Computer Physics Communications 249 (2020) 107065

Contents lists available at ScienceDirect

Computer Physics Communications

journal homepage: www.elsevier.com/locate/cpc

CPC 50th anniversary article

Questaal: A package of electronic structure methods based on the linear muffin-tin orbital technique☆

Dimitar Pashov [a], Swagata Acharya [a], Walter R.L. Lambrecht [b], Jerome Jackson [c,*], Kirill D. Belashchenko [d], Athanasios Chantis [e], Francois Jamet [a], Mark van Schilfgaarde [a]

[a] Department of Physics, King's College London, Strand, London WC2R 2LS, United Kingdom
[b] Department of Physics, Case Western Reserve University, Cleveland, OH 44106, USA
[c] Scientific Computing Department, STFC Daresbury Laboratory, Warrington WA4 4AD, United Kingdom
[d] Department of Physics and Astronomy and Nebraska Center for Materials and Nanoscience, University of Nebraska-Lincoln, Lincoln, NE 68588, USA
[e] American Physical Society, 1 Research Road, Ridge, NY 11961, USA

ARTICLE INFO

Article history:
Received 12 July 2019
Received in revised form 15 November 2019
Accepted 16 November 2019
Available online 3 December 2019

Keywords:
Questaal
Linear muffin tin orbital
Screening transformation
Density functional theory
Many-body perturbation theory

ABSTRACT

This paper summarises the theory and functionality behind Questaal, an open-source suite of codes for calculating the electronic structure and related properties of materials from first principles. The formalism of the linearised muffin-tin orbital (LMTO) method is revisited in detail and developed further by the introduction of short-ranged tight-binding basis functions for full-potential calculations. The LMTO method is presented in both Green's function and wave function formulations for bulk and layered systems. The suite's full-potential LMTO code uses a sophisticated basis and augmentation method that allows an efficient and precise solution to the band problem at different levels of theory, most importantly density functional theory, LDA+U, quasi-particle self-consistent GW and combinations of these with dynamical mean field theory. This paper details the technical and theoretical bases of these methods, their implementation in Questaal, and provides an overview of the code's design and capabilities.

Program summary

Program Title: Questaal
Program Files doi: http://dx.doi.org/10.17632/35jxxtzpdn.1
Code Ocean Capsule: https://doi.org/10.24433/CO.3778701.v1
Licensing provisions: GNU General Public License, version 3
Programming language: Fortran, C, Python, Shell
Nature of problem: Highly accurate *ab initio* calculation of the electronic structure of periodic solids and of the resulting physical, spectroscopic and magnetic properties for diverse material classes with different strengths and kinds of electronic correlation.
Solution method: The many electron problem is considered at different levels of theory: density functional theory, many body perturbation theory in the GW approximation with different degrees of self consistency (notably quasiparticle self-consistent GW) and dynamical mean field theory. The solution to the single-particle band problem is achieved in the framework of an extension to the linear muffin-tin orbital (LMTO) technique including a highly precise and efficient full-potential implementation. An advanced fully-relativistic, non-collinear implementation based on the atomic sphere approximation is used for calculating transport and magnetic properties.

1. Introduction

Different implementations of DFT are distinguished mainly by their basis set, which forms the core of any electronic structure method, and how they orthogonalise themselves to the core levels. Using these classifications most methods adopt one of four possible combinations shown in Fig. 1. In the vast major-

☆ This paper and its associated computer program are available via the Computer Physics Communication homepage on ScienceDirect (http://www.sciencedirect.com/science/journal/00104655)..

* Corresponding author.
E-mail address: jerome.jackson@stfc.ac.uk (J. Jackson).

https://doi.org/10.1016/j.cpc.2019.107065

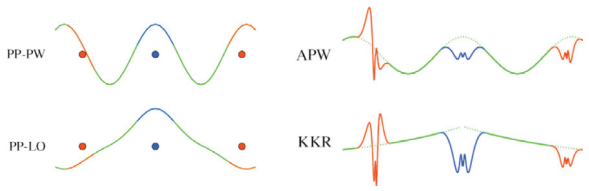

Fig. 1. 2 × 2 rubric for main classifications of basis set. Nuclei are shown as dots. The all-electron methods APW and KKR on the right substitute (augment) the envelope function (green) with numerical solutions of partial waves inside augmentation spheres (blue and red). Parts inside augmentation spheres are called "partial waves". The two figures on the left use a pseudopotential allowing their envelope functions to be smooth, with no augmentation needed. A pseudopotential's radius corresponds to a characteristic augmentation radius. The top two figures use plane waves for envelope functions; the bottom two use atom-centred local basis sets. We denote localised basis sets as "KKR", for the Korringa-Kohn-Rostocker method [3] as it plays a central role in this work; but there are other kinds, for example the Gaussian orbitals widely favoured among quantum chemists. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

