leq z \leq 20$. Note that, for the purpose of optimizing a radial function shape, it is not necessary for z to remain integer (in fact, integer z would cost us some variational flexibility), and we treat z as a continuous parameter instead. For light elements (H-Ar), we consider angular momenta $l=0,\dots,4$; for K (atomic number $Z=19$) and beyond, h functions ($l=5$) are also included.

Regarding the optimization target, we select the simplest possible chemical bonds formed by a given element: We minimize the total energy error of a set of N_d non-spinpolarized symmetric dimers at $N_d \sim 4$ -5 different bond distances

d_i , in DFT-LDA, given by

$$\Delta_{\text{basis}} = \frac{1}{N_d} \sum_{i=1}^{N_d} [\mathcal{E}_{\text{basis}}(d_i) - \mathcal{E}_{\text{cb}}(d_i)] \quad . \quad (11)$$

Here, $\mathcal{E}_{\text{basis}}(d_i)$ denotes the *non-selfconsistent* total energy per atom, for the dimer at bond distance d_i for a given basis set. The non-selfconsistent reference energy $\mathcal{E}_{\text{cb}}(d_i)$ for a converged basis set (cb) is obtained independent of the present procedure, by converging a very large, formally systematic basis set of confined atomic excited-state functions. We find that *non-selfconsistent* total energies are a fully sufficient optimization target, avoiding potential instabilities of the self-consistency cycle for pathological cases. The dimer distances d_i ($i=1, \dots, n_d$) are spread to sample the (self-consistent) LDA binding curve. We use the binding curve minimum itself, a shorter distance in the repulsive regime at $E \approx E_{\text{min}} + 2$ eV, and larger distances characteristic of chemical bonds and the large-distance tail of the binding curve.

During the optimization, the onset of the confining potential was chosen such that the reference total energies for the dimers from the independent, large systematic basis set above were fully (over-)converged (in practice, $r_{\text{onset}} = 5$ Å or larger was used during the basis optimization). This criterion for the cutoff is both rigorous and sufficient, since a basis optimized for a too narrow confining potential would be prevented from reaching the converged basis limit by construction. We note that the radial function dependent cutoff criterion Eq. (10) was not used for the basis set construction [although technically possible, any performance gains from Eq. (10) are not critical during the basis construction procedure].

Since the minimal basis functions are set up in DFT, we here benefit from the fact that total energies of (spherical) free atoms are already converged for the initial minimal basis. Our basis optimization procedure thus focuses exclusively on the wave function changes necessary to describe bond formation, rather than slow-converging single-atom contributions near the nucleus. As pointed out by others [74], dimers are actually a rather demanding test case for atom-centered orbitals, since no basis function overlap is available from other centers to improve the total energy accidentally.

Figure 2 illustrates the iterative basis construction process from qualitative tight-binding like accuracy to sub-meV total energy convergence for four elements: H, C, O, and Au. In each case, we show the convergence of the average non-selfconsistent total energy error of the sets of N_d symmetric dimers, Δ_{basis} [Eq. (11)], as the basis size increases. The initial full symbol indicates the minimal basis of occupied atomic radial functions. Each open symbol corresponds to one more selected radial function [with $(2l + 1)$ angular momentum functions]. According to the general prescription stated above, the LDA binding

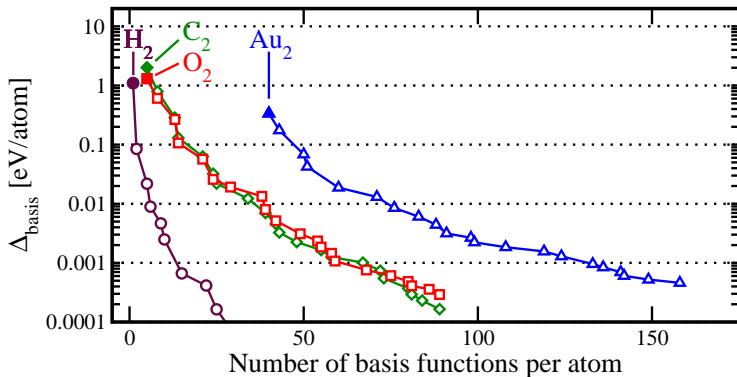


Fig. 2. Iterative construction of NAO basis sets: Development of Δ_{basis} [see Eq. 11] as a function of basis size for H_2 , C_2 , O_2 and Au_2 , averaged over several bond distances in each case (see text for details). Full symbols denote the minimal basis (occupied atomic core and valence functions only). Open symbols show the convergence of Δ_{basis} as one radial function after another is selected and added to the previous basis set.

curves for H_2 , C_2 , N_2 , and Au_2 lead to $d_i/\text{\AA}=\{0.5, 0.7, 1.0, 1.5, 2.5\}$ for H, $d_i/\text{\AA}=\{1.0, 1.25, 1.5, 2.0, 3.0\}$ for C, $d_i/\text{\AA}=\{1.0, 1.208, 1.5, 2.0, 3.0\}$ for O, and $d_i/\text{\AA}=\{2.1, 2.45, 3.0, 4.0\}$ for Au.

The radial functions thus selected in Fig. 2 are listed in Table 1. The lists are organized as different *tiers* or levels of basis function groups of different angular momenta: e.g., *spd-spdfg-spdf*... for O. While this type of grouping is often assumed intuitively (e.g., in standard Gaussian basis sets as described in Refs. [79,78,80] and references therein), *it here arises naturally as the output sequence of basis functions* from our automated basis construction process, not as input. We find this appearance of naturally ordered *tiers* across all elements, with only minor occasional exceptions. One such exception is indicated by a * symbol in the column for Au. This hydrogen-like $4f$ function appears between the p and s functions of the second *tier* in the sequence of Fig. 2; in order to achieve a more efficient second *tier* that is yet consistent with other elements, it is here listed as part of the third *tier* instead. It is important to note that this is merely a change of sequence, and that the entire list of available basis is of course always available for rigorous convergence tests.

We note that the precise point where one *tier* ends and another begins is sometimes not unique. This is the case, e.g., for the first *tier* of C and O, which might or might not include the following f function. For second-row elements as well as Na and Mg ($Z=11, 12$), the angular momenta of the first three selected basis functions are always $\{s, p, d\}$. We therefore label these basis functions as the first *tier* for elements 3-12 (Li-Mg) for consistency. This basis level has sometimes been labelled as “double-numeric plus polarization” (dnp) [10,81] or “double-zeta plus polarization” (dzp) [73] by others. Again, we note that, in our case, this sequence is not a designation based on mere

	H	C	O	Au
minimal	1s	[He]+2s2p	[He]+2s2p	[Xe]+6s5d4f
<i>tier 1</i>	H(2s,2.1)	H(2p,1.7)	H(2p,1.8)	Au ²⁺ (6p)
	H(2p,3.5)	H(3d,6.0)	H(3d,7.6)	H(4f,7.4)
		H(2s,4.9)	H(3s,6.4)	Au ²⁺ (6s)
				H(5g,10)
				H(6h,12.8)
				H(3d,2.5)
<i>tier 2</i>	H(1s,0.85)	H(4f,9.8)	H(4f,11.6)	H(5f,14.8)
	H(2p,3.7)	H(3p,5.2)	H(3p,6.2)	H(4d,3.9)
	H(2s,1.2)	H(3s,4.3)	H(3d,5.6)	H(3p,3.3)
	H(3d,7.0)	H(5g,14.4)	H(5g,17.6)	H(1s,0.45)
		H(3d,6.2)	H(1s,0.75)	H(5g,16.4)
				H(6h,13.6)
<i>tier 3</i>	H(4f,11.2)	H(2p,5.6)	O ²⁺ (2p)	H(4f,5.2)*
	H(3p,4.8)	H(2s,1.4)	H(4f,10.8)	H(4d,5.0)
	H(4d,9.0)	H(3d,4.9)	H(4d,4.7)	H(5g,8.0)
	H(3s,3.2)	H(4f,11.2)	H(2s,6.8)	H(5p,8.2)
				H(6d,12.4)
				H(6s,14.8)
	

Table 1

Radial functions selected during the basis optimization for H, O, and Au, as illustrated in Fig. 2. The first line (“minimal”) summarizes the free-atom radial functions used (noble-gas configuration of the core and quantum numbers of the additional valence radial functions). “H(nl, z)” denotes a hydrogen-like basis function for the bare Coulomb potential z/r , including its radial and angular momentum quantum numbers, n and l . $X^{2+}(nl)$ denotes a n, l radial function of a doubly positive free ion of species X . The asterisk denotes one radial function that is listed out of sequence to retain the otherwise consistent ordering into successive angular momentum shells (*tiers*; see text).

intuition, but rather a *result* of our basis optimization strategy. The same logic also indicates that, for heavier elements, *tier 1* should comprise more than just an s , p , and d function: For Al ($Z=13$) and beyond, the first f function occurs indeed before the $\{s,p,d\}$ group is complete, and is thus included in *tier 1*.

Finally, we note that in Fig. 2, the absolute convergence level achieved with a given *tier* is element dependent. For example, the convergence level of the *tier* 1 basis sets for C and O is ≈ 0.1 eV for the non-selfconsistent *total* energy. Compared to the minimal basis, the error is thus reduced by more than a factor of ten. For energy differences between bonded structures, much of the remaining error will also cancel. In practice, *tier* 1 for light elements thus guarantees accurate geometry (pre-)relaxations, but *tier* 2 (with a convergence accuracy of ≈ 0.01 eV for the non-selfconsistent total energy in Fig. 2) is required to guarantee meV-converged energy differences. Comparing this to Au, we see that already *tier* 1 yields ≈ 0.01 eV convergence, and is sufficient in practice for well-converged energy differences. Our findings regarding energy differences are illustrated quantitatively in Sect. 3.5 below.

3.4 Transferability of NAO basis sets: Basis set superposition errors

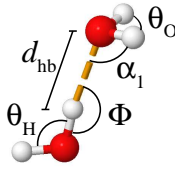
A particular concern for overlapping atom-centered basis sets is the possibility of *basis set superposition errors* (BSSE) [82–84]. Consider, for example, the atomization energy of an N -atom compound:

$$\Delta E^{\text{atm}} = E_{\text{tot}}^{\text{compound}} - \sum_{i=1}^N E_{\text{tot}}^{\text{atom } i}. \quad (12)$$

Most LCAO basis set types (Gaussians, Slater-type orbitals, ...) do not describe the free atom exactly. Hence, basis functions situated on one atom may variationally improve the total energy contributions originating from other atom in the compound. Since this improvement is possible only in the compound, not in each atomic reference calculation (without the additional basis functions from the other atoms), the result is an artificially lowered atomization energy ΔE^{atm} . The crux is that apart from depending on the basis set, this BSSE is also structure dependent, i.e., denser structures will typically suffer a larger BSSE than sparser ones.

By construction, our NAO basis sets does not incur any BSSE at all in Eq. (12) for non-spinpolarized, spherical reference atoms: The exact valence wave functions of these atoms are already part of the basis (the minimal basis), and no variational improvement is possible by adding further basis functions. On the other hand, this argument no longer strictly holds for spin-polarized and/or non-spherical [85] reference atoms, and perhaps even more importantly for *intermolecular* binding energies. The wave function for each individual molecular fragment is certainly not exact, and the question of an intermolecular BSSE must still be addressed.

We exemplify the role of BSSE for NAOs for a prototype intermolecular geometry and binding energy, that of the hydrogen-bonded H₂O dimer (for a



Basis set	d_{hb} [Å]	Φ [°]	θ_{H} [°]	θ_{O} [°]	α_1 [°]	α_2 [°]
min	1.74	163.8	100.6	101.4	95.2	94.3
<i>tier 1</i>	1.91	172.3	104.5	104.6	109.0	108.5
<i>tier 2</i>	1.92	172.3	104.6	104.8	108.4	108.0
<i>tier 3</i>	1.92	172.3	104.6	104.8	108.4	108.0

Fig. 3. Selected equilibrium geometry parameters of the H_2O dimer in DFT-PBE, as a function of the basis set. The table on the right lists the optimized values of bond distances and angles as defined by the figure on the left. One of the tilt angles of the upper H_2O molecule, α_2 (the equivalent of α_1 for the other H atom in this molecule), is hidden and therefore not indicated in the figure.

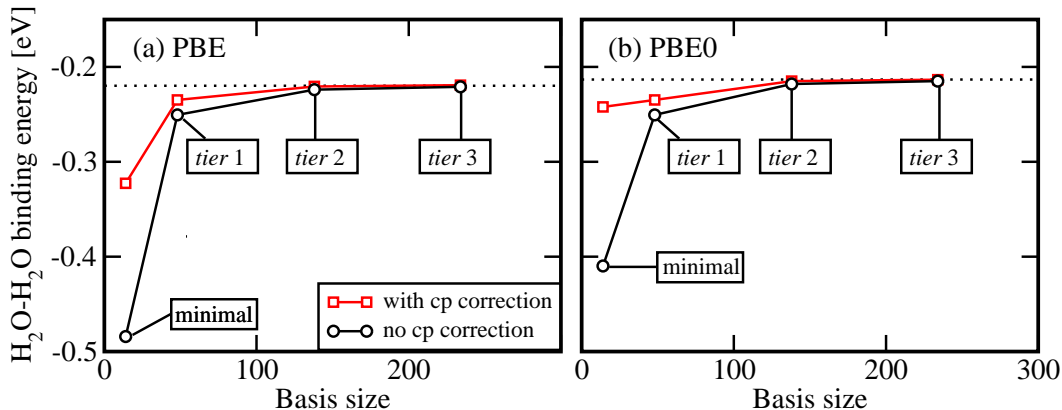


Fig. 4. Uncorrected (circles) and counterpoise-corrected (squares) binding energy of the H_2O dimer as a function of basis set as defined in Table 1. (a) DFT-PBE. (b) DFT-PBE0.

description and further benchmarks of this system, see Ref. [86] and references therein). The convergence of the geometry parameters using the Perdew-Burke-Ernzerhof (PBE) [58] exchange-correlation functional is summarized first in Fig. 3. The only significant geometry deviations from the converged value are found for the minimal basis; already for the small *tier 1* basis, even sensitive tilt angles between both molecules are essentially correct.

In Fig. 4, we show the convergence with basis set size of the $(\text{H}_2\text{O})_2$ binding energy for the DFT-PBE [58] (subfigure a) and -PBE0 [30] (subfigure b) exchange-correlation functionals and for different basis sets. Both graphs employ the same optimized geometries of Fig. 3.² For both functionals, the black

² While the present paper deals mostly with DFT-LDA and -GGA, we here include results for the PBE0 hybrid functional, to illustrate the point that the performance of our *orbital* basis sets for ground-state calculations is qualitatively independent of whether a normal or generalized Kohn-Sham scheme (for the single-determinant

curves (circles) show the overall convergence of the straight energy difference

$$E_b = E_{\text{relaxed}}[(\text{H}_2\text{O})_2] - 2E_{\text{relaxed}}[\text{H}_2\text{O}] \quad (13)$$

for the *tier* basis sets of Table 1, choosing an (over-)converged confining potential ($r_{\text{onset}}=5.0 \text{ \AA}$, $w=2.5 \text{ \AA}$). E_b is qualitatively converged (to $\approx 10 \%$) for both functionals already at the *tier* 1 level, and is quantitatively converged for *tier* 2 (within 5 meV of the converged values, dotted lines: $E_{\text{PBE,conv}}=-219.5 \text{ meV}$ and $E_{\text{PBE0,conv}}=-213.3 \text{ meV}$, verified independently of the *tier* basis sets shown here). Together with some residual basis incompleteness error cancellation between $E_{\text{relaxed}}[(\text{H}_2\text{O})_2]$ and $E_{\text{relaxed}}[\text{H}_2\text{O}]$, the relative magnitudes of the remaining error in E_b with increasing basis set size are directly consistent with the total energy convergence levels of the H and O basis sets indicated in Fig. 2, and illustrate the transferability of these basis sets.

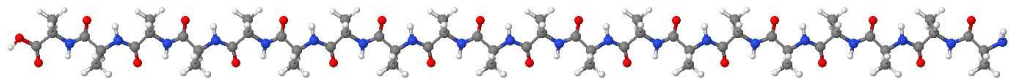
The red lines in Figs. 4a and b address the magnitude of the residual BSSE in our calculations by way of the counterpoise (CP) correction [83,84] of quantum chemistry. The CP correction compensates for the BSSE by calculating the total energies of both H_2O fragments in the dimer separately, once by including only the basis functions of the fragment in question ($E_{\text{fragment}}[\text{H}_2\text{O}]$), and once for the same fragment geometry, but including as “ghost” basis functions on empty sites the basis functions of the respective other fragment ($E_{\text{fragment}}^{\text{CP}}[\text{H}_2\text{O}]$). The resulting correction term is

$$\Delta^{\text{CP}} = \sum_{\text{fragments}} [E_{\text{fragment}}^{\text{CP}}[\text{H}_2\text{O}] - E_{\text{fragment}}[\text{H}_2\text{O}]]. \quad (14)$$

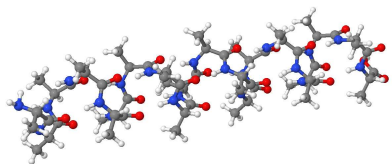
$E_b - \Delta^{\text{CP}}$ is given by the red lines in Fig. 4. It is immediately obvious that Δ^{CP} is of appreciable magnitude only for the minimal and *tier* 1 basis sets, i.e., those where a noticeable binding energy error exists in the first place. However, the correction is not enough to fully compensate for the basis set error, indicating that the BSSE is not the dominant aspect of basis set incompleteness in both cases. While the correction shows the right trend also for *tiers* 2 and 3, it is there almost negligible in magnitude, consistent with the already good total energy convergence for the larger basis sets.

The above observations for $(\text{H}_2\text{O})_2$ in DFT match our experience with many other systems. We conclude that BSSE corrections for ground-state DFT are not critical for the well converged NAO-based total energies and energy difference that can be achieved already at moderate basis size (e.g., *tier* 2). We note that this holds only for the local, semilocal and hybrid functionals. In contrast, a CP correction may be critical for explicitly correlated methods (e.g., MP2), as will be further discussed in Ref. [6].

exchange part of PBE0) is used. Implementation details, specifically regarding the auxiliary basis set used to expand products of orbital basis functions to compute the two-electron Coulomb operator, will be given in a forthcoming paper [6].



(a) Fully extended Ala₂₀



(b) α -helical Ala₂₀

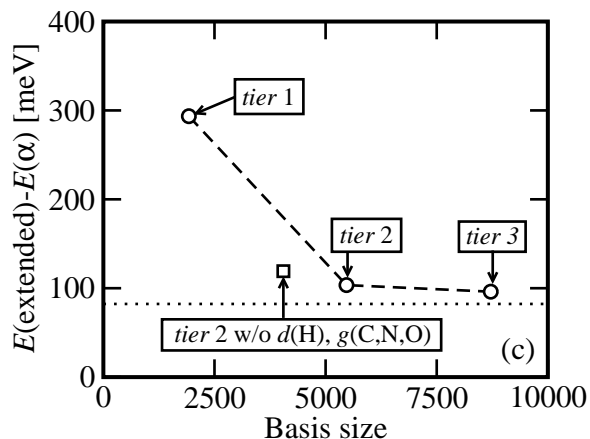


Fig. 5. (a) Fully extended and (b) α -helical conformer of a 203-atom polyaniline molecule Ala₂₀. (c) Convergence of the conformational difference between both conformers with NAO basis size, for successive basis *tiers* (Table 1). The dotted line indicates the result for a very large basis set. Also shown (square symbol) is the result using *tier 2*, but omitting high angular momentum functions (*d* for H, *g* for C, N, and O).

3.5 Converging NAO basis sets in practice

With a consistent and general basis set construction strategy, we still require benchmark examples that demonstrate how basis set convergence is achieved in practice. The present section illustrates the convergence of energy differences and materials properties with basis size, for two characteristic classes of systems: large light-element molecule conformers, and a heavy-element periodic solid (fcc Au). We also address the relevance of high-angular momentum basis functions in our production basis sets. For example, the basis sets in Table 1 contain comparatively high angular momenta (*d* for H, *tier 2*; *g* for C and O, *tier 2*; or, *g* and *h* for Au, *tier 1*). Since including such basis functions means significant computational effort, it is important to verify their impact for specific tasks.

Strictly, the conclusions below hold for the specific cases shown (a larger series of tests is beyond the scope of this paper), but in our experience they are indicative of general trends. They also indicate how routine convergence tests can be conducted for the key results of any particular physical study, analogous to, e.g., verifying the cutoff energy in any application of plane wave basis sets.

We first turn to the convergence of a large-molecule conformational energy difference in Fig. 5. Part (c) shows the difference between a fully extended [subfigure (a)] and an α -helical [subfigure (b)] conformer of a polypeptide Ala₂₀ (203 light-element atoms H, C, N and O) as a function of basis size. Both structures are fixed-geometry cutouts from fully relaxed, infinite-chain models of the same structure elements [87], with relaxed terminating groups -NH₂ and -COOH.

Convergence with basis size: For a small *tier 1* basis set, Figure 5c shows a remaining error of ≈ 210 meV in the energy difference. In fact, this corresponds to “only” ≈ 1 meV/atom, but is clearly farther away from the converged limit than typical relevant energy scales kT . Still, this basis level performs very well for geometry optimizations (consistent with Fig. 3 above). A larger *tier 2* basis set improves the calculated energy difference significantly. For it, the remaining error is ≈ 25 meV, i.e., 0.12 meV/atom from the converged limit. As in the case of (H₂O)₂, the *relative* magnitudes of the errors in the energy difference are consistent with the relative magnitudes of the total energy convergence levels of the respective *tiers* achieved during the basis set construction procedure for dimers, cf. Fig. 2.

Effect of high angular momentum basis functions: The square data point in Fig. 5c shows the convergence level achieved for this example when the highest angular momentum functions from the *tier 2* basis sets of each element are omitted (*d* in the case of H, *g* in the case of C, N, and O). Compared to the full *tier 2*, this modification incurs an additional error of ≈ 20 meV, which is rather moderate in view of the size of the investigated systems, especially in view of the basis size reduction achieved (≈ 25 % in this example). For many production calculations, this basis set level is thus sufficient to guarantee high-accuracy conformational energy differences at a rather competitive computational cost. On the other hand, the *total energy* error incurred by omitting these functions is ≈ 1.4 eV, compared to a remaining total energy error of ≈ 0.5 eV for the full *tier 2*. The omitted high-*l* basis functions thus are not negligible when investigating quantities without error cancellation, such as atomization energies in the case of our NAOs.

In Fig. 6, we complement our findings for light elements with the NAO convergence properties for a heavy transition metal, Au ($Z=79$). Specifically, we illustrate the convergence properties of the lattice parameter a_{fcc} , the bulk modulus B , and the cohesive energy E_{coh} (obtained from a fit of the Murnaghan equation of state $E(V)$ [88]) with basis size for fcc Au in DFT-PBE.

Convergence with basis size: Figure 6 shows that a_{fcc} , B and E_{coh} converge rapidly with basis size. Consistent with the ≈ 0.01 eV convergence level achieved in the basis construction (Fig. 2), the *tier 1* basis set gives essentially converged results. For this basis set, $a_{\text{fcc}}=4.159$ Å and $B=1.37$ MBar agree very

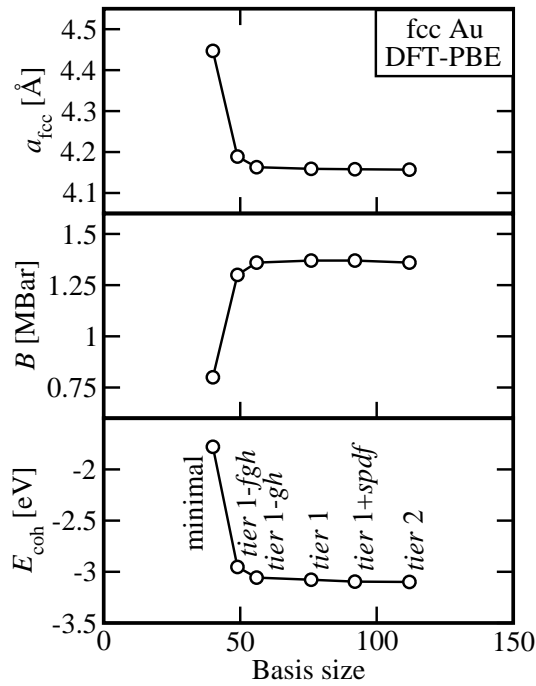


Fig. 6. Convergence of the lattice parameter a_{fcc} , bulk modulus B , and cohesive energy E_{coh} for fcc Au in DFT-PBE as a function of basis size. *tier 1-fgh* denotes *tier 1* as described in Table 1, but omitting the f , g , and h high-angular momentum functions; likewise for *tier 1-gh* but omitting only the g and h functions; finally, *tier 1+spdf* adds only the s , p , d , and f function of *tier 2* to the complete *tier 1*.

well with their literature counterparts (4.158 Å and 1.42 MBar, respectively [89]). Regarding the cohesive energy, *tier 2* yields only ≈ 20 meV improvement over the *tier 1* result. We find that similar convergence is achieved at the *tier 1* level also for other structures including Au. Again, the different convergence levels achieved in Fig. 2 are rather consistent with our conclusion that *tier 2* is required for light elements (H, C, N, O), while *tier 1* already performs very well for Au.

Effect of high angular momentum basis functions: Figure 6 also shows several basis sets that omit high angular momentum functions from the *tier* basis sets of Au. Omitting the g and h functions from *tier 1* does not significantly affect a_{fcc} and B . Especially for large-scale simulations, this basis set level thus enables rather accurate results at competitive cost (see Sect. 5 below). In contrast, visible changes arise when also the f function is left away. Similar conclusions hold also for energy differences other than a dense fcc bulk, e.g. for large-scale reconstructed Au surfaces [90].

In total, both test cases demonstrate how systematic convergence is achieved by using the successive *tiers* of Table 1 as basis sets. They also substantiate our observations regarding the element-specific accuracy of *tiers* near the end of Sec. 3.3. For example, *tier 1* is recommended to achieve a fast prerelax-

ation for light elements, but *tier 2* is needed for meV-level converged energy differences. In contrast, the (larger) *tier 1* for Au yields quite well converged energies, and its *spdf* section even suffices for faster “quick and dirty” calculations. In our own implementation [17], we therefore provide a hierarchy of predefined settings “really coarse”, “coarse”, and “standard” for all elements. These choices provide easy access to basis set levels and other settings at which fast (pre-)relaxations can safely be performed, up to the “standard” level where calculated energy differences can be expected to be converged at a level of few meV even for large systems. Beyond these predefined levels, the further *tiers* of our tabulated basis sets always provide a systematic path to verify the convergence of any calculated quantities explicitly.

4 Kohn-Sham equations, total energies and energy derivatives from NAOs: Numerical implementation

4.1 Numerical integration

Using NAO’s, Eqs. (2-7) require the performance of numerical integrations in three dimensions for a variety of tasks, most prominently in the setup of the Hamilton and overlap matrices, Eq. (5). As it stands, this task faces two challenges: First, it formally scales as $O(N^3)$ with system size, as all pairs of basis functions ($\sim N^2$) must be integrated across the entire system ($\sim N$). Since each basis function is non-zero only inside a fixed volume controlled by the confinement potential, this formal scaling is reduced to $O(N)$ in the limit of large systems, but the associated effort is significant and can be dominant for common system sizes (e.g., below ≈ 100 atoms for molecules). Second, one must integrate both the fast-varying localized wave function parts near the nuclei and the extended, smoother parts in interstitial and far-field regions with the same accuracy.

Several efficient integration schemes suitable for all-electron NAOs have been proposed in the literature (e.g., Refs. [10,24,26,91,92]). In practice, we opt to integrate all basis functions in a single sweep, using the now-standard partitioned integration technique on overlapping, atom-centered grids described by Becke, Delley, and others [10,25,91], which allows tightly converged calculations even for rather small grid sizes (e.g., $\approx 10,000$ grid points per atom for light elements).

In this scheme, each extended integrand $[\varphi_i(\mathbf{r})\hat{h}^{\text{KS}}\varphi_j(\mathbf{r})]$ (or any other integrand) is formally divided into localized atom-centered pieces by a “partition

of unity”:

$$\int d^3r \varphi_i(\mathbf{r}) \hat{h}^{\text{KS}} \varphi_j(\mathbf{r}) = \sum_{\text{at}} \int d^3r p_{\text{at}}(\mathbf{r}) \varphi_i(\mathbf{r}) \hat{h}^{\text{KS}} \varphi_j(\mathbf{r}), \quad (15)$$

where the sum of all atom-centered *partition functions* $p_{\text{at}}(\mathbf{r})$ equals one everywhere, by defining [10]

$$p_{\text{at}}(\mathbf{r}) = \frac{g_{\text{at}}(\mathbf{r})}{\sum_{\text{at}'} g_{\text{at}'}(\mathbf{r})}. \quad (16)$$

The normalizing sum over at' in the denominator runs over all atoms in the system, and g_{at} is a strongly peaked function about its originating atom. For example, Delley [10] suggested a simple intuitive approach similar to Hirshfeld partitioning [93],

$$g_{\text{at}}(\mathbf{r}) = \frac{n_{\text{at}}^{\text{free}}(r)}{r^2}, \quad (17)$$

where $n_{\text{at}}^{\text{free}}(r)$ is the electron density of non-spinpolarized, spherical free atoms. In contrast, more intricate schemes based on purely geometric considerations have also been put forward in the literature, most notably the detailed study by Stratmann and coworkers [25]. For brevity, we refer to their original work for implementation details, but note that this latter scheme allows a system-dependent reduction of $\approx 10\text{-}20\%$ in grid points compared to Eq. (17). While both prescriptions are of similar accuracy for many problems, the method of Ref. [25] proved to be strikingly more accurate than Eq. (17) for corner cases of very open systems (e.g., the fully extended Polyalanine structure of Fig. 5a), allowing additional computational savings through reduced overall integration grids.

Each single-atom integrand is integrated on its own grid of N_r spherical integration shells $r(s)$ ($s = 1, \dots, N_r$), on which angular integration points Ω_t with tabulated integration weights $w_{\text{ang}}(t)$ are distributed so as to integrate angular momentum functions up to a certain order exactly [94,95] (often called Lebedev grids [96–98]). Specifically, we employ a version provided by Delley with an accuracy of 17 digits [99]. The most efficient location of radial grid shells has been discussed by several authors [91,100,101]; our choice again goes back to Delley [101],

$$r(s) = r_{\text{outer}} \cdot \frac{\log\{1 - [s/(N_r + 1)]^2\}}{\log\{1 - [N_r/(N_r + 1)]^2\}}. \quad (18)$$

The radial integration weights are obtained from

$$w_{\text{rad}}(s)ds = r(s)^2 dr(s) = r(s)^2 \frac{dr}{ds}(s)ds. \quad (19)$$

The radius of the outermost shell, r_{outer} , and N_r specify the accuracy. Since $r(0) = 0$ and $r(N_r + 1) = +\text{infinity}$ by construction, a simple uniform accuracy

increase can be obtained by placing additional shells at integer fractions of the original grid, e.g., at $s = \frac{1}{2}$, $s = \frac{3}{2}$, ..., $s = 2N_r + 1/2$.

In production calculations, the number of angular grid points on a given radial shell r_s must still be specified. In principle, it is possible to determine the required angular grid for any given structure rigorously and *adaptively*, by converging the initial (non-selfconsistent) matrix elements s_{ij} and h_{ij} to better than a given threshold on each radial shell. We have implemented this algorithm for cluster-type geometries. In production calculations, it is, however, advantageous to specify conservative integration grids that are independent of the particular structure, in order to retain some error cancellation in energy differences. As in the case of basis sets, we therefore use symmetric dimers to determine the necessary grid density per radial shell for a given element adaptively, and then use the results as predefined, fixed grids in production calculations.

It is important to note that near-nuclear shells require significantly fewer points than those that touch the interstitial region. The number of angular grid points per shell thus increases from some value n_{\min} near the nucleus to a maximum n_{\max} on faraway shells. We find that a conservative choice $n_{\min}=110$ ensures noise-free forces also for high angular momentum components in the Hartree potential. For light-element molecules, we find $n_{\max}=302$ to be sufficient for accurate geometry relaxations, and $n_{\max}=590$ points to give total energies converged to $\sim 10^{-5}$ meV/atom even for very large molecules. The increase between n_{\min} and n_{\max} is then controlled through the adaptive scheme above.

In total, Eq. (15) is thus shortened to a sum over integration points

$$\int d^3r \varphi_i(\mathbf{r}) \hat{h}^{\text{KS}} \varphi_j(\mathbf{r}) = \sum_{\text{at}, s, t} w(\mathbf{r}) \varphi_i(\mathbf{r}) \cdot [\hat{h}^{\text{KS}} \varphi_j(\mathbf{r})] \quad , \quad (20)$$

where $\mathbf{r} = \mathbf{r}(\text{at}, s, t)$ and $w(\mathbf{r}) = p_{\text{at}}(\mathbf{r}) \cdot w_{\text{rad}}(s) \cdot w_{\text{ang}}(t)$. As we will discuss in detail in Ref. [102], the actual evaluation of Eq. (20) proceeds then in the following steps:

(1) The full integration grid is subdivided into localized *batches* of $O(100)$ grid points each, using a recursive algorithm: We split the full grid by finding a cut-plane which minimizes the sum of distance squares of all grid points to that plane [103]. The same operation is then applied to the two resulting groups of points, etc., until a predetermined maximum batch size is reached in each case.

(2) In each batch, the N_{nz} basis functions are found that exceed a given threshold η_{wave} on at least one point in the batch (called the “non-zero” functions out of the the system-wide N_{b} basis functions below). If the basis size *per atom* is fixed, N_{nz} approaches a constant with growing system size N , leading to $O(N)$ -like scaling for all integrals.

(3) For each batch, the vectors $\varphi_i(\mathbf{r})$ and $\hat{h}^{\text{KS}}\varphi_j(\mathbf{r})$ are evaluated point by point, but only if a given basis function is non-zero at that point.

(4) Within each batch, Eq. (15) is evaluated for all points at once, using a matrix multiplication over \mathbf{r} for the N_{nz} basis functions:

$$h_{ij}^{\text{batch}} = \sum_{\mathbf{r} \text{ in batch}} a_{ir} b_{rj} \quad , \quad (21)$$

with $a_{ir} = w(\mathbf{r}) \cdot \varphi_i(\mathbf{r})$ and $b_{rj} = \hat{h}^{\text{KS}}\varphi_j(\mathbf{r})$.

(5) The $(N_{\text{nz}} \times N_{\text{nz}})$ matrix elements h_{ij}^{batch} are added into the system-wide $(N_{\text{b}} \times N_{\text{b}})$ Hamilton matrix h_{ij} .

Steps (1-5) are trivially parallelized by distributing the batches across all available CPUs so as to ensure an even workload distribution. The only communication required are global matrix sums at the end of each integration. In addition, any large array stored on the grid [e.g., $n(\mathbf{r})$] is automatically memory-parallel. In massively parallel environments and/or in the limit of very large systems, it is important to note that the system-wide Hamilton matrix h_{ij} grows in memory as $O(N^2)$. To minimize communication, step (5) requires that a copy of h_{ij} is kept on every single CPU. We alleviate this memory bottleneck by using compressed row storage for h_{ij} , and by ensuring that each individual MPI task handles only batches that are spatially close together. Step (5) can then be performed using an efficient sorting algorithm and only for the fraction of matrix elements h_{ij} that are touched by the present CPU.