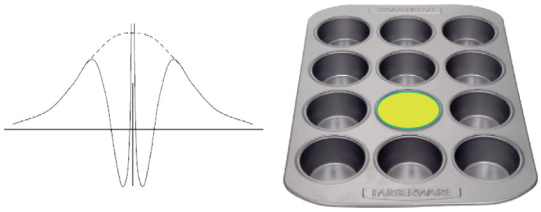

Fig. 2. Left: basis function in the vicinity of a nucleus, showing orthogonalisation to core states (solid line), and corresponding basis function in a pseudopotential (dashed line). Right: a muffin-tin potential: the potential is flat in the interstitial region between sites, and spherically symmetric in a volume around each site. Blue depicts an augmentation radius s_R around a sphere centred at some nuclear position **R** where the interstitial and augmented regions join.

ity of cases, basis sets consist of either atom-centred spatially localised functions (lower panel of Fig. 1), or plane waves (PW) (upper). As for treatment of the core, it is very common to substitute an effective repulsive (pseudo) potential to simulate its effect, an idea initially formulated by Conyers Herring [1]. Pseudopotentials make it possible to avoid orthogonalisation to the core, which allows the (pseudo) wave functions to be nodeless and smooth. For methods applied to condensed matter, the primary alternative method, formulated by Slater in 1937 [2], keeps all the electrons. Space is partitioned into non-overlapping spheres centred at each nucleus, with the interstitial region making up the rest. The basis functions are defined by plane waves in the interstitial, which are replaced ("augmented") by numerical solutions of the Schrödinger equations (partial waves) inside the augmentation spheres. The two solutions must be joined smoothly and differentiably on the augmentation sphere boundary (minimum conditions for a non-singular potential). Slater made a simplification: he approximated the potential inside the augmentation spheres with its spherical average, and also the interstitial potential with a constant. This is called the Muffin Tin (MT) approximation; see Fig. 2.

Solutions to spherical potentials are separable into radial and angular parts, $\phi_\ell(\varepsilon, r)Y_L(\hat{\mathbf{r}})$. The ϕ_ℓ are called partial waves and Y_L are the spherical harmonics. Here and elsewhere, angular momentum labelled by an upper case letter refers to both the ℓ and m parts. A lower case symbol refers to the orbital index only (ℓ is the orbital part of $L = (\ell, m)$). The ϕ_ℓ are readily found by numerical integration of a one-dimensional wave equation (Section 2), which can be efficiently accomplished.

An immense amount of work has followed the original ideas of Herring and Slater. The Questaal package is an all-electron implementation in the Slater tradition, so we will not further discuss the vast literature behind the construction of a pseudopotential, except to note there is a close connection between pseudopotentials and the energy linearisation procedure to be described below. Blöchl's immensely popular Projector Augmented-Wave method [4] makes a construction intermediate between pseudopotentials and APWs. Questaal uses atom-centred envelope functions instead of plane waves (Section 3), and an augmentation scheme that resembles the PAW method but can be converged to an exact solution for the reference potential, as Slater's original method did. The spherical approximation is still almost universally used to construct the basis set, and thanks to the variational principle, errors are second order in the nonspherical part of the potential. The nonspherical part is generally quite small, and this is widely thought to be a very good approximation, and the Questaal codes adopt it.

For a MT potential (V_{MT} taken to be 0 for simplicity), the Schrödinger equation for energy ε has locally analytical solutions: in the interstitial the solution can expressed as a plane wave $e^{i\mathbf{k}\cdot\mathbf{r}}$, with $\varepsilon = \hbar^2k^2/2m$. (We will use atomic Rydberg units throughout, $\hbar = 2m = e^2/2 = 1$.) In spherical coordinates envelope functions can be Hankel functions $H_L(E, \mathbf{r}) = h_\ell(kr)Y_L(\hat{\mathbf{r}})$ or Bessel functions $j_\ell(kr)Y_L(\hat{\mathbf{r}})$, except that Bessel functions are excluded as envelope functions because they are not bounded in space. Inside the augmentation spheres, solutions consist of some linear combination of the ϕ_ℓ.