For the case of periodic boundary conditions, the Kohn-Sham equations, Eq. (4), become \mathbf{k} -space dependent, with separate matrices $h(\mathbf{k})$, $s(\mathbf{k})$, and solutions $\psi_{l,\mathbf{k}}(\mathbf{r})$ for different \mathbf{k} -points in the first Brillouin zone. This situation is handled by formally defining Bloch-like generalized basis functions $\chi_{i,\mathbf{k}}(\mathbf{r})$ from real-space basis functions $\varphi_i(\mathbf{r})$ that are centered in unit cells shifted by translation vectors $\mathbf{T}(\mathbf{N})$ [$\mathbf{N} = (N_1, N_2, N_3)$]:

$$\chi_{i,\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{N}} \exp [i\mathbf{k}\mathbf{T}(\mathbf{N})] \cdot \varphi_i [\mathbf{r} - \mathbf{R}_{\text{at}} + \mathbf{T}(\mathbf{N})] \quad . \quad (22)$$

Formally, this definition leads to \mathbf{k} -dependent matrix elements

$$h_{ij}(\mathbf{k}) = \langle \chi_{i,\mathbf{k}} | \hat{h}^{\text{KS}} | \chi_{j,\mathbf{k}} \rangle = \sum_{\mathbf{M}, \mathbf{N}} \exp \{i\mathbf{k} [\mathbf{T}(\mathbf{N}) - \mathbf{T}(\mathbf{M})]\} \langle \phi_{i,\mathbf{M}} | \hat{h}^{\text{KS}} | \phi_{j,\mathbf{N}} \rangle \quad (23)$$

[with the usual real-space basis functions $\phi_{i,\mathbf{M}}$ and $\phi_{j,\mathbf{N}}$ centered in different unit cells \mathbf{M} and \mathbf{N} , and equivalent for $s_{ik}(\mathbf{k})$]. If the integrals $\langle \phi_{i,\mathbf{M}} | \hat{h}^{\text{KS}} | \phi_{j,\mathbf{N}} \rangle$

were evaluated as they stand (with integration volumes extending over possibly several unit cells), one of the lattice sums in Eq. (23) would break down immediately due to periodicity (all terms depend only on $\mathbf{N} - \mathbf{M}$). In any case, only a finite number of real-space matrix elements are non-zero, since all basis functions are bounded by the confinement potential.

However, from a computational point of view it is more convenient to integrate (numerically) the matrix elements $\langle \phi_{i,\mathbf{M}} | \hat{h}^{\text{KS}} | \phi_{j,\mathbf{N}} \rangle$ not as single integrals extended over their whole respective integration volume, but rather in separate pieces, as partial integrals that extend over the volume of just one unit cell:

$$h_{ij}^{\text{uc}}(\mathbf{N}, \mathbf{M}) = \int_{\text{unit cell}} d^3r \phi_{i,\mathbf{M}}(\mathbf{r}) \hat{h}^{\text{KS}} \phi_{j,\mathbf{N}}(\mathbf{r}). \quad (24)$$

In practice, the appropriate integration grid for one unit cell is obtained by taking the overlapping atom-centered grid points (spread out over several unit cells), and mapping each grid point back to the first unit cell. We can then tabulate those basis functions $\{\phi_{i,\mathbf{N}}(\mathbf{r})\}$ that touch the first unit cell at all, to obtain all non-zero, inequivalent pieces of the real-space Hamiltonian matrix, $h_{ij}^{\text{uc}}(\mathbf{N}, \mathbf{M})$, within the whole structure. The advantage of this procedure is that we can then use the exact same integration steps (1-5) as in the cluster case. Using $h_{ij}^{\text{uc}}(\mathbf{N}, \mathbf{M})$, Eq. (23) can be rearranged to the formula

$$h_{ij}(\mathbf{k}) = \sum'_{\mathbf{M}, \mathbf{N}} \exp \{i\mathbf{k} [\mathbf{T}(\mathbf{N}) - \mathbf{T}(\mathbf{M})]\} \cdot h_{ij}^{\text{uc}}(\mathbf{N}, \mathbf{M}) \quad (25)$$

at each \mathbf{k} point. In contrast to Eq. (23), it is important to note that the sums \sum' over \mathbf{M} and \mathbf{N} in Eq. (25) are *not* lattice sums, but rather run only over those unit cells whose basis functions $\phi_{i,\mathbf{M}}$ and $\phi_{j,\mathbf{N}}$ have a non-zero piece in the first unit cell at all.

4.2 (Parallel) eigenvalue solution

In Kohn-Sham DFT, an accurate solution of the single-particle eigenvalue problem Eq. (4) is required. Traditional linear algebra for dense systems based on standard libraries (Lapack or ScaLapack [104,105]) provides a straightforward and robust solution for this step, but scales as $O(N^3)$ with system size in practice. The NAO basis set is particularly appealing in this respect in that the total number of basis functions N_b for a given structure is rather small compared to other types of basis functions even for converged production calculations, typically only a single-digit factor larger than the number of occupied states, N_{occ} . Thus, the use of standard eigenvalue solvers is not a problem for most relevant system sizes, despite the unfavorable scaling. Only for rather large systems (e.g., $\approx 1,000$ atoms for light elements as shown in

Sect. 5) does the solution time for Eq. (4) become dominant compared to all other computational steps.

In serial computations, we thus presently simply employ standard Lapack [104] library subroutines: first, a Cholesky decomposition $L_{ik}L_{jk}^T$ of s_{ij} to transform the non-orthonormal generalized eigenvalue problem to a standard one using the matrix $L_{ik}^{-1}h_{kr}L_{rj}^{-T}$. Then follows a transformation to tridiagonal form before finding the eigenvalues and eigenvectors.

In (massively) parallel computations, the key challenge is to preserve the low relative time requirements for Eq. (4) by ensuring near-ideal scaling to as many CPUs as possible. In our implementation, the exact eigenvalue solution is still performed through Scalapack [105] library routines, albeit with some customizations and ongoing work towards scalability improvements over many thousands of CPUs, as will be described in more detail in Ref. [106]. In brief, we note here that it is advantageous to employ the divide-and-conquer variant of the actual eigensolver (ScaLapack subroutine `pdstedc`). In addition, using backward / forward substitution to calculate the product of the inverse Cholesky factors with h_{ij} does not scale beyond a few hundred CPU's. This limitation follows from the fact that backward / forward substitution always employs only a few rows / columns of the respective matrices at the very beginning or end of the calculations. We thus compute the inverse of the Cholesky factor L_{ij} for this step. Doing so lets us employ standard matrix multiplication which scales much better. Together with compressed storage of h_{ij} in other phases of the calculation, only its local subsets needed for integration are kept on each CPU, and the procedure is fully memory-parallel.

There are two separate limits to our current (Sca)Lapack implementation: First (and trivially), very large systems (thousands of atoms and beyond) for which the simple $O(N^3)$ scaling becomes dominant, and second, massively parallel code execution on thousands of CPU's, which reaches ScaLapack's scaling limits. Currently, the ScaLapack tridiagonalisation step in the procedure is the limiting factor in the parallel scaling, as also noted by others [107]. To overcome the scaling limits of (Sca)Lapack and similar generic solutions, many alternative approaches have successfully been introduced for more specific contexts. Among these, iterative eigenvalue solver strategies (e.g., Refs. [108–113,107]) have been particularly successful in electronic structure theory, and we are working towards such an implementation [114].

A basis of NAOs centered at different sites is necessarily non-orthogonal (i.e., $s_{ij} \neq \delta_{ij}$). As the basis set grows, different linear combinations of basis functions can thus be expressed increasingly well by one another, ultimately threatening an ill-conditioned eigenvalue problem Eq. (4). In general, we do *not* observe any such ill-conditioning for clusters or large periodic cells even for very large production basis sets (*tier* 3 or larger in our nomenclature of Table

1), although this might be different for highly compressed materials that occur at the center of planets and stars. Nevertheless, ill-conditioned eigenvalue problems may arise in two distinct scenarios: (i) very large basis sets which are deliberately increased towards (over-)completeness, e.g., for stringent convergence tests; (ii) small unit-cell periodic calculations with very dense \mathbf{k} -meshes and large basis sets. Case (ii) is apparently a consequence of “pipelining” many different \mathbf{k} -points through Eq. (25), in each case combining only slightly different phase factors with the same real-space matrix elements h_{ij}^{uc} .

In any case, a standard safeguard against accidental ill-conditioning is simple: First, transform s_{ij} to its diagonal form explicitly. Then, project out any of its eigenvectors with suspiciously small eigenvalues (e.g., 10^{-5} or below). In addition, very large basis sets can also be stabilized by simply increasing other accuracy settings, e.g., the integration grids. This reduces any minor numerical errors that could be “collected” into spurious eigenvectors of s_{ij} .

4.3 Electron density and local potentials

Based on updated Kohn-Sham wave functions $\psi_l(\mathbf{r})$, the next step is to obtain an updated electron density $n(\mathbf{r})$, and, from it, the local potentials $v_{\text{es}}(\mathbf{r})$ and $v_{\text{xc}}(\mathbf{r})$.

There are two numerical routes to obtain the electron density $n(\mathbf{r})$ from a given set of Kohn-Sham eigenvectors c_{il} : either an eigenstate-based update procedure that scales as $O(N^2)$, or a density matrix based update that scales as $O(N)$ for large systems but requires more numerical effort for small systems.

Eigenstate-based density update: The straightforward way to the electron density is to evaluate each occupied Kohn-Sham orbital $\psi_l(\mathbf{r})$ ($l = 1, \dots, N_{\text{occ}}$, where N_{occ} denotes the highest state with a non-zero occupation number f_l) at each integration point, and then build the electron density by a sum of squares:

$$n(\mathbf{r}) = \sum_l^{N_{\text{occ}}} f_l |\psi_l(\mathbf{r})|^2. \quad (26)$$

Derivatives of $n(\mathbf{r})$ can be obtained in the exact same way, except that derivatives of $\psi_l(\mathbf{r})$ are now required, with each component (three for the gradient and six for the Hessian) as expensive as the density itself. In practice, Eq. (26) requires a matrix product $\psi_l(\mathbf{r}) = \sum_i c_{il} \varphi_i(\mathbf{r})$ for each batch of grid points \mathbf{r} . While the number of non-zero basis functions per point becomes constant with system size, the same is not true for the extended Kohn-Sham eigenstates. The simultaneous growth of the grid and of N_{occ} with system size thus render the effort for Eq. (26) as $O(N^2)$.

Density matrix based density update: As soon as N_{occ} becomes larger than the number of non-zero basis functions per integration point, it is favorable to switch to an electron density update based on the density matrix,

$$n_{ij} = \sum_l^{\text{occ.}} f_l c_{il} c_{jl}. \quad (27)$$

The electron density is then obtained by an operation which involves only locally non-zero basis functions:

$$n(\mathbf{r}) = \sum_{ij} \varphi_i(\mathbf{r}) n_{ij} \varphi_j(\mathbf{r}) \quad . \quad (28)$$

The resulting single matrix product per batch of grid points \mathbf{r} and single dot product per grid point scale as $O(N)$.

In periodic systems, the sum over occupied states in Eq. (27) implicitly includes \mathbf{k} -points and phase factors. Equation (28) is then evaluated entirely on the real-space grid, summing over all real-space basis functions φ_i that touch the first unit cell.

From n and its gradients, we can compute the local electron-electron interaction potentials, $v_{\text{es}}(\mathbf{r})$ and $v_{\text{xc}}(\mathbf{r})$. LDA or GGA exchange-correlation functionals $E_{\text{xc}}[n] = \int d^3r f_{\text{xc}}[n(\mathbf{r}), |\nabla n(\mathbf{r})|^2]$ are straightforward to compute per integration point and not a time or memory bottleneck. However, the evaluation of $|\nabla n(\mathbf{r})|$ makes GGAs overall slightly more expensive than LDA (since the expensive operations in the electron density update must be replicated for each of the three gradient components as well).

For GGAs, the *full* XC potential is yet more expensive because it depends explicitly on the Hessian (with six independent components) of n :

$$v_{\text{xc}} = \frac{\delta E_{\text{xc}}[n]}{\delta n} = \frac{\partial f_{\text{xc}}}{\partial n} - 2\nabla \left[\frac{\partial f_{\text{xc}}}{\partial |\nabla n|^2} \cdot \nabla n \right]. \quad (29)$$

As long as only total energies and eigenvalues are required, the Hessian can be avoided through an integration by parts for h_{ij} and the XC potential energy [115], i.e.,

$$\begin{aligned} & \int d^3r \varphi_i \varphi_j \cdot \nabla \left\{ (\nabla n) \cdot \frac{\partial}{\partial |\nabla n|^2} [f_{\text{xc}}(n, |\nabla n|)] \right\} \\ &= \int d^3r \nabla [\varphi_i \varphi_j] \cdot (\nabla n) \cdot \frac{\partial}{\partial |\nabla n|^2} f_{\text{xc}}(n, |\nabla n|) \quad . \end{aligned} \quad (30)$$

The explicit Hessian of n is thus only needed for energy derivatives (forces).

Since the electron density $n(\mathbf{r})$ is tabulated on the non-uniform, atom-centered grid, the electrostatic potential $v_{\text{es}}(\mathbf{r}) = \int d^3r' \frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}$ is evaluated directly in an atom-centered form, using partitioned overlapping electron densities that are localized around each atom as described by Delley [10,116] for cluster-type and periodic geometries. We begin by subtracting a superposition of non-spinpolarized spherical free-atom densities from the full electron density,

$$\delta n(\mathbf{r}) = n(\mathbf{r}) - \sum_{\text{at}} n_{\text{at}}^{\text{free}}(|\mathbf{r} - \mathbf{R}_{\text{at}}|) \quad . \quad (31)$$

The free-atom electron densities $n_{\text{at}}^{\text{free}}(|\mathbf{r} - \mathbf{R}_{\text{at}}|)$ and their electrostatic potential counterparts $v_{\text{at}}^{\text{es,free}}(|\mathbf{r} - \mathbf{R}_{\text{at}}|)$ are accurately known as cubic spline functions on dense logarithmic grids. The remaining difference density $\delta n(\mathbf{r})$ is decomposed into atom-centered multipole components, using expressions exactly equivalent to the partitioned integration formalism Eq. (15),

$$\delta \tilde{n}_{\text{at},lm}(r) = \int_{r=|\mathbf{r}-\mathbf{R}_{\text{at}}|} d^2\Omega_{\text{at}} p_{\text{at}}(\mathbf{r}) \cdot \delta n(\mathbf{r}) \cdot Y_{lm}(\Omega_{\text{at}}) \quad . \quad (32)$$

The multipole components $\delta \tilde{n}_{\text{at},lm}(r)$ are then spline-interpolated onto a dense logarithmic grid. The Green's-function solution to Poisson's Equation for multipoles from classical electrostatics [69] connects $\tilde{n}_{\text{at},lm}(r)$ to electrostatic multipole potentials $\tilde{v}_{\text{at},lm}(r)$:

$$\delta \tilde{v}_{\text{at},lm}(r) = \int_0^r dr_{<} r_{<}^2 g_l(r_{<}, r) \delta \tilde{n}_{\text{at},lm}(r_{<}) + \int_r^\infty dr_{>} r_{>}^2 g_l(r, r_{>}) \delta \tilde{n}_{\text{at},lm}(r_{>}) \quad . \quad (33)$$

The Green function for the unscreened Hartree potential is $g_l(r_{<}, r_{>}) = r_{<}^l / r_{>}^{l+1}$.

We use a third-order linear multistep integrator [117] to compute $\delta \tilde{v}_{\text{at},lm}(r)$, which are then splined and reassembled to the full electrostatic potential of $\delta n(\mathbf{r})$,

$$\begin{aligned} \delta v_{\text{es}}(\mathbf{r}) &= \sum_{\text{at},lm} \delta \tilde{v}_{\text{at},lm}(|\mathbf{r} - \mathbf{R}_{\text{at}}|) Y_{lm}(\Omega_{\text{at}}), \text{ and} \\ v_{\text{es}}(\mathbf{r}) &= \sum_{\text{at}} v_{\text{at}}^{\text{es,free}}(|\mathbf{r} - \mathbf{R}_{\text{at}}|) + \delta v_{\text{es}}(\mathbf{r}) \quad . \end{aligned} \quad (34)$$

The sum runs over all atoms and m values, with l bounded by a specified cutoff $l_{\text{max}}^{\text{es}}$.

The main computational effort in determining $v^{\text{es}}(\mathbf{r})$ is given by the sum in Eq. (34), which grows as $(l_{\text{max}}^{\text{es}})^2$ and which must be evaluated separately at every grid point \mathbf{r} . Thus, $l_{\text{max}}^{\text{es}}$ should be chosen as low as possible while still avoiding any significant cutoff errors. Fortunately, quadratic total energy convergence with the electron density expansion error for finite $l_{\text{max}}^{\text{es}}$ can be achieved by

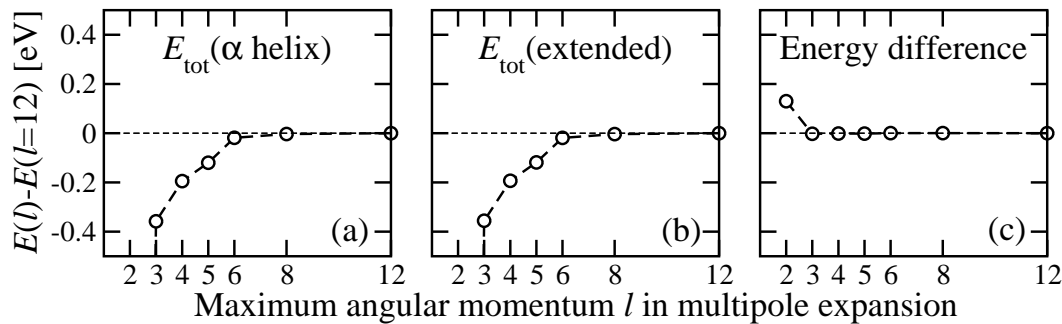


Fig. 7. Convergence of total energies and the energy difference of two different conformers of the polypeptide Ala₂₀ (Fig. 5a and b) with $l_{\text{max}}^{\text{es}}$ [Eq. (34)], at the *tier* 1 basis set level in DFT-PBE (see Fig. 5c). Reference energy in all cases: $l_{\text{max}}^{\text{es}}=12$. (a) Total energy convergence of the α helix. (b) Total energy convergence of the extended structure. (c) Convergence of the energy difference between α -helical and extended conformer.

computing the electrostatic double-counting correction in Eq. (7) as [39]

$$\frac{1}{2} \int d^3r n_{\text{MP}}(\mathbf{r}) \cdot v_{\text{es}}(\mathbf{r}), \quad (35)$$

where

$$n_{\text{MP}}(\mathbf{r}) = \sum_{\text{at}} n_{\text{at}}^{\text{free}}(|\mathbf{r} - \mathbf{R}_{\text{at}}|) + \sum_{\text{at}, lm} \delta \tilde{n}_{\text{at}, lm}(|\mathbf{r} - \mathbf{R}_{\text{at}}|) Y_{lm}(\Omega_{\text{at}}), \quad (36)$$

i.e., using the multipole expansion $n_{\text{MP}}(\mathbf{r})$ (capped at $l_{\text{max}}^{\text{es}}$) for the electron density also in the double-counting expression.

A detailed convergence analysis including also the electron density and potential components themselves has also been given by Delley [10]. For our own implementation, we display the energy convergence with $l_{\text{max}}^{\text{es}}$ in Fig. 7 for the molecular total energies and energy difference of the fully extended vs. the α -helical Ala₂₀ conformers already discussed in Sect. 3.5 (Fig. 5) above. The total energies for both conformers are converged to ≈ 0.2 eV at $l_{\text{max}}^{\text{es}}=4$ (i.e., ≈ 1 meV/atom). This total energy error is drastically diminished at $l_{\text{max}}^{\text{es}}=6$ (remaining error: 0.025 eV ≈ 0.1 meV/atom). Finally, the total energy error cancels practically completely in the energy *difference* between both conformers already at $l_{\text{max}}^{\text{es}} \gtrsim 3$. We have found these specific observations to hold across a wide range of systems. In practice, we find that $l_{\text{max}}^{\text{es}}=4$ is normally sufficient for meV/atom-converged energy differences, while truly tight convergence can be achieved with $l_{\text{max}}^{\text{es}}=6$. Similar expansion orders for local atomic potential components are found, e.g., in the full-potential (linearized) augmented plane wave [FP-(L)APW] [118] and other muffin-tin based methods. In any case, the convergence of $l_{\text{max}}^{\text{es}}$ can always be verified explicitly for key results.

Since Eq. (34) invokes all atoms on all grid points, it formally scales as $O(N^2)$ with system size. However, the extent of all electron density components

$\delta\tilde{n}_{\text{at},lm}(r)$ is bounded by the partition function $p_{\text{at}}(\mathbf{r})$, which in turn is limited by the free-atom electron density or other method used to create it [Eqs. (15) and (17), or Ref. [25]]. It is thus prudent to limit the range of the free-atom density itself by a separate, adjustable cutoff radius. In practice, its value can be chosen equal to the largest cutoff radius for any basis function associated with the atom in question. Outside the adjustable free-atom cutoff radius $\delta\tilde{v}_{\text{at},lm}(r)$ is then a simple analytical field that decays as $m_{\text{at},lm}r^{-(l+1)}$. Here, $m_{\text{at},lm}$ are the finite, classical multipole moments for the potential components $\delta\tilde{v}_{\text{at},lm}(r)$ outside the radius where the density component $\delta\tilde{n}_{\text{at},lm}(r)$ becomes zero. As pointed out in Ref. [26], the single slow-decaying $l = 0$ component normally does not die away anywhere even for the largest systems. Nonetheless, the mere elimination of the fastest-dying, most expensive high- l components improves the scaling of Eq. (34) significantly in practical calculations for large systems.

For periodic boundary conditions, the real-space electrostatic potential can be handled in much the same way, but with possible finite multipole moments $m_{\text{at},lm}$ for the difference density components $\delta\tilde{v}_{\text{at},lm}(r)$, long-range components from distant unit cells must now be accounted for. For the superposition of (neutral) free-atom potentials $v_{\text{at}}^{\text{es,free}}(r)$, this problem does not arise: The electronic charge is always compensated by the nucleus, and the sum $v_{\text{at}}^{\text{es,free}}(r) + Z/r$ is bounded by the extent of the free-atom density $n_{\text{at}}^{\text{free}}(r)$ (see also Sec. 4.6 for the resulting, “neutralized” total energy expressions). However, the same is not true for $\delta\tilde{v}_{\text{at},lm}(r)$ —in fact, for charged unit cells, the electrostatic potential would even diverge.

We follow Ref. [116] and solve the long-range electrostatic potential in periodic systems using Ewald’s method of adding and subtracting smooth compensating Gaussian charge densities around each multipole. While the compensating densities are added only formally, the respective compensating *potentials* are actually evaluated in real and reciprocal space. Effectively, this procedure splits each analytical potential tail $r^{-(l+1)}$ into a long-ranged and a short-ranged part, e.g.,

$$\frac{1}{r} = \frac{\text{erf}(r/r_0) + \text{erfc}(r/r_0)}{r} \quad (37)$$

(and similar for $l > 0$, with somewhat more complex expressions as given in detail in Ref. [116]; multipolar Ewald sums for overlapping densities have also been used in other implementations [12,119,120], including a full presentation in Ref. [121]). Only the short-range part is then accounted for in real space, while the long-range part is conveniently treated in reciprocal space.

In periodic systems, the zero level of the electrostatic potential is not necessarily unique in periodic boundary conditions (for a discussion, see Ref. [122]). We therefore set the potential zero to be the spatial average of $\delta v_{\text{es}}(\mathbf{r})$ [Eq. (34)].

For charged unit cells, any charge in the potential is effectively neutralized by leaving away the $\mathbf{G} = 0$ component of the reciprocal-space potential part. In effect, this procedure corresponds to adding a constant, neutralizing background charge density (“jellium background”) n_0 to the entire system. However, unlike in fully reciprocal space methods (where also the density is handled in reciprocal space), n_0 must then also be accounted for in those total-energy expressions that involve the real-space density (see Sec. 4.6).

The scaling parameter r_0 determines the smoothness of the splitting, and thus the maximum reciprocal lattice vector in the Fourier part of the potential, \mathbf{G}_{\max} . We estimate \mathbf{G}_{\max} by demanding that a conservative estimate of the smallest Fourier series component in the expansion still remains smaller than a given error threshold η . By direct comparison with the exact potential expression Eq. (4) in Ref. [116], we obtain

$$\frac{4\pi}{\Omega_{\text{cell}}} \cdot \frac{1}{\mathbf{G}_{\max}^2} \cdot \exp\left(-\frac{r_0^2 \mathbf{G}_{\max}^2}{4}\right) \cdot |\mathbf{G}_{\max}|^{l_{\max}^{\text{es}}} < \eta. \quad (38)$$

In principle, r_0 and η are adjustable parameters. In practice, we find that, e.g., $r_0 = 3.0$ bohr and $\eta = 5 \cdot 10^{-10}$ Ha guarantee full accuracy across a wide range of systems, without becoming a bottleneck in production calculations.

Like the Hamiltonian integration and electron density update, Eq. (34) parallelizes naturally with the integration grid distributed across different CPUs. However, each grid point \mathbf{r} needs access to most of the splined atomic potential components $\delta\tilde{v}_{\text{at},lm}(\mathbf{r})$ in the structure. Since this array grows linearly with system size, keeping a full copy on every CPU quickly becomes a memory bottleneck on low-memory, massively parallel architectures. Therefore, we optionally trade some additional communication for full memory parallelism memory of $\delta\tilde{v}_{\text{at},lm}(\mathbf{r})$: We first distribute $\delta\tilde{v}_{\text{at},lm}(\mathbf{r})$ across separate processes for different atoms. On every CPU, we then fetch the contribution from each atom one by one, thus building the full $v_{\text{es}}(\mathbf{r})$ and $n_{\text{es}}(\mathbf{r})$ step by step.

4.4 Achieving self-consistency

The chicken-egg problem of electronic structure theory is *self-consistency*: The correct electronic Hamiltonian h_{ij} should yield as output the same Kohn-Sham orbitals, electron density and potentials as were used to construct it. This subject continues to be a focus of much effort in the community; we here summarize the algorithms employed in our own implementation, including a generalization of the Kerker preconditioner [123–125] to the present all-electron case, using the electrostatic-potential framework outlined in the preceding section.

In the standard self-consistent field (s.c.f.) method, the problem is tackled it-

eratively. Beginning from an initial guess for the electron density, Kohn-Sham orbitals, or potential, a trial Hamiltonian $h_{ij}^{(0)}$ is constructed. The output wave functions $|\psi_l^{(1)}\rangle$ are used to construct an updated electron density, potentials, and thus an updated Hamiltonian $h_{ij}^{(1)}$, and so forth. The procedure is considered converged when the total energy, sum of eigenvalues, electron density difference norm and/or other quantities no longer change by more than a certain threshold.

When beginning a calculation from scratch, we form the initial electron density $n^{(0)}$ from a superposition of spherical free-atom electron densities:

$$n^{(0)}(\mathbf{r}) = \sum_{\text{at}} n_{\text{at}}^{\text{free}}(|\mathbf{r} - \mathbf{R}_{\text{at}}|). \quad (39)$$

For spin-polarized and / or charged systems, $n_{\text{at}}^{\text{free}}$ may be replaced by the electron density of free ions or spin-polarized atoms. By default, a spin-polarized calculation is initialized using spin-polarized free atoms occupied following Hund’s rules. For Hartree-Fock or hybrid functionals, the initial Hamiltonian is created with an initial DFT-LDA or -GGA s.c.f. cycle. The resulting Kohn-Sham orbitals $|\psi_l^{(1)}\rangle$ initialize the non-local exchange operator from the first Hartree-Fock (or hybrid) iteration onwards.

Alternatively, when a self-consistent result for a nearby structure is already available (e.g., from an earlier relaxation of molecular dynamics step), it is advantageous to reinitialize the electron density using that existing set of Kohn-Sham orbitals, after reorthonormalizing it for the new geometry.

Presently, we combine three separate approaches to ensure stable self-consistency:

1. Damping of electron density oscillations. During the s.c.f. cycle, we obtain updated Kohn-Sham orbitals $|\psi_l^{(\mu)}\rangle$ from a Hamiltonian $h_{ij}^{(\mu-1)}$, which is based on an electron density $n^{(\mu-1)}$. Before constructing the next Hamiltonian $h_{ij}^{(\mu)}$, we first dampen the density change between s.c.f iterations $(\mu - 1)$ and μ . To that end, we base $h_{ij}^{(\mu)}$ not directly on $n_{\text{KS}}^{(\mu)} := \sum_l f_l |\psi_l|^2$, but rather on:

$$n_{\text{dmp}}^{(\mu)} = n^{(\mu-1)} + \hat{G}^1 (n_{\text{KS}}^{(\mu)} - n^{(\mu-1)}). \quad (40)$$

If \hat{G}^1 is a constant, this is simple linear mixing between the present and previous electron densities. However, \hat{G}^1 may also be a non-local *preconditioner* (see below), damping, e.g., long-range density changes stronger than short-range ones [123–126].

2. Electron density mixing. In addition to simple damping, the s.c.f. procedure is further stabilized by mixing in earlier electron density iterations. We employ Pulay’s *direct inversion of the iterative subspace* (DIIS) [127] scheme in the modified form of Ref. [126], which keeps the electron density norm constant

by construction, rather than by way of a separate constraint. Based on the *actual* electron density residual between the present and previous iteration,

$$R^{(\mu)} = n_{\text{KS}}^{(\mu)} - n^{(\mu-1)} \quad ; \quad (41)$$

the goal is to estimate a linear combination of all previous electron densities that *would* have minimized the residual:

$$n_{\text{opt}} = n^{(\mu-1)} + \sum_{k=1}^{N_p} \bar{\alpha}_k (n^{(\mu-k+1)} - n^{(\mu-k)}) . \quad (42)$$

Kresse and Furthmüller [126] showed how to rewrite Pulay’s [127] original linear algebra to obtain $\bar{\alpha}_k$ from scalar products between differences of electron density residuals, $\langle R^{(\nu)} - R^{(\nu-1)} | R^{(\kappa)} - R^{(\kappa-1)} \rangle$. In practice, we combine n_{opt} with electron density damping to read

$$n^{(\mu)} = \hat{G}^1 \left[R^{(\mu)} + \sum_{k=1}^{N_p} \bar{\alpha}_k \left(R^{(\mu-k+1)} - R^{(\mu-k)} \right) \right] \quad (43)$$

$$+ n^{(\mu-1)} + \sum_{k=1}^{N_p} \bar{\alpha}_k \left(n^{(\mu-k)} - n^{(\mu-k-1)} \right) .$$

To retain a consistent electron density description, precisely the same mixing operations must be applied to both spin components, to density gradients, and to the density matrix in the case of single-determinant exchange (for hybrid functionals or Hartree-Fock).

3. Occupation broadening at the Fermi level. In metallic systems with many states close to the Fermi level, additional stability is achieved by occupying the states $|\psi_l\rangle$ by a distribution f_l which is slightly broadened about the Fermi level, e.g., the Fermi-function [48], Gaussian [128], or Methfessel-Paxton [129] broadening schemes. It is worth noting that an entropy function $S(f_l)$ can be associated with each partially occupied level (see, e.g., Refs. [126,130,131] for details). Its significance is that any energy derivatives (forces) that do not explicitly account for the change of f_l with atomic positions are no longer exact derivatives of the total energy Eq. (7). Rather, they correspond to the electronic *free energy* $F = E_{\text{tot}} - \sigma \sum_l S(f_l)$, where σ is the width of the broadening distribution.

In practice, Eq. (43) brings the number of required s.c.f. cycles for tight convergence of total energies and gradients to ≈ 10 -20 in standard cases (e.g., 14 s.c.f. cycles for 10^{-7} eV convergence of E_{tot} for the Al₂₀ molecules of Fig. 5 above), with the “damping operator” \hat{G}^1 chosen to be a simple constant. There is, however, no convergence guarantee, and certain pathological cases (e.g., certain metals or cluster geometries with crossing levels) may converge

significantly slower or not at all when \hat{G}^1 remains constant. In such cases, a more sophisticated choice of \hat{G}^1 provides significant additional flexibility to ensure a faster and more stable convergence [123–125].

In the plane wave community, a successful approach is to damp long-range oscillations in $R^{(\mu)}$ stronger than short ranged ones, using $\hat{G}^1 = Ak^2/(k^2 + k_0^2)$ (Kerker preconditioning [125]). A is the simple linear damping factor mentioned above and k_0 is a cutoff wavelength for the damping. We have implemented a real-space version of this method, formally:

$$\hat{G}^1 = A \left(1 - \frac{k_0^2}{k^2 + k_0^2} \right) \equiv A \left(1 + \frac{k_0^2}{\nabla^2 - k_0^2} \right) \quad (44)$$

[we here use operator language to clarify Eq. (46) below, i.e., $1/(\nabla^2 - k_0^2)$ should be read as the inverse of an *operator*]. The long-wave length components in a real-space electron density residual $R^{(\mu)}$ are thus reduced by adding a correction term $k_0^2 R_{\text{corr}}$:

$$\hat{G}^1 R^{(\mu)}(\mathbf{r}) = R^{(\mu)}(\mathbf{r}) + k_0^2 R_{\text{corr}} \quad (45)$$

R_{corr} is given by the modified Helmholtz equation,

$$(\nabla^2 - k_0^2) R_{\text{corr}}(\mathbf{r}) = R^{(\mu)}(\mathbf{r}). \quad (46)$$

This can be solved in real space by using the same multipole expansion method as for the electrostatic potential [Eqs. (31-34)]. The Green function for the solution of the preconditioned residual is [69]

$$g_l(r_<, r_>) = i_l(k_0 r_<) k_l(k_0 r_>) \quad . \quad (47)$$

The functions i_l and k_l are modified spherical Bessel functions of half-integer order, which can be easily calculated from recursion relations or series expansions for small arguments [117].

The numerical effort for our damping scheme is thus formally equivalent to that of the Hartree potential; however, since the main goal is to dampen the long-range components, the effort can be reduced by choosing a smaller l_{max} in Eq. (34) for the angular momentum expansion of \hat{G}^1 . In practice, e.g., in large metallic slabs, we find that even $l_{\text{max}}=0$ (damping the monopole component on each atom only) already provides the required stabilization effect of G^1 . We also find that $k_0 \approx 1.5 \text{ bohr}^{-1}$ often gives good performance, although the exact optimum choice for k_0 remains somewhat system-dependent.

Finally, we comment briefly on the details of evaluating the total energy, Eq. (7), during the approach to self-consistency. The Kohn-Sham energy functional is defined in Eq. (6) as

$$E_{\text{KS}} := T_s[n] + V_{\text{ext}}[n] + E_{\text{es}}[n] + E_{\text{xc}}[n] + E_{\text{nuc-nuc}} \quad (48)$$

for the self-consistent electron density n . When evaluating the same expression before self-consistency is reached, an ambiguity arises: Since the evaluation of the single-particle kinetic energy T_s requires the knowledge of a set of Kohn-Sham eigenstates consistent with n , we can generally evaluate T_s only for the *output* density from a given iteration, $n_{\text{KS}}^{(\mu)}$, but not for the (in general) mixed and preconditioned *input* density $n^{(\mu-1)}$ used to construct the Hamiltonian. Conversely, we compute the electrostatic potential v^{es} routinely for the input density after mixing, $n^{(\mu-1)}$, but not for the output density $n_{\text{KS}}^{(\mu)}$. In order to avoid having to evaluate the point-wise electrostatic potential, Eq. (34) twice, E_{es} is thus most economically evaluated for $n^{(\mu-1)}$, and not for $n_{\text{KS}}^{(\mu)}$.

The total energy expression that we thus evaluate while approaching self-consistency was first discussed by Wendel and Martin [132] and is commonly referred to as the *Harris functional* [133,134]. It combines precisely the pieces that are available anyway in the course of a regular self-consistency cycle:

$$E_{\text{Harris}}^{(\mu)} = \sum_{l=1}^{N_{\text{states}}} f_l^{(\mu)} \epsilon_l^{(\mu)} - \int d^3r [n^{(\mu-1)}(\mathbf{r}) v_{\text{xc}}[n^{(\mu-1)}](\mathbf{r})] + E_{\text{xc}}[n^{(\mu-1)}] \quad (49)$$

$$- \frac{1}{2} \int d^3r [n^{(\mu-1)}(\mathbf{r}) v_{\text{es}}^{(\mu-1)}(\mathbf{r})] + E_{\text{nuc-nuc}}.$$

Here, the single-particle eigenvalues supply the kinetic energy for the *output* electron density $n_{\text{KS}}^{(\mu)}$, while the double-counting corrections are strictly evaluated for the mixed and preconditioned electron density used to set up the Hamiltonian, $n^{(\mu-1)}$. Eq. (49) obviously agrees with the Kohn-Sham functional Eq. (6) for the final, self-consistent electron density. Moreover, Eq. (49) can be shown [135] to converge noticeably faster towards the final self-consistent energy for large and/or metallic systems than Eq. (6) when evaluated strictly for the same density, $n_{\text{KS}}^{(\mu)}$.

To check the s.c.f. convergence in practice, a convergence criterion of 10^{-6} eV for the variational total energy $E_{\text{Harris}}^{(\mu)}$ between successive iterations usually guarantees the convergence this quantity, but not necessarily the sufficient convergence of any non-variational quantities (sum of Kohn-Sham eigenvalues, forces). In particular, converged forces (typically, to 10^{-4} eV/Å between successive s.c.f. iterations) must thus be ensured explicitly. Since computing the full force expression (Sect. 4.7 below) in each s.c.f. cycle is expensive, we do not compute forces until *all* energy-based quantities are converged, which includes specifically also the non-variational sum of eigenvalues. We find that, once this quantity is sufficiently converged, only two additional s.c.f. iterations with forces (the minimal number to ensure explicit force convergence) are normally required.