The all-electron basis sets "APW" (augmented plane wave) and "KKR" [3] are both instances of augmented-wave methods: both generate arbitrarily accurate solutions for a muffin-tin potential. They differ in their choice of envelope functions (plane waves or Hankel functions), but they are similar in that they join onto solutions of partial waves in augmentation spheres. Both basis sets are energy-dependent, which makes them very complicated and their solution messy and slow. This difficulty was solved by O.K. Andersen in 1975 [5]. His seminal work paved the way for modern "linear" replacements for APW and KKR, the LAPW and Linear Muffin Tin Orbitals (LMTO) methods. By making a linear approximation to the energy dependence of the partial waves inside the augmentation spheres (Section 2.4), the basis set can be made energy-independent and the eigenfunctions and eigenvalues of the effective one-particle equation obtained with standard diagonalisation techniques. LAPW, with local orbitals added to widen the energy window over which the linear approximation is valid (Section 3.7.3) is widely viewed to be the industry gold standard for accuracy. Several well known codes exist:for example FLEUR (http://www.flapw.de), and WIEN2K (http://susi.theochem.tuwien.ac.at) and its descendants such as the Exciting code (http://exciting-code.org). A recent study [6] established that these codes all generate similar results when carefully converged. Questaal's main DFT code is a generalisation of the LMTO method (Section 3), using for the envelope functions the more flexible smooth Hankel functions (Section 3.1) instead of standard Hankels. Accuracy of the smooth-Hankel basis is also high (Section 3.13), and though not quite reaching the standard of the LAPW methods, it is vastly more efficient. If needed, Questaal can add APW's to the basis to converge it to the LAPW standard (Section 3.10).

1.1. Questaal's history

Questaal has enjoyed a long and illustrious history, originating in O.K. Andersen's group in the 1980s as the standard "Stuttgart" code. It has undergone many subsequent evolutions, e.g. an early all-electron full-potential code [7], which was used in one of the first *ab initio* calculations of the electron–phonon interaction for superconductivity [8], an efficient molecules code [9] which was employed in the first *ab initio* description of instanton dynamics [10], one of the first noncollinear magnetic codes and the first

ab initio description of spin dynamics [11], first implementation of exact exchange and exact exchange+correlation [12], one of the first all-electron *GW* implementations [13], and early density-functional implementations of non-equilibrium Green's functions for Landauer–Buttiker transport [14]. In 2001 Aryasetiawan's *GW* was extended to a full-potential framework to become the first all-electron *GW* code [15]. Soon after the concept of quasiparticle self-consistency was developed [16], which has dramatically improved the quality of *GW*. Its most recent extension is to combine QS*GW* with DMFT. It and the code of Kutepov et al. [17] are the first implementations of QS*GW*+dynamical mean field theory (DMFT); and to the best of our knowledge Questaal has the only implementation of response functions (spin, charge, superconducting) within QS*GW*+DMFT.

1.2. Main features of the Questaal package

Ideally a basis set is *complete*, *minimal*, and *short ranged*. We will use the term *compact* to mean the extent to which a basis can satisfy all these properties: the faster a basis can yield converged results for a given rank of Hamiltonian, the more compact it is. It is very difficult to satisfy all these properties at once. KKR is by construction *complete* and *minimal* for a "muffin-tin" potential, but it is not short-ranged. In 1984 it was shown (by Andersen once again! [18]) how to take special linear combinations of muffin-tin orbitals ("screening transformation") to make them short ranged. Andersen's screening transformation was derived for LMTOs, in conjunction with his classic Atomic Spheres Approximation, (ASA, Section 2.7), and screening has subsequently been adopted in KKR methods also. The original Questaal codes were designed around the ASA, and we develop it first in Section 2.5. Its main code no longer makes the ASA approximation, and generalises the LMTOs to more flexible functions; that method is developed in Section 3. These functions are nevertheless long-ranged and cannot take advantage of very desirable properties of short-ranged basis sets. Very recently we have adapted Andersen's screening transformation to the flexible basis of full-potential method (Section 2.9). Screening provides a framework for the next-generation basis of "Jigsaw Puzzle Orbitals" (JPOs) that will be a nearly optimal realisations of the three key properties mentioned above.

Most implementations of *GW* are based on plane waves (PWs), in part to ensure completeness, but also because implementation is greatly facilitated by the fact that the product of two plane waves is analytic. *GW* has also been implemented recently in tight-binding forms using e.g., a numerical basis [19], or a Gaussian basis [20]. None of these basis sets is very compact. The FHI AIMS code *can* be reasonably well converged, but only at the expense of a large number of orbitals. Gaussian basis sets are notorious for becoming over-complete before they become complete. Questaal's JPO basis – still under development – should bring into a single framework the key advantages of PW and localised basis sets.

Questaal implements:

1. density-functional theory (DFT) based on common LDA and GGA exchange–correlation functionals (other LDA or GGA, but not meta-GGA, variants are available via `libxc`[21]). There is a standard full-potential DFT code, `lmf` (Section 3), and also three codes (`lm`, `lmgf`, `lmpg`) that implement DFT in the classical Atomic Spheres Approximation [5,18], presented in Section 2.7. The latter use the screened, tight-binding form (Section 2.10). `lm`, a descendant of Andersen's standard ASA package (Stuttgart code), is an approximate, fast form of `lmf`, useful mainly for close-packed magnetic metals; `lmgf` (Section 2.12) is a Green's function implementation closely related to the KKR Green's function, parameterised so that it can be linearised. Section 2.13 shows how this is accomplished, resulting in an efficient, energy-independent Hamiltonian `lm` uses. `lmgf` has two useful extensions: the coherent potential approximation (CPA, Section 2.18) and the ability to compute magnetic exchange interactions. `lmpg` (Section 2.14) is a principal layer Green's function technique similar to `lmgf` but designed for layer geometries (periodic boundary conditions in two dimensions). `lmgf` is particularly useful for Landauer–Buttiker transport [22–25], and it includes a nonequilibrium Keldysh technique [14];
2. density functional theory with local Hubbard corrections (LDA+U) with various kinds of double-counting correction (Section 3.8);
3. the *GW* approximation based on DFT. Questaal's *GW* package is separate from the DFT code; there is an interface (`lmfgwd`) that supplies DFT eigenfunctions and eigenvalues to it. It was originally formulated by Aryasetiawan in the ASA [26], derived from the Stuttgart ASA code; and it was the first all-electron *GW* implementation. Kotani and van Schilfgaarde extended it to a full-potential framework in 2002 [15]. A shell script `lmgw` runs the *GW* part and manages links between it and `lmf`;
4. the Quasiparticle Self-Consistent *GW* (QS*GW*) approximation, first formulated by Faleev, Kotani and van Schilfgaarde [16]. Questaal's QS*GW* is a descendent of Kotani's original code, which with some modest modifications can be found at https://github.com/tkotani/ecalj/. QS*GW* also works synchronously with `lmf`, yielding either a static QS*GW* self-energy Σ_0 which `lmf` reads as an effective external potential added to the Hamiltonian, or a dynamical self-energy (diagonal part) which a post-processing code `lmfgws`, uses to construct properties of an interacting Green's function, such as the spectral function. See Section 4;
5. extensions to Dynamical Mean Field Theory using a bath based on either DFT or QS*GW*. Questaal does not have its own implementation of DMFT, but has interfaces to Haule's CTQMC solver [27], and to the `TRIQS` library [28]. See Section 5;
6. an empirical tight-binding model (`tbe`). The interested reader is referred to Refs. [29,30] and the Questaal web site; and
7. a large variety of other codes and special-purpose editors to supply input the electronic structure methods, or to process the output.