4.5 Scalar Relativity

The single-particle kinetic energy operator of Schrödinger’s equation is a simple differential operator on a scalar wave function

$$\langle \mathbf{r} | \hat{t}_s | \psi \rangle = -\frac{\nabla^2}{2} \psi(\mathbf{r}). \quad (50)$$

In contrast, the correct relativistic description of electrons is given by Dirac’s four-component equation,

$$\begin{pmatrix} v & c\boldsymbol{\sigma} \cdot \mathbf{p} \\ c\boldsymbol{\sigma} \cdot \mathbf{p} & -2c^2 + v \end{pmatrix} \begin{pmatrix} \Psi \\ \chi \end{pmatrix} = \epsilon \begin{pmatrix} \Psi \\ \chi \end{pmatrix}. \quad (51)$$

where $\mathbf{p} = -i\nabla$ is the momentum operator, $\boldsymbol{\sigma}$ is a vector of Pauli spin matrices, and Ψ and χ themselves have two components each. For $\epsilon - v(\mathbf{r}) \ll 2c^2$, this reduces to Schrödinger’s Equation, which is obviously satisfied for valence-like electrons in the spatial region relevant for chemical bonding. However, $|\epsilon - v(\mathbf{r})|$ easily approaches $\approx 2c^2$ near the nuclear Coulomb singularity for *all* electrons. More, specifically, beyond the lightest elements, any wave functions with a non-negligible part near the nucleus are thus modified appreciably. Relativity thus affects not just deep core states, but also valence states (e.g., $6s$ and $6p$ like states in $5d$ transition metals), whereby chemical bonding properties are significantly affected as well (see Ref. [136] for a detailed discussion of relativistic effects).

To describe chemical properties correctly in a full-potential framework, our primary interest is to capture the *effect* of near-nuclear relativity on valence- and semicore-like electrons, while retaining the computational effort of a Schrödinger-like equation. We adopt here a one-component, scalar-relativistic scheme, which in its most general form is found by solving one of the sub-equations of Eq. (51) for χ and then eliminating χ from the other one. By neglecting spin-orbit coupling, we obtain:

$$\left(\mathbf{p} \cdot \frac{c^2}{2c^2 + \epsilon_l - v} \cdot \mathbf{p} + v \right) \psi_l = \epsilon_l \psi_l. \quad (52)$$

This expression is, e.g., the familiar one-dimensional Koelling-Harmon expression for scalar relativity [137]. Obviously, the presence of ϵ_l in the kinetic energy part is a complication, since this defines a different Hamiltonian $\hat{h}^{\text{KS}}(l)$ for each eigenstate ψ_l . [Note that the kinetic energy indeed reduces to the non-relativistic, ϵ_l -independent form Eq. (50) for $(\epsilon_l - v) \ll 2c^2$].

There are (at least) two ways to create a manageable yet accurate scalar-relativistic kinetic energy operator. One is to separate core and valence states

as done, e.g., in muffin-tin based methods (FP-(L)APW [138], augmented spherical waves [139], linearized muffin-tin orbitals [140], Korringa-Kohn-Rostoker [141]), i.e., solving separately for each localized core state and setting $\epsilon_l = 0$ uniformly for each valence state. The other is to approximate the scalar-relativistic Hamiltonian in Eq. (52) as a whole, and then correct for the approximation by a subsequent perturbative treatment.

We presently pursue the second option, using the scaled zeroth order regular approximation (ZORA) suggested by van Lenthe *et al.* [142]. In this scheme, self-consistent solutions $|\psi_l\rangle$ are first obtained using the ZORA [143–145] kinetic energy operator

$$\hat{t}_{\text{ZORA}} = \mathbf{p} \cdot \frac{c^2}{2c^2 - v} \cdot \mathbf{p}, \quad (53)$$

corresponding to $|\epsilon_l| \ll 2c^2$, a potentially harsh approximation for deep core states. After self-consistency is reached, each eigenvalue is rescaled using the “scaled ZORA” expression

$$\epsilon_l^{\text{scaled}} = \frac{\epsilon_l^{\text{ZORA}}}{1 + \langle \psi_l | \mathbf{p} \frac{c^2}{(2c^2 - v)^2} \mathbf{p} | \psi_l \rangle}. \quad (54)$$

For total energies [Eq. (7)], the sum of ZORA eigenvalues ϵ_l^{ZORA} is then replaced with the sum of scaled ZORA eigenvalues $\epsilon_l^{\text{scaled}}$.

One important feature of *scaled* ZORA is that it restores almost exactly the gauge invariance against shifts of the potential zero, which is lost in Eq. (53). Van Lenthe *et al.* [142] showed how this rescaling can be justified by way of a Foldy-Wouthuysen transform [146] of the Dirac Hamiltonian. As a consequence, the scaled ZORA Hamiltonian is for example exact for a one-electron hydrogen-like atom (this is relevant because the near-nuclear potential resembles this situation in any atomic structure). For completeness, we note that it is also possible to perform a simple separation between localized core states and extended semicore/valence states, enabling an accurate single-site treatment of core-state scalar relativity according to Eq. (52), as done, e.g., in the FP-(L)APW method [147].

It has been noted [148,149] that the missing gauge invariance in ZORA is also restored by substituting in Eq. (53) for v only the on-site free-atom potential $v_{\text{at}(j)}^{\text{free}}$ at the atomic center [at(j)] associated with a basis function j :

$$\hat{t}_{\text{at.ZORA}}|\varphi_j\rangle = \mathbf{p} \cdot \frac{c^2}{2c^2 - v_{\text{at}(j)}^{\text{free}}} \cdot \mathbf{p}|\varphi_j\rangle. \quad (55)$$

We refer to Eq. (55) as “atomic ZORA”. Because $\hat{t}_{\text{at.ZORA}}$ now depends on

basis function j , an explicit symmetrization is needed for matrix elements:

$$T_{ij} = \frac{1}{2} \left(\langle \varphi_i | \hat{t}_{\text{at.ZORA}} | \varphi_j \rangle + \langle \varphi_j | \hat{t}_{\text{at.ZORA}} | \varphi_i \rangle \right). \quad (56)$$

As is illustrated further below, our present experience is that relaxed atomic geometries from this approximation correspond practically exactly to both scaled ZORA results and to FP-(L)APW benchmark results.

In practical calculations, we can use the kinetic energy expression Eq. (53) directly in the construction of the free-atom “minimal basis” functions, by integrating Eq. (8) on an accurate logarithmic grid as in the non-relativistic case. However, a subtle issue arises for three-dimensional integrations on a sparser radial grid, such as Eq. (18). Compared to the non-relativistic case, a direct integral

$$T_{ij} = - \int d^3r \left\{ \varphi_i(\mathbf{r}) \cdot \nabla \left[\frac{c^2}{2c^2 - v(\mathbf{r})} \cdot \nabla \varphi_j(\mathbf{r}) \right] \right\} \quad (57)$$

would require a derivative of the local potential $v(\mathbf{r})$. We can circumvent this by an integration by parts, [142]

$$T_{ij} = - \int d^3r \left\{ [\nabla \varphi_i(\mathbf{r})] \cdot \frac{c^2}{2c^2 - v(\mathbf{r})} \cdot [\nabla \varphi_j(\mathbf{r})] \right\}, \quad (58)$$

and the latter expression can be evaluated without potential derivatives. For GGA’s, the contribution of the explicit gradient derivative in Eq. (29) to the local potential is not evaluated during a normal s.c.f. cycle, since it is itself already treated by an integration by parts. Since this is small compared to $[2c^2 - v(\mathbf{r})]$, we omit it in the denominator of Eq. (58).

The direct expression Eq. (57) benefits from an additional cancellation of integration errors near the nucleus. In this region, we have for all free-atom like basis functions $|\varphi_{\text{at}}\rangle$ (e.g., all core functions):

$$(\hat{t}[v] + v)|\varphi_{\text{at}}\rangle \approx (\hat{t}[v_{\text{at}}^{\text{free}}] + v)|\varphi_{\text{at}}\rangle = (\epsilon_{\text{at}} - v_{\text{at}}^{\text{free}} + v)|\varphi_{\text{at}}\rangle. \quad (59)$$

The difference $v - v_{\text{at}}^{\text{free}}$ cancels the deep Coulomb singularity at the origin; in the non-relativistic case, this is exact. The constant free-atom eigenvalue ϵ_{at} is accurately known from solving Eq. (8) on a logarithmic grid. In contrast, Eq. (58) does not allow for the same cancellation, and would therefore need a denser integration grid near the nucleus than the non-relativistic case.

In order to restore the integration accuracy on the standard radial integration grid, Eq. (18), we therefore integrate instead the difference between the kinetic energy for the full potential and that of a sum of free-atom potentials:

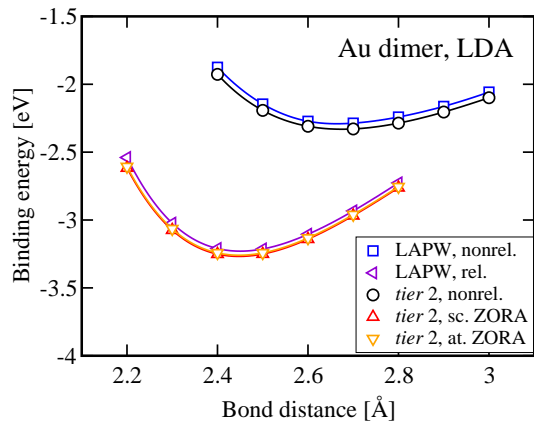


Fig. 8. Different levels of relativistic theory at the *tier 2* NAO basis set level compared to benchmark FP-(L)APW calculations [118,150] for the binding curve for Au₂ in DFT-LDA. The “atomic ZORA” [Eq. (55)] and “scaled ZORA” [Eq. (53)] for self-consistency, with perturbative post-processing according to (54)] approximations agree closely with relativistic Wien2k benchmark results. For comparison, the non-relativistic binding curves (same level of agreement) are also shown.

$$T_{ij} = - \int d^3r \left\{ [\nabla\varphi_i(\mathbf{r})] \cdot \left[\frac{c^2}{2c^2 - v(\mathbf{r})} - \frac{c^2}{2c^2 - \sum_{\text{at}} v_{\text{at}}^{\text{free}}(\mathbf{r})} \right] \cdot \nabla\varphi_j(\mathbf{r}) \right\} \quad (60)$$

$$+ \int d^3r \left\{ \varphi_i(\mathbf{r}) \cdot \nabla \left[\frac{c^2}{2c^2 - \sum_{\text{at}} v_{\text{at}}^{\text{free}}(\mathbf{r})} \cdot \nabla\varphi_j(\mathbf{r}) \right] \right\}.$$

The first integrand (the difference term) is now small by construction. The computational effort for this term can additionally be reduced by evaluating it only at grid points where the integrand is larger than a certain threshold. The second integrand exhibits the desired error cancellation, and can be integrated as it stands since the analytic derivative of the sum of free-atom potentials $\sum_{\text{at}} v_{\text{at}}^{\text{free}}(\mathbf{r})$ is accurately known on a logarithmic grid.

Figure 8 illustrates the quality of our present relativistic treatment for a simple testcase, the Au dimer in DFT-LDA, compared to benchmark relativistic full-potential (L)APW calculations [118,150]. The NAO binding curves of Au₂ were obtained using the well-converged *tier 2* basis set of Table 1 (cf. also Fig. 2). The relativistic equilibrium distance at the scaled ZORA and atomic ZORA levels is 2.45 Å, with binding energies of −3.27 eV and −3.26 eV, respectively. These results and binding curves compare almost exactly to the FP-(L)APW results (2.45 Å/−3.23 eV), which are also shown in the figure. We note in passing that the same level of agreement is attained between both methods for the non-relativistic case.

4.6 Non-divergent total energy for non-periodic, neutral and charged periodic systems

As already noted, we use Eq. (7) viz. Eq. (49) for practical total energy computations. As written, both equations contain as separate terms (i) the double-counted classical interelectronic repulsion energy, and (ii) the internuclear repulsion energy of the entire system:

$$E_{\text{double}} + E_{\text{nuc-nuc}} = -\frac{1}{2} \int d^3r [n(\mathbf{r})v_{\text{es}}(\mathbf{r})] + \frac{1}{2} \sum_{\text{at} \neq \text{at}'} \frac{Z_{\text{at}}Z_{\text{at}'}}{|\mathbf{R}_{\text{at}} - \mathbf{R}_{\text{at}'}|}. \quad (61)$$

Both terms are straightforward to compute in cluster systems, but unfortunately, each one on its own diverges with increasing system size, and outrightly for periodic boundary conditions.

To ensure a stable total energy expression, Eq. (61) is rewritten [12] to exploit the combined charge neutrality of all electrostatic contributions (electronic and nuclear) associated with a given atom. Similar ideas have been employed, e.g., in early muffin-tin based work [151–153], and for *overlapping* electron density partitions in Refs. [119–121], closer to the procedure summarized below.

The (electrostatic) electron+nuclear potential for each atom reads

$$v_{\text{at}}^{\text{es,tot}}(r) = v_{\text{at}}^{\text{es}}(|\mathbf{r} - \mathbf{R}_{\text{at}}|) + \frac{Z_{\text{at}}}{|\mathbf{r} - \mathbf{R}_{\text{at}}|}, \quad (62)$$

where

$$v_{\text{at}}^{\text{es}}(|\mathbf{r} - \mathbf{R}_{\text{at}}|) = v_{\text{at}}^{\text{es,free}}(|\mathbf{r} - \mathbf{R}_{\text{at}}|) + \sum_{lm} \delta \tilde{v}_{\text{at},lm}(|\mathbf{r} - \mathbf{R}_{\text{at}}|) Y_{lm}(\Omega_{\text{at}}) \quad (63)$$

sums over all (partitioned) electronic electrostatic contributions due to atom at. It is important to note that, by construction, the free-atom like part of $v_{\text{at}}^{\text{es,tot}}(r)$ is strictly localized within the radius of the electron density of a free atom, which in turn can be restricted by a global confining potential.

We rewrite Eq. (61) to reflect separately the energy of the electronic charge density and the energy of all nuclei in the combined electron-nuclear potentials $v_{\text{at}}^{\text{es,tot}}(r)$. Turning to the electronic part first, we add

$$\begin{aligned} - \int d^3r n(\mathbf{r})v_{\text{es}}(\mathbf{r}) &\rightarrow - \int d^3r n(\mathbf{r})v_{\text{es}}(\mathbf{r}) - \int d^3r n(\mathbf{r}) \sum_{\text{at}} \frac{Z_{\text{at}}}{|\mathbf{r} - \mathbf{R}_{\text{at}}|} = \\ &\quad - \int d^3r n(\mathbf{r}) \left[\sum_{\text{at}} v_{\text{at}}^{\text{es,tot}}(|\mathbf{r} - \mathbf{R}_{\text{at}}|) \right] \end{aligned} \quad (64)$$

by straightforward algebraic manipulation (but note the comment regarding charged periodic systems below).

The nuclear part of Eq. (61) is equally straightforward by exploiting the definition of the electronic electrostatic potential at each nucleus,

$$\int d^3r \frac{n(\mathbf{r})}{|\mathbf{r} - \mathbf{R}_{\text{at}}|} = v_{\text{es}}(\mathbf{R}_{\text{at}}). \quad (65)$$

Then,

$$\begin{aligned} \sum_{\text{at}, \text{at}' \neq \text{at}} \frac{Z_{\text{at}} Z_{\text{at}'}}{|\mathbf{R}_{\text{at}} - \mathbf{R}_{\text{at}'}|} &\rightarrow \sum_{\text{at}, \text{at}' \neq \text{at}} \frac{Z_{\text{at}} Z_{\text{at}'}}{|\mathbf{R}_{\text{at}} - \mathbf{R}_{\text{at}'}|} + \int d^3r n(\mathbf{r}) \sum_{\text{at}} \frac{Z_{\text{at}}}{|\mathbf{r} - \mathbf{R}_{\text{at}}|} = \\ &\sum_{\text{at}} Z_{\text{at}} \left[\sum_{\text{at}' \neq \text{at}} \frac{Z_{\text{at}'}}{|\mathbf{R}_{\text{at}} - \mathbf{R}_{\text{at}'}|} + v_{\text{es}}(\mathbf{R}_{\text{at}}) \right] = \\ &\sum_{\text{at}} Z_{\text{at}} \left[v_{\text{at}}^{\text{es}}(0) + \sum_{\text{at}' \neq \text{at}} v_{\text{at}'}^{\text{es, tot}}(|\mathbf{R}_{\text{at}} - \mathbf{R}_{\text{at}'}|) \right]. \quad (66) \end{aligned}$$

For *neutral* periodic boundary systems, the formalism is identical, except that the long-range electronic potential is additionally split into a real- and reciprocal space part. The electronic potential energy Eq. (64) is reduced to an energy per unit cell by simply restricting the integration volume to the unit cell only. The nuclear contribution per unit cell, Eq. (66), is obtained by evaluating the outer sum over nuclei (\sum_{at}) within the unit cell only.

As mentioned in Sec. 4.3, the use of Ewald's method naturally accounts for *charged* unit cells in the (otherwise divergent) electrostatic *potential*, by leaving away the $\mathbf{G} = 0$ component of the reciprocal-space potential term. For a unit cell charge q , this procedure amounts to the addition of a constant, neutralizing background charge density $n_0 = -q/\Omega$ (Ω is the unit cell volume) to the system. Since Eq. (64) above is handled entirely in real space, this constant density offset does not vanish, but must be accounted for explicitly, most conveniently by evaluating the average electrostatic potential of electrons *and* nuclei. The respective term in Eq. (64) becomes:

$$-\int d^3r n_0 \left[\sum_{\text{at}} v_{\text{at}}^{\text{es, tot}}(|\mathbf{r} - \mathbf{R}_{\text{at}}|) \right] =: q \cdot v_{\text{avg}}^{\text{es}}. \quad (67)$$

4.7 Total-energy derivatives (forces)

For structure optimizations, vibrational properties (via finite differences), or *ab initio* molecular dynamics, we require total energy derivatives (forces)

$$\mathbf{F}_{\text{at}} = -\frac{\partial}{\partial \mathbf{R}_{\text{at}}} E_{\text{tot}}. \quad (68)$$

All needed derivatives follow directly from Eq. (6), with technical adjustments due to Eqs. (64) and (66). We here discuss the derivatives needed for DFT-LDA and -GGA, which we provide in our own implementation, `FHI-aims`[17].

Analytical energy derivatives with atom-centered basis sets have been discussed by several previous authors (e.g., Refs. [154–157]), emphasizing the fact that explicit derivatives beyond simple Hellmann-Feynman forces [158,159] are required. For the specific derivatives arising for atom-centered basis sets in DFT-LDA/GGA, see, e.g., Refs. [101,115,160,161]; we here only summarize the salient expressions in our own notation.

Hellmann-Feynman forces: These are the forces that arise by embedding each nucleus into the electrostatic fields of the electronic charge density and all other nuclei [158,159]:

$$\begin{aligned} \mathbf{F}_{\text{at}}^{\text{HF}} = & \sum_{\substack{\text{at}' \\ \text{at}' \neq \text{at}}} \frac{Z_{\text{at}} Z_{\text{at}'}}{|\mathbf{R}_{\text{at}} - \mathbf{R}_{\text{at}'}|^3} (\mathbf{R}_{\text{at}} - \mathbf{R}_{\text{at}'} \\ & - \int d^3r n(\mathbf{r}) \sum_{\text{at}} \frac{Z_{\text{at}}}{|\mathbf{r} - \mathbf{R}_{\text{at}}|^3} (\mathbf{r} - \mathbf{R}_{\text{at}}). \end{aligned} \quad (69)$$

In periodic systems, electronic and nuclear terms cannot be summed up separately, but only combined as follows:

$$\begin{aligned} \mathbf{F}_{\text{at}}^{\text{HF,conv}} = & Z_{\text{at}} \cdot \\ & \left[\nabla_{\text{at}} v_{\text{at}}^{\text{es}}(0) + \sum_{\substack{\text{at}' \\ \text{at}' \neq \text{at}}} \nabla_{\text{at}} \left(\frac{Z_{\text{at}'}}{|\mathbf{R}'_{\text{at}} - \mathbf{R}_{\text{at}}|} + v_{\text{at}'}^{\text{es}}(|\mathbf{R}'_{\text{at}} - \mathbf{R}_{\text{at}}|) \right) \right]. \end{aligned} \quad (70)$$

Electrostatic multipole derivatives: A first correction term arises because the electrostatic potential expression Eq. (34) is always truncated beyond a given $l_{\text{max}}^{\text{es}}$ and the additional derivative of the double-counting correction Eq. (35) has to be taken into account. The missing multipole terms “move” with \mathbf{R}_{at} , and thus give rise to an extra force term [160]

$$\mathbf{F}_{\text{at}}^{\text{MP}} = - \int d^3r [n(\mathbf{r}) - n_{\text{MP}}(\mathbf{r})] \cdot \nabla_{\text{at}} [v_{\text{at}}^{\text{es}}(|\mathbf{r} - \mathbf{R}_{\text{at}}|)] . \quad (71)$$

Here, n_{MP} is the multipole expansion of the full electron density, Eq. (36), used to define the electrostatic potential. Since the periodic Hellmann-Feynman expression Eq. (70) involves the electronic electrostatic potential, a corresponding multipole error arises also there; this is compensated by redefining

$$\mathbf{F}_{\text{at}}^{\text{MP,conv}} = - \int d^3r [n(\mathbf{r}) - n_{\text{MP}}(\mathbf{r})] \cdot \nabla_{\text{at}} \left[\frac{Z_{\text{at}}}{|\mathbf{r} - \mathbf{R}_{\text{at}}|} + v_{\text{at}}^{\text{es}}(|\mathbf{r} - \mathbf{R}_{\text{at}}|) \right] \quad (72)$$

for the periodic case. For production settings ($l_{\text{max}}^{\text{es}} \geq 4$), $\mathbf{F}_{\text{at}}^{\text{MP}}$ or $\mathbf{F}_{\text{at}}^{\text{MP,conv}}$ are small compared to $\mathbf{F}_{\text{at}}^{\text{HF}}$ or $\mathbf{F}_{\text{at}}^{\text{HF,conv}}$, but never negligible.

Pulay forces: A further class of significant derivatives arises because the basis functions “move” with \mathbf{R}_{at} [154–156], often called Pulay forces for short. These terms vanish for basis sets whose Hilbert space is independent of the atomic positions, e.g., by including basis derivatives outrightly [162,163] in the basis set. However, the goal of a (reasonably) efficient basis set renders this remedy impractical in an all-electron description that relies on atom-centered radial functions. Through $n(\mathbf{r}) = \sum_l f_l |\psi_l(\mathbf{r})|^2$, $\epsilon_l = \langle \psi_l | \hat{h}^{\text{KS}} | \psi_l \rangle$, and up to the local (density-only) parts of exchange-correlation, the respective derivative of Eq. (7) is

$$\mathbf{F}_{\text{at}}^{\text{P}} = - 2 \sum_{lij} f_l \int d^3r c_{li} [\nabla_{\text{at}} \varphi_i(\mathbf{r})] (\hat{h}^{\text{KS}} - \epsilon_l) c_{lj} \varphi_j(\mathbf{r}) . \quad (73)$$

For GGAs, the explicit density gradient derivative in Eq. (29) yields a further term. We formally separate this term out because it requires the explicit Hessian matrix of n , and is thus numerically costly:

$$\begin{aligned} \mathbf{F}_{\text{at}}^{\text{GGA}} = & - 4 \int d^3r \frac{\partial f_{xc}}{\partial |\nabla n|^2} \sum_{lij} \{ f_l c_{li} c_{lj} [\nabla_{\text{at}} \varphi_i(\mathbf{r})] [\nabla \varphi_j(\mathbf{r}) \cdot \nabla n(\mathbf{r})] \\ & + f_l c_{li} c_{lj} \varphi_i(\mathbf{r}) [\nabla_{\text{at}} \nabla \varphi_j(\mathbf{r}) \cdot \nabla n(\mathbf{r})] \} . \end{aligned} \quad (74)$$

In periodic systems, all sums over states l also include \mathbf{k} . Both for $\mathbf{F}_{\text{at}}^{\text{P}}$ and $\mathbf{F}_{\text{at}}^{\text{GGA}}$, any explicit sums over l within the integrations are avoided by first evaluating separately the density matrix Eq. (27) and its equivalent weighted by ϵ_l .

Scalar-relativistic derivatives (atomic ZORA): Since, in general, a scalar-relativistic “atomic ZORA” produces very good geometries, it is convenient to evaluate scalar relativistic energy gradients at this level of theory [Eq. (55) above]. Because $\hat{t}_{\text{at,ZORA}}$ depends on $v_{\text{at}}^{\text{free}}$, it no longer commutes with ∇_{at} [unlike

in the non-relativistic Pulay force expression, Eq. (73)]. The kinetic energy expression thus incurs an explicit derivative. Together with the explicit symmetrization in Eq. (56), this leads to a combined expression for the Pulay forces,

$$\begin{aligned} \mathbf{F}_{\text{at}}^{\text{P,at.ZORA}} = & -2 \sum_{lij} f_l \int d^3r c_{li} c_{lj} \\ & \left\{ \frac{1}{2} [\nabla_{\text{at}} \varphi_i(\mathbf{r})] \hat{t}_{\text{at.ZORA}} \varphi_j(\mathbf{r}) + \frac{1}{2} \varphi_j(\mathbf{r}) \nabla_{\text{at}} [\hat{t}_{\text{at.ZORA}} \varphi_i(\mathbf{r})] \right. \\ & \left. + [\nabla_{\text{at}} \varphi_i(\mathbf{r})] [v(\mathbf{r}) - \epsilon_l] \varphi_j(\mathbf{r}) \right\}, \end{aligned} \quad (75)$$

where the operator $\hat{t}_{\text{at.ZORA}}$ always acts to the right. The important point is that we know the expression $\hat{t}_{\text{at.ZORA}} \varphi_j(\mathbf{r})$ accurately on each atom’s logarithmic grid, from the solution of the defining radial equation Eq. (8). From its splined form, we can easily determine its derivative $\nabla_{\text{at}} [\hat{t}_{\text{at.ZORA}} \varphi_j(\mathbf{r})]$, and thus have all needed pieces to evaluate Eq. (75) as it stands.

In principle, analytical derivatives can also be formulated for the scaled ZORA total energy expression, which is based on the perturbatively corrected eigenvalue expression Eq. (54). However, their implementation is complicated by the fact that higher derivatives of the potential $v(\mathbf{r})$ are needed. Instead, and as already noted above, we are presently developing [147] a scalar-relativistic scheme based on a separation of core and valence states, for which the corresponding relativistic force expression will be revisited.

Atom-centered (“mobile”) integration grids: For completeness, we note that integrations of the form of Eq. (15) formally incur another derivative [101,161]: Since all integration points move with their atomic centers, their integration weights depend on \mathbf{R}_{at} through the partition function $p_{\text{at}}(\mathbf{r})$ [Eq. (16)]. While some workers place great emphasis on this term [161], others [101] find it to be largely irrelevant for reasonably converged integration grids.

In our experience, noticeable grid-dependent derivatives would arise *only* if the simple, non-neutralized total energy expression of Eq. (61) is used, leading for example to inconsistent force and energy landscapes very close to a local structure minimum (typically, residual forces $\gtrsim 5 \cdot 10^{-3}$ eV/Å for light-weight angular grids). In contrast, any inconsistencies at this level do not occur for the neutralized total energy expression of Eqs. (64) and (66). Apparently, the separate Hartree energy of the charged electronic subsystem in the naive expression Eq. (61) increases the missing grid derivative to a noticeable level, compared to the correct neutralized electron-nucleus expression. For practical purposes, we conclude that no formal “grid” derivative along the lines of Refs. [101,161] appears to be required at least for tightly converged structure optimizations.

4.8 Using energy gradients: Structure optimization and Born-Oppenheimer *ab initio* molecular dynamics

Once total energy gradients are available, a plethora of uses becomes available that, in principle, no longer depends on the numerical origin of the forces (from NAOs or otherwise): geometry relaxation, *ab initio* molecular dynamics, harmonic or anharmonic vibrational properties, or nudged-elastic band transition state searches (to list only those uses presently available in our own code, `FHI-aims`[17]). Although technically independent of the use of NAOs, we nevertheless summarize briefly the specific choices made in our implementation of structure relaxations and molecular dynamics.

For structure optimization, we employ the now-standard Broyden-Fletcher-Shanno-Goldfarb (BFGS) algorithm [164]. In brief, the BFGS method starts with a steepest descent from an initial set of atomic positions $\mathbf{R}_{\text{at},0}$, and is iterated from there:

$$\mathbf{R}_{\text{at},i+1} = \mathbf{R}_{\text{at},i} + l_i \mathbf{S}_{\text{at},i} \quad (76)$$

(\mathbf{S}_i is the search direction and l_i a dimensionless linestep). The initial search direction $\mathbf{S}_{\text{at},0}$ is equal to the forces in dimensionless units, and $l_0=1$. The following relaxation steps successively refine an estimate for the local Hessian of the total energy, \hat{H}_i , and use this to update the search direction $\mathbf{S}_{\text{at},i+1}$ by inverting

$$\hat{H}_{i+1} \mathbf{S}_{\text{at},i+1} = \Delta \mathbf{F}_{\text{at},i+1}, \quad (77)$$

where $\Delta \mathbf{F}_{\text{at},i+1} = \mathbf{F}_{\text{at},i} - \mathbf{F}_{\text{at},i+1}$. Initially, \hat{H}_0 is set to unity (corresponding to pure steepest descent), and then updated according to [164]

$$\hat{H}_{i+1} = \hat{H}_i + \frac{\Delta \mathbf{F}_{\text{at},i+1} (\Delta \mathbf{F}_{\text{at},i+1})^T}{\Delta \mathbf{F}_{\text{at},i+1} \cdot (l_i \mathbf{S}_{\text{at},i})} - \frac{(\hat{H}_i \mathbf{S}_{\text{at},i}) (\hat{H}_i \mathbf{S}_{\text{at},i})^T}{\mathbf{S}_{\text{at},i} \hat{H}_i \mathbf{S}_{\text{at},i}}. \quad (78)$$

By default, we keep the line step to update the atomic positions at the estimate for a harmonic total energy surface ($l_i = 1$). We also enforce a predefined maximum coordinate displacement, typically 0.2 Å.

After each line step l_i , we can estimate *a posteriori* the step length l_{opt} that *would* have been optimal, i.e., where the force component along the search direction would have been zero according to a simple linear interpolation:

$$l_{\text{opt},i} = \frac{\mathbf{F}_{\text{at},i-1} \cdot \mathbf{S}_{\text{at},i}}{\mathbf{F}_{\text{at},i-1} \cdot \mathbf{S}_{\text{at},i} - \mathbf{F}_{\text{at},i} \cdot \mathbf{S}_{\text{at},i}} l_i. \quad (79)$$

Depending on this estimate, we adapt our version of BFGS for three special scenarios:

(1) For serious discrepancies between l_i and $l_{\text{opt},i}$, the search step is repeated with $l_{\text{opt},i}$ instead of l_i . The pertinent criterion is $|l_{\text{opt},i}/l_i - 1| < \eta$ (by default, $\eta = 3$).

(2) If the total energy increased by more than a small tolerance (e.g., 10^{-4} eV) in the last relaxation step, we always repeat the line search with $l_{\text{opt},i}$. If the total energy then still increases, we conclude that either the potential-energy surface $E(\mathbf{R}_{\text{at}})$ is not nearly quadratic, or the present Hessian \hat{H} is insufficient to locally represent the energy hypersurface near \mathbf{R}_{at} . In that case, we reset the Hessian matrix to unity and restart with a steepest descent.

(3) If the *a posteriori* estimate of $l_{\text{opt},i}$ is negative, the potential energy surface must have a negative curvature in the search direction, which also means that the Hessian matrix is not positive definite. In order to exit from the negative curvature region with as few relaxation steps as possible, we then reinitialize the Hessian to unity and enforce the maximally allowed relaxation step size in the direction of the forces.

For Born-Oppenheimer *ab initio* molecular dynamics (MD) simulations, we integrate the classical Newton equations of motion for a system of atoms,

$$m_{\text{at}}\ddot{\mathbf{R}}_{\text{at}} = \mathbf{F}(\mathbf{R}_{\text{at}}) \quad . \quad (80)$$

To solve for the molecular trajectories, we use by default the velocity-Verlet algorithm [165]. Our current implementation allows for MD in the *NVE* and *NVT* ensembles. For constant temperature simulations, we provide a Berendsen thermostat [166], and (for the canonical *NVT*-ensemble) a simple, stochastic Andersen thermostat [167], and a Nosé-Hoover thermostat [168,169]. Work to implement further thermostats and constant pressure simulations is ongoing as well. Finally, a fourth-order symplectic integrator (Ref. [170], referred to as “SI4” in Ref. [171]) is available for very accurate *NVE* trajectories. Regardless of the integrator used, we note that rather tight s.c.f. convergence criteria must be applied in Born-Oppenheimer molecular dynamics to avoid slow energy drifts in *NVE*.

Both for structure optimization and for molecular dynamics, we routinely remove any spurious residual translations and rotations of the system as a whole for cluster geometries (or pure translations for periodic systems). In principle, no such rotations or translations should arise, but very small deviations may occur due to the finite accuracy of the integration grids in the numerical calculation of energies and forces. While usually a small effect, these components could introduce unwanted slow rotations in long molecular dynamics runs. In practice, we therefore formally project out such residual components from the forces, using the Eckart-Sayvetz conditions [172,173]. In molecular dynamics, these conditions may also be enforced for the velocities themselves, in order to correct for the effects of a finite time step.

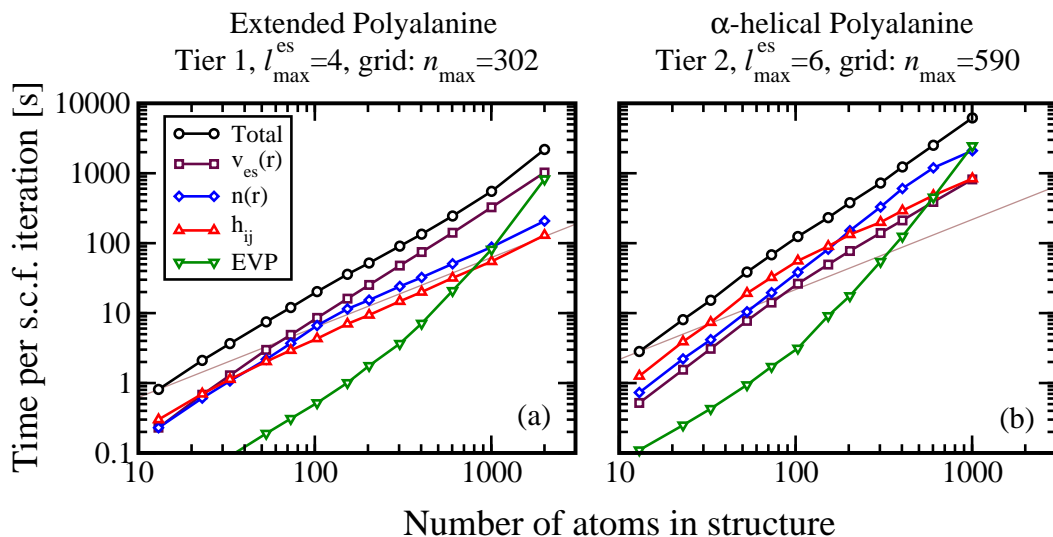


Fig. 9. Wall clock time per s.c.f. iteration with system size for two series of polyalanine conformers Ala_n in DFT-PBE, on 16 IBM p5+ CPU cores. Subtimings are shown for the individual parts of each s.c.f. iteration: Electrostatic potential summation $v_{\text{es}}(\mathbf{r})$, electron density update $n(\mathbf{r})$, h_{ij} matrix element integration, and solution of the eigenvalue problem (EVP) using ScaLapack. Thin brown line: slope for linear scaling. (a) Fully extended conformers ($n=1-200$), using settings appropriate for accurate geometry pre-relaxations: *tier 1* basis set, $l_{\max}^{\text{es}}=4$, $n_{\max}=302$ for Lebedev grids on radial integration shells. (b) α -helical conformers ($n=1-100$), using settings appropriate for tightly converged energy differences: *tier 2* basis set, $l_{\max}^{\text{es}}=6$, $n_{\max}=590$ for Lebedev grids on radial integration shells.

5 Performance in practice

In this section, we quantify the performance of the algorithms presented here (scaling with size of the investigated system, and with number of CPU cores on parallel architectures), using our own code `FHI-aims` on modern computer hardware, and for two classes of physical systems: (i) a series of fully extended vs. α -helical polyalanine molecules Ala_n , characteristic of large light-element molecules in cluster-type geometries ($n=1-200$, i.e., 13-2,003 atoms; see Figs. 5a and b for the respective Ala_{20} conformers); and (ii) a series of $\text{Au}(100)-(5 \times n)$ surface slabs, characteristic of large-scale, heavy-element periodic systems ($n=1-20$, i.e., 26-526 atoms per unit cell). The computers employed range from commodity PC cluster architectures up to special, massively parallel architectures (IBM BlueGene/P).

5.0.1 Polyalanine conformers Ala_n ($n=1-200$)

We first turn to the light-element series of molecules, (i). Figure 9 shows the wall clock time required as a function of system size (number of atoms), for a single s.c.f. iteration and its parts, executed in parallel on 16 CPU cores of an

Operation	a_1 (extended Ala $_n$)	a_1 (α helical Ala $_n$)
Total s.c.f. iteration	1.9	1.8
Electrostatic potential $v_{\text{es}}(\mathbf{r})$	1.7	1.5
Electron density $n(\mathbf{r})$	1.2	1.1
h_{ij} matrix elements	1.2	1.1
Eigenvalue problem	3.3	3.3

Table 2

Fitted CPU time scaling exponents a_1 for the largest polyalanine conformers Ala $_n$ included in Figs. 9a and b, respectively, assuming $t = a_0 N^{a_1}$.

IBM Power5+ computer (1.9 GHz clock frequency). Approximately 15 s.c.f. iterations are required for tight force convergence ($\lesssim 10^{-4}$ eV/Å) in each case. Figure 9a refers to a series of fully extended conformers of Ala $_n$ (Fig. 5a). Due to their spatial extent, these structures should benefit most from exploiting the localisation of basis functions on a real-space grid, particularly the electron density update and the integration of matrix elements h_{ij} . The computational settings are characteristic of accurate geometry pre-relaxations, with small *tier* 1 basis sets for all elements (confining potential: $r_{\text{onset}}=3.5$ Å, $w=2.0$ Å, respectively), $l_{\text{es}}^{\text{max}}=4$ for the electrostatic potential $v_{\text{es}}(\mathbf{r})$, and integration grids totalling $\approx 10,800$ points per atom (with a maximum of 302 angular integration points per radial integration shell). Scaling exponents a_1 were fitted to the three largest conformers shown, assuming a separate polynomial scaling law $t = a_0 N^{a_1}$ for each operation of the s.c.f. iteration; these exponents are listed in Table 2.

With respect to the scaling exponents, it is gratifying to see that those operations which should in principle exhibit strict $O(N)$ scaling, namely the updates of the electron density $n(\mathbf{r})$ and the Hamilton matrix elements h_{ij} , come close to this expected behaviour already for rather manageable molecule sizes ($a_1=1.2$ for $\gtrsim 100$ atoms). The crossover where a density matrix based update of $n(r)$ becomes favorable according to Eq. (28) occurs just above 100 atoms for this structure and basis choice. With these two operations out of the picture, the scaling exponent $a_1=1.7$ for the electrostatic potential update, Eq. (34), dominates the available system sizes; its low-angular momentum components simply fall off too slowly with system size. Beyond system sizes of the order of $\approx 1,000$ atoms, the $\sim O(N^3)$ scaling of the ScaLapack eigenvalue solver takes over as expected.

Figure 9b contrasts these findings with a series of more compact α -helical conformers. At the same time, the computational settings are now chosen to reflect tight convergence of conformational energy differences (e.g., ≈ 0.12 meV/atom for the energy difference discussed in Fig. 5): *tier* 2 basis sets, $l_{\text{es}}^{\text{max}}=6$ for the electrostatic potential $v_{\text{es}}(\mathbf{r})$, and dense integration grids totalling $\approx 16,400$

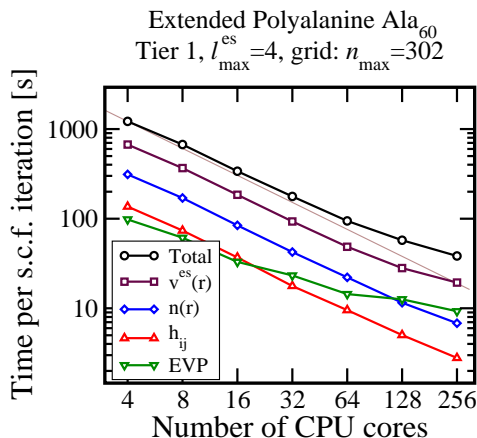


Fig. 10. Parallel scalability of FHI-aims for a 603 atom fully extended molecular conformer Ala₆₀ on a current Linux cluster architecture (see Fig. 9a for computational parameters and timing details). Each CPU core is a 64bit AMD Opteron with 2.6 GHz clock frequency; nodes of four cores each are interconnected with Infiniband. The thin brown line corresponds to ideal scaling.

points per atom. The immediate difference between Fig. 9b and Fig. 9a is that the cross-over to linear scaling in both the density update and the h_{ij} update happens significantly later for the more compact helices. For the example of the density update, the density matrix based update formula becomes favorable only between 400 and 600 atoms in the structure. Nonetheless, the large-scale scaling exponents a_1 for all individual operations in Table 2 are quite similar for both sets of conformers, leading again to an approximate crossover of the ScaLapack $O(N^3)$ scaling at $\approx 1,000$ atoms.

In addition to good scaling of CPU time requirements with system size, it is very important (particularly for applications such as molecular dynamics) to be able to reduce the wall clock time requirements for individual s.c.f. cycles efficiently by way of (massive) parallelization. For the 603-atom, extended Ala₆₀ conformer included in Fig. 9 (same computational settings), we demonstrate the efficiency of this parallel scaling on a PC-based cluster architecture type that is widely available today. The computer system consists of Infiniband-connected compute nodes with four 2.6 GHz AMD Opteron CPU cores each, i.e., makes heavy use of internode (rather than shared memory) communication. Nonetheless, the parallel scalability for this system (shown in Fig. 10) is quite satisfactory: Nearly no overall performance loss occurs up to 64 CPU cores. The grid based operations [$n(\mathbf{r})$ and h_{ij} integrals] scale with no loss also beyond. The scalability limit of ScaLapack is here approached at ≈ 128 CPU cores (with 5724×5724 , the size of the matrices h_{ij} and s_{ij} is still small). In addition, some communication limitations due to the distribution of the real-space multipole components of the Hartree potential across different cores becomes noticeable at ≈ 256 CPU cores.

Regarding the thus identified scaling limits of the current implementation, we

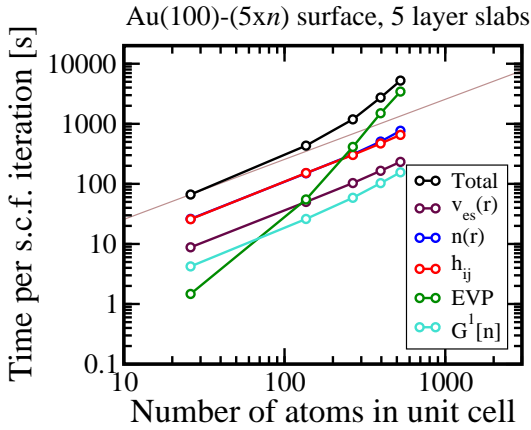


Fig. 11. CPU time scaling with system size for a series of large Au(100)-($5 \times n$) surface unit cells ($n=1,5,10,15,20$) in DFT-PBE, for five layer slabs: *spdf* section of the *tier 1* basis set, $l_{\max}^{\text{es}}=6$, and $\approx 14,000$ points per atom (max. $n_{\max}=590$ angular points per radial shell). The calculations were performed on 32 IBM Power5+ CPU cores (1.9 GHz clock speed). Subtimings are reported for the electrostatic potential $v_{\text{es}}(\mathbf{r})$, the electron density $n(\mathbf{r})$, the integration of h_{ij} , and the ScaLapack eigenvalue solver (EVP). In addition, we use a Kerker preconditioner G^1 in the density mixing stage.

note the following areas to be tackled in the future: (1) In principle, near-linear scaling solutions for a long-range real-space electrostatic potential with free boundary conditions are available in the context of electronic structure theory (e.g., Refs. [174–177]), and such an implementation will alleviate the main scaling bottleneck in Fig. 9a. (2) For very large systems, the scaling limits of ScaLapack can be overcome by an iterative eigenvalue solver strategy [108–113,107], as mentioned above. (3) Finally, the use of difference techniques between successive s.c.f iterations should greatly reduce the overall computational effort expended on the grid based operations [$n(\mathbf{r})$, h_{ij} , and $v_{\text{es}}(\mathbf{r})$] for any of the molecules shown in Fig. 9.

5.0.2 Au(100)-($5 \times n$) reconstructed surface slabs

As a characteristic example for periodic solid-state calculations involving heavy elements, we consider the Au(100) large-scale surface reconstruction. This surface forms a quasi-hexagonal layer on top of a quadratic substrate,[178–181] leading to a large coincidence period which can be approximated as $(5 \times n)$, where $n \geq 20$. Figure 11 shows the CPU time scalings for this system class: the pieces of a s.c.f. iteration on 32 IBM Power5+ CPU cores (1.9 GHz clock speed), for DFT-PBE and production-quality numerical parameters: $l_{\max}^{\text{es}}=6$, $\approx 14,000$ integration points per Au atom, and the *spdf* section of the *tier 1* basis set (shown to be a sufficient basis choice in Sect. 3.5) with converged settings for the confining potential ($r_{\text{onset}}=4.0 \text{ \AA}$, $w=2.0 \text{ \AA}$, yielding an outer radius $r_{\text{cutoff}}=6.0 \text{ \AA}$). Two \mathbf{k} -points were considered for each slab. The latter

setting is appropriate from $n \geq 10$ onwards, and is kept here for consistency also for the smaller unit cells [(5×1) and (5×5)].

The important point is that, for periodic systems, *all* operations on the real-space grid [$n(\mathbf{r})$, h_{ij} , and $v_{\text{es}}(\mathbf{r})$] scale as $O(N)$ from the outset. This is because already the smallest geometry takes into account the full overlap of basis functions also from neighbouring unit cells. In addition, the long-range part of the Hartree potential is treated by Ewald’s method and using Fourier transform, which eliminates the less favorable scaling for this part that was observed for molecules. Due to the all-electron nature of the problem (79 electrons per Au atom), the ScaLapack based eigensolver here becomes dominant already for ≈ 400 atoms; in the near future, a separation of core and valence states together with an iterative solver strategy for the latter should alleviate this part.

The time required for the smaller cells in Fig. 11 remains manageable (≈ 30 - 40 s.c.f. iterations are required to converged forces to 10^{-4} eV/Å), although even the (5×1) unit cell is already large by conventional standards. However, significant CPU resources are clearly required in order to study very large surface slabs, such as (5×20) (526 atoms) or larger. By exploiting the good parallel scaling properties of the algorithms presented here, even such large problems can still be addressed successfully on modern, dedicated massively parallel computer architectures. For example, we have ported FHI-aims to IBM’s BlueGene/P architecture. This architecture unites thousands of CPU cores (in the case of the BlueGene/P system “genius” at the Garching compute center of the Max-Planck-Society, currently 16,384 CPU cores) that are individually relatively slow (850 MHz PowerPC 450 processors), and connects them with a sophisticated communication interface. Since furthermore each individual MPI task must be limited to less than 512 MB working memory, the efficient use of this modern architecture is only enabled by the full memory parallelization of all important operations. A description of the porting effort and scaling properties will be given elsewhere [106].

6 Conclusion

In the present work, we summarize a complete set of algorithms that enables first-principles electronic structure theory (here based on DFT-LDA and -GGA) using numeric atom-centered orbitals as basis functions. These numerical algorithms and basis functions form the base of our all-electron, full-potential electronic structure code FHI-aims, intended to be a versatile application and development tool for all aspects of current electronic structure theory. While many of the individual algorithms have been developed by a community of researchers over the years, our observations made in the

context of their implementation should be of use not just to users of a single code (**FHI-aims**), but to the broader community of practitioners that employ NAOs for other purposes as well. Our examples demonstrate the capabilities and reliability of the algorithms presented here for a wide range of tasks, including system sizes up to thousands of atoms, and good performance from PC-like single-CPU up to massively parallel computational architectures with thousands of CPU cores without compromises on accuracy. Further examples with respect to new developments in electronic structure theory itself will be given separately [6], concerning particularly our implementation of the two-electron Coulomb operator needed for, e.g., hybrid functionals, Hartree-Fock, MP2 or *GW* many-body perturbation theories. We have also outlined focal points for what we believe are the most important future development directions, namely, an FP-LAPW like implementation of scalar relativity [147], or an iterative eigenvalue solver strategy to enable improved scaling towards significantly more CPUs and even larger system sizes [114].

We believe that the sustained and continued development of the algorithms presented here, whether embodied in **FHI-aims** or elsewhere, represents a significant opportunity to leverage a powerful, efficient yet accurate framework for future developments of any aspect of electronic structure theory, and this is where our efforts will be focused.

Acknowledgments

Our development efforts have benefitted from numerous fruitful discussions with and contributions from many fellow researchers. Regarding the algorithms presented here, we especially wish to acknowledge Dr. Rainer Johanni (Rechenzentrum Garching of the Max-Planck-Society, with Dr. Hermann Lederer) for significant parallel scaling contributions (ScaLapack and BlueGene architecture), Mariana Rossi (FHI Berlin) for adding several LDA, GGA, and hybrid XC functionals; Dr. Alexandre Tkatchenko (FHI Berlin) for continued benchmarking and testing efforts, and implementing a C_6/R^6 van der Waals correction for DFT [62]; and Dr. Patrick Rinke (FHI Berlin / UCSB, Santa Barbara, CA) for continued support and ideas towards all many-body formalisms building on the resolution of the identity of the two-electron Coulomb operator. These contributions will also be the subject of individual publications.

We further acknowledge discussions with Dr. Jörg Behler (FHI, now Ruhr-Universität, Bochum) during the early development stages of **FHI-aims**; Dr. Christoph Freysoldt (FHI, now Max-Planck-Institut für Eisenforschung Düsseldorf) regarding algorithms for electronic structure self-consistency; and Dr. Lars Ismer (*ibid.*) regarding algorithms for structure relaxation.

We are also indebted to Dr. Bernard Delley (Paul-Scherrer-Institut, Bern) for use of his code version of Lebedev-style angular integration grids; Prof. Gustavo Scuseria (Rice University, Houston, TX) for providing us with the original version of the Heyd-Scuseria-Ernzerhof (HSE) long-range exchange gradient functional; and Andreas Dolfen (Forschungszentrum Jülich) for providing useful implementation parts regarding 3D plot output functionality and electrostatic moment calculations.

Ville Havu acknowledges the use of the “Murska” cluster of the Finnish IT center for sciences, CSC.

References

- [1] P. Hohenberg, W. Kohn, *Phys. Rev. B* 136 (1964) 864.
- [2] W. Kohn, L. Sham, *Phys. Rev.* 140 (1965) A1133.
- [3] L. Hedin, *Phys. Rev.* 139 (1965) A796.
- [4] G. Onida, L. Reining, A. Rubio, *Rev. Mod. Phys.* 74 (2002) 601.
- [5] P. Rinke, A. Qteish, J. Neugebauer, C. Freysoldt, M. Scheffler, *New J. Phys.* 7 (2005) 126.
- [6] X. Ren, A. Sanfilippo, V. Blum, P. Rinke, K. Reuter, M. Scheffler, to be published.
- [7] F. Averill, D. Ellis, *J. Chem. Phys.* 59 (1973) 6412.
- [8] A. Zunger, A. Freeman, *Phys. Rev. B* 15 (1977) 4716.
- [9] B. Delley, D. Ellis, *J. Chem. Phys.* 76 (1982) 1949.
- [10] B. Delley, *J. Chem. Phys.* 92 (1990) 508.
- [11] G. te Velde, F. Bickelhaupt, E. Baerends, C. F. Guerra, S. van Gisbergen, J. Snijders, T. Ziegler, *J. Comput. Chem.* 22 (2001) 931.
- [12] K. Koepernik, H. Eschrig, *Phys. Rev. B* 59 (1999) 1743.
- [13] A. Horsfield, *Phys. Rev. B* 56 (1991) 6594.
- [14] O. Sankey, D. Niklewski, *Phys. Rev. B* 40 (1989) 3979.
- [15] J. Soler, E. Artacho, J. Gale, A. García, J. Junquera, P. Ordejón, D. Sánchez-Portal, *J. Phys.: Condens. Matter* 14 (2002) 2745.
- [16] T. Ozaki, H. Kino, J. Yu, M. Han, N. Kobayashi, M. Ohfuti, F. Ishii, T. Ohwaki, User’s manual of OpenMX, <http://www.openmx-square.org>, 2008.

- [17] V. Blum, M. Scheffler, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter, M. Scheffler, The Fritz Haber Institute *ab initio* molecular simulations package (FHI-aims), 2009, <http://www.fhi-berlin.mpg.de/aims>.
- [18] J. Talman, J. Chem. Phys. 80 (1984) 2000.
- [19] J. Talman, J. Chem. Phys. 84 (1986) 6879.
- [20] J. Talman, Int. J. Quantum Chem. 93 (2003) 72.
- [21] J. Talman, Int. J. Quantum Chem. 95 (2003) 442.
- [22] J. Talman, Collect. Czech Chem. Commun. 70 (2005) 1035.
- [23] J. Talman, Int. J. Quantum Chem. 107 (2007) 1578.
- [24] J. M. Perez-Jorda, W. Yang, Chemical Physics Letters 241 (1995) 469.
- [25] R. Stratmann, G. Scuseria, M. Frisch, Chem. Phys. Lett. 257 (1996) 213.
- [26] C. Fonseca Guerra, J. Snijders, G. te Velde, E. Baerends, Theor. Chem. Acc. 99 (1998) 391.
- [27] A. Becke, J. Chem. Phys. 98 (1993) 1372.
- [28] P. Stephens, F. Devlin, C. Chabalowski, M. Frisch, J. Phys. Chem. 98 (1994) 11623.
- [29] J. Perdew, M. Ernzerhof, K. Burke, J. Chem. Phys. 105 (1996) 9982.
- [30] C. Adamo, V. Barone, J. Chem. Phys. 110 (1999) 6158.
- [31] J. Heyd, G. E. Scuseria, M. Ernzerhof, J. Chem. Phys. 118 (2003) 8207.
- [32] J. Heyd, G. E. Scuseria, M. Ernzerhof, J. Chem. Phys. 124 (2006) 219906.
- [33] A. V. Krukau, O. A. Vydrov, A. F. Izmaylov, G. E. Scuseria, J. Chem. Phys. 125 (2006) 224106.
- [34] A. Szabo, N. Ostlund, Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory, Dover, 1996.
- [35] D. Langreth, J. Perdew, Phys. Rev. B 15 (1977) 2884.
- [36] O. Gunnarson, B. Lundqvist, Phys. Rev. B 13 (1976) 4274.
- [37] X. Ren, P. Rinke, M. Scheffler, Phys. Rev. B, submitted.
- [38] S. Boys, I. Shavitt, University of Wisconsin Rept. WIS-AF-13.
- [39] B. Dunlap, J. Connolly, J. Sabin, J. Chem. Phys. 71 (1979) 3396.
- [40] J. W. Mintmire, J. R. Sabin, S. B. Trickey, Phys. Rev. B 26 (1982) 1743.
- [41] C. V. Alsenoy, J. Comput. Chem. 9 (1988) 620.
- [42] O. Vahtras, J. Almlöf, M. Feyereisen, Chem. Phys. Lett. 213 (1993) 514.

- [43] K. Eichkorn, O. Treutler, H. Öhm, M. Häser, R. Ahlrichs, *Chem. Phys. Lett.* 240 (1995) 283.
- [44] D. Hartree, *Proc. Camb. Phil. Soc.* 24 (1928) 111.
- [45] D. Hartree, W. Hartree, *Proc. Roy. Soc. London* 150 (1935) 9.
- [46] G. Pratt, *Phys. Rev.* 88 (1952) 1217.
- [47] F. Herman, S. Skillman, *Atomic Structure Calculations*, Prentice-Hall, Englewood Cliffs, NJ, 1963.
- [48] N. Mermin, *Phys. Rev.* 137 (1965) A1441.
- [49] C. Roothaan, *Rev. Mod. Phys.* 23 (1951) 69.
- [50] C. Hall, *Proc. Roy. Soc. London* A205 (1951) 541.
- [51] A. Seidl, A. Görling, P. Vogl, J. Majewski, M. Levy, *Phys. Rev. B* 53 (1996) 3764.
- [52] S. Vosko, L. Wilk, M. Nusair, *Can. J. Phys.* 58 (1980) 1200.
- [53] J. Perdew, A. Zunger, *Phys. Rev. B* 23 (1981) 5048.
- [54] J. Perdew, Y. Wang, *Phys. Rev. B* 45 (1992) 13244.
- [55] G. Scuseria, V. Staroverov, Progress in the development of exchange-correlation functionals, in: C. Dykstra, G. Frenking, K. Kim, G. Scuseria (Eds.), *Theory and Applications of Computational Chemistry: The First 40 Years*, Elsevier, Amsterdam, 2005, Ch. 24, the parameters for "VWN"-LDA as implemented in the Gaussian code are given in Table 1.
- [56] A. Becke, *J. Chem. Phys.* 88 (1988) 1053.
- [57] C. Lee, W. Yang, R. Parr, *Phys. Rev. B* 37 (1988) 785.
- [58] J. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* 77 (1996) 3865.
- [59] Y. Zhang, W. Yang, *Phys. Rev. Lett.* 80 (1998) 890.
- [60] B. Hammer, L. Hansen, J. Nørskov, *Phys. Rev. B* 59 (1999) 7413.
- [61] J. Perdew, A. Ruzsinszky, G. Csonka, O. Vydrov, G. Scuseria, L. Constantin, X. Zhou, K. Burke, *Phys. Rev. Lett.* 100 (2008) 136406.
- [62] A. Tkatchenko, M. Scheffler, *Phys. Rev. Lett.* 102 (2009) 073005.
- [63] J. Behler, B. Delley, K. Reuter, M. Scheffler, *Phys. Rev. B* 75 (2007) 115409.
- [64] H. Jónsson, G. Mills, K. W. Jacobsen, *Nudged Elastic Band Method for Finding Minimum Energy Paths of Transitions*, World Scientific, 1998, pp. 385–404.
- [65] G. Henkelman, B. P. Uberuaga, H. Jónsson, *J. Chem. Phys.* 113 (2000) 9901.

- [66] G. Henkelman, H. Jónsson, *J. Chem. Phys.* 113 (2000) 9978.
- [67] M.-P. Gaigeot, M. Martinez, R. Vuilleumier, *Mol. Phys.* 105 (2007) 2857.
- [68] M. Fuchs, M. Scheffler, *Comp. Phys. Comm.* 119 (1999) 67.
- [69] G. B. Arfken, H. J. Weber, *Mathematical Methods for Physicists*, 5th Edition, Harcourt/Academic Press, 2001.
- [70] H. Eschrig, I. Bergert, *phys. stat. sol. (b)* 90 (1978) 621.
- [71] H. Eschrig, *Optimized LCAO method and the electronic structure of extended systems*, Akademie Verlag and Springer, Berlin, 1988.
- [72] B. Delley, *J. Chem. Phys.* 113 (2000) 7756.
- [73] J. Junquera, O. Paz, D. Sanchez-Portal, E. Artacho, *Phys. Rev. B* 64 (2001) 235111.
- [74] T. Ozaki, H. Kino, *Phys. Rev. B* 69 (2004) 195113.
- [75] V. Blum, M. Scheffler, in preparation.
- [76] E. Artacho, D. Sánchez-Portal, P. Ordejón, A. García, J. Soler, *phys. stat. sol. b* 215 (1999) 809.
- [77] J. Slater, Atomic shielding constants, *Phys. Rev.* 36 (1930) 57.
- [78] A. Wilson, T. van Mourik, J. T.H. Dunning, *J. Mol. Struct. (Theochem)* 388 (1996) 339.
- [79] M. Frisch, J. Pople, J. Binkley, *J. Chem. Phys.* 80 (1984) 3265.
- [80] F. Weigend, R. Ahlrichs, *Phys. Chem. Chem. Phys.* 7 (2005) 3297.
- [81] S. Kenny, A. Horsfield, H. Fujitani, *Phys. Rev. B* 62 (2000) 4899.
- [82] E. Clementi, *J. Chem. Phys.* 46 (1967) 3851.
- [83] H. Jansen, P. Ros, *Chem. Phys. Lett.* 3 (1969) 140.
- [84] S. Boys, F. Bernardi, *Mol. Phys.* 19.
- [85] F. Kutzler, G. Painter, *Phys. Rev. Lett.* 59 (1987) 1285.
- [86] B. Santra, A. Michaelides, M. Scheffler, *J. Chem. Phys.* 127 (2007) 184104.
- [87] J. Ireta, J. Neugebauer, M. Scheffler, A. Rojo, M. Galván, *J. Phys. Chem. B* 107 (2003) 1432.
- [88] F. Murnaghan, *Proc. Nat. Acad. Sci.* 30 (1947) 244.
- [89] S. Kurth, J. Perdew, P. Blaha, *Int. J. Quantum Chem.* 75 (1999) 889.
- [90] P. Havu, V. Blum, P. Rinke, M. Scheffler, in preparation.
- [91] A. Becke, *J. Chem. Phys.* 88 (1988) 2547.

- [92] G. te Velde, E. Baerends, *J. Comput. Phys.* 99 (1992) 84.
- [93] F. Hirshfeld, *Theor. Chim. Acta (Berl.)* 44 (1977) 129.
- [94] A. McLaren, *Math. Comput.* 17 (1963) 361.
- [95] A. Stroud, *Approximate Calculation of Multiple Integrals*, Prentice-Hall, Englewood Cliffs, 1971.
- [96] V. Lebedev, *Zh. Vychisl. Mat. mat. Fiz.* 15 (1975) 48.
- [97] V. Lebedev, *Zh. Vychisl. Mat. mat. Fiz.* 16 (1976) 293.
- [98] V. Lebedev, D. Laikov, *Doklady Mathematics* 59 (1999) 477.
- [99] B. Delley, *J. Comp Chem.* 17 (1995) 1152.
- [100] O. Treutler, R. Ahlrichs, *J. Chem. Phys.* 102 (1995) 346.
- [101] J. Baker, J. Andzelm, A. Scheiner, B. Delley, *J. Chem. Phys.* 101 (1994) 8894.
- [102] V. Havu, V. Blum, P. Havu, M. Scheffler, in preparation.
- [103] W. Kahan, Separating clouds by a plane, lecture notes, CS Division, UC Berkeley.
- [104] E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. D. Croz, A. Greenbaum, S. Hammarling, A. McKenney, D. Sorensen, *Lapack Users' Guide*, 3rd Edition, SIAM, Philadelphia, PA, 1999, <http://www.netlib.org/lapack>.
- [105] L. Blackford, J. Choi, A. Cleary, E. D'Azevedo, J. Demmel, I. Dhillon, J. Dongarra, S. Hammarling, G. Henry, A. Petitet, K. Stanley, D. Walker, R. Whaley, *ScaLAPACK Users' Guide*, SIAM, Philadelphia, PA, 1997, <http://www.netlib.org/scalapack>.
- [106] R. Johanni, H. Lederer, V. Blum, P. Havu, V. Havu, M. Scheffler, to be published.
- [107] M. Rayson, P. Briddon, *Comput. Phys. Comm.* 178 (2008) 128.
- [108] D. Wood, A. Zunger, *J. Phys. A: Math. Gen.* 18 (1985) 1343.
- [109] Y. Saad, *Numerical Methods for Large Eigenvalue Problems*, Manchester University Press, 1992.
- [110] M. Payne, M. Teter, D. Allan, T. Arias, J. Joannopoulos, *Rev. Mod. Phys.* 64 (1992) 1045.
- [111] G. Kresse, J. Furthmüller, *Phys. Rev. B* 54 (1996) 11169.
- [112] C. Gan, P. Haynes, M. Payne, *Comp. Phys. Comm.* 134 (2001) 33.
- [113] Y. Zhou, M. Tiago, Y. Saad, J. Chelikowski, *J. Comput. Phys.* 219 (2006) 172.
- [114] V. Havu, P. Havu, V. Blum, M. Scheffler, in preparation.

- [115] J. Pople, P. Gill, B. Johnson, Chem. Phys. Lett. 199 (1992) 557.
- [116] B. Delley, J. Chem. Phys. 100 (1996) 6107.
- [117] M. Abramowitz, I. A. Stegun (Eds.), Handbook of Mathematical Functions, Dover Publications, Inc. New York, 1970.
- [118] P. Blaha, K. Schwarz, G. Madsen, D. Kvasnicka, J. Luitz, Wien2k, an Augmented Plane Wave Plus Local Orbitals Program for Calculating Crystal Properties, Ed. Wien2k_03, TU Wien, 2003.
- [119] U. Birkenheuer (1994).
- [120] J. C. Boettger, Int. J. Quantum Chem. S29.
- [121] S. Trickey, J. Alford, J. Boettger, Methods and implementation of robust, high-precision gaussian basis dft calculations for periodic systems: the gtoff code, in: J. Leszczynski (Ed.), Computational Materials Science, Vol. 102, Elsevier, Amsterdam, 2004, p. 171.
- [122] V. Saunders, C. Freyria-Fava, R. Dovesi, L. Salasco, C. Roetti, Mol. Phys. 77 (1992) 629.
- [123] M. Manninen, R. Nieminen, P. Hautojärvi, Phys. Rev. B 12 (1975) 4012.
- [124] R. Nieminen, J. Phys. F 7 (1977) 375.
- [125] G. Kerker, Phys. Rev. B 23 (1981) 3082.
- [126] G. Kresse, J. Furthmüller, Comp. Mat. Sci. 6 (1996) 15.
- [127] P. Pulay, Chem. Phys. Lett. 73 (1980) 393.
- [128] C.-L. Fu, K.-M. Ho, Phys. Rev. B 28 (1983) 5480.
- [129] M. Methfessel, A. Paxton, Phys. Rev. B 40 (1989) 3616.
- [130] M. Gillan, J. Phys.:Condens. Matter 1 (1989) 689.
- [131] F. Wagner, T. Laloyaux, M. Scheffler, Phys. Rev. B 57 (1998) 2102.
- [132] H. Wendel, R. Martin, Phys. Rev. B 19 (1979) 5251.
- [133] J. Harris, Phys. Rev. B 31 (1985) 1770.
- [134] W. Foulkes, R. Haydock, Phys. Rev. B 39 (1989) 12520.
- [135] A. Read, R. Needs, J. Phys.: Condens. Matter 1 (1989) 7565.
- [136] P. Pyykkö, Chem. Rev. 88 (1988) 563.
- [137] D. Koelling, B. Harmon, J. Phys. C 10 (1977) 3107.
- [138] D. Singh, L. Nordstrom, Planewaves, pseudopotentials, and the LAPW method, 2nd Edition, Springer, 2006.
- [139] V. Eyert, The Augmented Spherical Wave Method, Springer, 2007.

- [140] H. Skriver, *The LMTO method*, Springer, 1984.
- [141] M. Asato, A. Settels, T. Hosjino, T. Asada, S. Blügel, R. Zeller, P. Dederichs, *Phys. Rev. B* 60 (1999) 5202, and references therein.
- [142] E. van Lenthe, E. Baerends, J. Snijders, *J. Chem. Phys.* 101 (1994) 9783.
- [143] C. Chang, M. Péliissier, M. Durand, *Phys. Scr.* 34 (1986) 394.
- [144] J.-L. Heully, I. Lindgren, E. Lindroth, S. Lundquist, A.-M. Martensson-Pendrill, *J. Phys. B* 19 (1986) 2799.
- [145] E. van Lenthe, E. Baerends, J. Snijders, *J. Chem. Phys.* 99 (1993) 4597.
- [146] L. Foldy, S. Wouthuysen, *Phys. Rev.* 78 (1950) 29.
- [147] V. Blum, P. Havu, M. Scheffler, in preparation.
- [148] C. van Wüllen, *J. Chem. Phys.* 109 (1999) 392.
- [149] J. van Lenthe, S. Faas, J. Snijders, *Chem. Phys. Lett.* 328 (2000) 107.
- [150] Computational parameters used in the Wien2k [118] calculations for Au₂: Supercell size 20×20×25 bohr³, 5s and 4f states included as semicore states, $R_{MT}=2.0$ bohr, $l_{\max}=12$ for the wave function expansion and $l_{\text{nsmax}}=6$ for the potential expansion, plane-wave cutoff $k_{\max}^2=36$ Ry, potential cutoff $G_{\max}^2=400$ Ry, and all computations performed at the Γ point.
- [151] J. Janak, *Phys. Rev. B* 10 (1974) 3985.
- [152] A. Williams, J. Kübler, C. Gelatt, *Phys. Rev. B* 19 (1979) 6094.
- [153] M. Weinert, E. Wimmer, A. Freeman, *Phys. Rev. B* 26 (1982) 4571.
- [154] J. Gerratt, I. Mills, *J. Chem. Phys.* 49 (1968) 1719.
- [155] J. Gerratt, I. Mills, *J. Chem. Phys.* 49 (1968) 1730.
- [156] P. Pulay, *Mol. Phys.* 17 (1969) 197.
- [157] M. Scheffler, J. Vigneron, G. Bachelet, *Phys. Rev. Lett.* 49 (1982) 1765.
- [158] H. Hellmann, *Z. Phys.* 85 (1933) 180.
- [159] R. Feynman, *Phys. Rev.* 56 (1939) 340.
- [160] B. Delley, *J. Chem. Phys.* 94 (1991) 7245.
- [161] B. Johnson, P. Gill, J. Pople, *J. Chem. Phys.* 98 (1993) 5612.
- [162] H. Nakatsui, K. Kanda, T. Yonezawa, *Chem. Phys. Lett.* 75 (1980) 340.
- [163] H. Nakatsui, T. Hayakawa, M. Hada, *Chem. Phys. Lett.* 80 (1981) 94.
- [164] W. K. Press, S. A. Teukolsky, W. T. Vetterlin, B. T. Flannery, *Numerical Recipes*, 3rd Edition, Cambridge University Press, 2007.

- [165] D. Frenkel, B. Smit, 2nd Edition, Academic Press, 2002.
- [166] H. Berendsen, J. Postma, W. van Gunsteren, A. DiNola, J. Haak, *J. Chem. Phys.* 81 (1984) 3684.
- [167] H. Andersen, *J. Chem. Phys.* 72 (1980) 2384.
- [168] S. Nosé, *J. Chem. Phys.* 81 (1984) 511.
- [169] W. Hoover, *Phys. Rev. A* 31 (1985) 1695.
- [170] H. Yoshida, *Phys. Lett. A* 150 (1990) 262.
- [171] H. Ishida, Y. Nagai, A. Kidera, *Chem. Phys. Lett.* 282 (1998) 115.
- [172] C. Eckart, *Phys. Rev.* 47 (1935) 552.
- [173] A. Sayvetz, *J. Chem. Phys.* 7 (1939) 383.
- [174] M. Strain, G. Scuseria, M. Frisch, *Science* 271 (1996) 51.
- [175] P. Maslen, C. Ochsenfeld, C. White, M. Lee, M. Head-Gordon, *J. Phys. Chem.* 102 (1998) 2215.
- [176] L. Füsti-Molnar, P. Pulay, *J. Chem. Phys.* 116 (2002) 7795.
- [177] L. Genovese, T. Deutsch, A. Neelov, S. Goedecker, G. Beylkin, *J. Chem. Phys.* 125 (2006) 074105.
- [178] D. Fedak, N. Gjøstein, *Surf. Sci.* 8 (1967) 1967.
- [179] M. van Hove, R. Koestner, P. Stair, J. Biberian, L. Kesmodel, I. Bartos, G. Somorjai, *Surf. Sci.* 103 (1981) 189.
- [180] G. Binnig, H. Rohrer, C. Gerber, E. Stoll, *Surf. Sci.* 144 (1984) 321.
- [181] D. Gibbs, B. Ocko, D. Zehner, S. Mochrie, *Phys. Rev. B* 42 (1990) 7330.