1.3. Outline of the paper

The aim of the paper is to provide a consistent presentation of the key expressions and ideas of the LMTO method, and aspects of electronic structure theory as they are implemented in the different codes in the Questaal suite. Necessarily this presentation is rather lengthy; the paper is organised as follows. Section 2 describes the LMTO basis, assuming the muffin-tin and atomic sphere descriptions of the potential. Together with tail cancellation and linearisation this comprises therefore a derivation of the traditional ("second generation") LMTO method. The transformation to a short range tight-binding-like basis, and to other representations is described. The formulation of the crystal Green's function in terms of the LMTO potential parameters is derived, allowing the use of coherent potential approximation alloy theory. Non-collinear magnetism and fully-relativistic LMTO techniques are presented. Section 3 describes the full-potential code: its basis, augmentation method, core treatment and other technical aspects are described in detail. The LDA+*U*

method, relativistic effects, combined LMTO+APW, and numerical precision are also discussed. Section 4 describes the *GW* approximation: the importance of self-consistency, the nature of QS*GW* in particular and its successes and limitations. Section 5 describes Questaal's interface to different DMFT solvers; Section 6 discusses calculation of spin and charge susceptibilities. Our perspective on realising a high fidelity solution to the many-body problem for solids is described in Section 7. Finally, software aspects of the Questaal project are outlined in Section 8. Several appendices are provided with a number of useful and important relations for the LMTO methodology.

2. The muffin-tin potential and the atomic spheres approximation

As noted earlier, the KKR method solves the Schrödinger equation in a MT potential to arbitrarily high accuracy. We develop this method first, and show how the LMTO method is related to it. In Section 3.1 we show how the basis in `lmf` is related to LMTOs. The original KKR basis set consists of spherical Hankel (Neumann) functions $H_{RL}(E,\mathbf{r}) = h_{R\ell}(kr)Y_L(\hat{\mathbf{r}})$ as envelope functions with $k^2{=}E$, augmented by linear combinations of partial waves $\phi_{R\ell}(\varepsilon,r)Y_{RL}(\hat{\mathbf{r}})$ inside augmentation spheres (Fig. 3). (See the Appendix for definitions and the meaning of subscripts R and L.) The envelope must be joined continuously and differentiably onto augmentation parts, since the kinetic energy cannot be singular. Note that for large ℓ, $\phi_{R\ell} \to \text{const}\times r^\ell$. This is because the angular part of the kinetic energy becomes dominant for large ℓ.

2.1. One-centre expansion of Hankel functions

An envelope function $H_L(E,\mathbf{r})$ has a "head" centred at the origin where it is singular and tails at sites around the origin. In addition to the head, tails must be augmented by linear combinations of $\phi_\ell(\varepsilon,r)$ centred there, which require H_L be expanded in functions centred at another site. If the remote site is at $\mathbf{R}$ relative to the head, the one-centre expansion can be expressed as a linear combination of Bessel functions

$$H_L(E,\mathbf{r}) = \sum_M S_{ML}(E,\mathbf{R})J_M(E,\mathbf{r}-\mathbf{R}) \tag{1}$$

This follows from the fact that both sides satisfy the same second order differential equation $(\nabla^2 + E) = 0$. The two functions centred at the origin satisfying this equation are the Hankel and Bessel functions. Hankel functions have a pole at $r = 0$, whereas Bessel functions are regular there, so this relation must be true for all $r < |\mathbf{R}|$. The larger the value of r, the slower the convergence with M.

Expressions for the expansion coefficients S can be found in various textbooks; they of course depend on how H and J are defined. Our standard definition is

$$S_{MK}(E,\mathbf{R}) = 4\pi \sum_L C_{KLM}(-1)^k(-E)^{(k+m-\ell)/2}H_L(E,\mathbf{R}) \tag{2}$$

The C_{KLM} are Gaunt coefficients (Eq. (A.8)).

To deal with solids with many sites, we write the one-centre expansion using subscript R to denote a nucleus and $\mathbf{r}$ relative to the nuclear coordinate:

$$H_{RL}(E,\mathbf{r}) = \sum_M S_{R'M,RL}(E)J_{R'M}(E,\mathbf{r}) \tag{3}$$

Envelope functions then have two ℓ cutoffs: ℓ_b for the head at R, and ℓ_a for the one-centre expansion at R'. These need not be the same: ℓ_b determines the rank of the Hamiltonian, while ℓ_a is a cutoff beyond which $\phi_{R'\ell}\times\phi_{R'\ell'}$ is well approximated by const $\times$ $r^{\ell+\ell'}$. A reasonable rule of thumb for reasonably well converged calculations is to take ℓ_b one number larger than the highest ℓ character in the valence bands. Thus $\ell_b = 2$ is reasonable for *sp* elements, $\ell_b = 3$ for transition metal elements, $\ell_b = 4$ for *f* shell elements. In the ASA, reasonable results can be obtained for $\ell_b = 2$ for transition metal elements, and $\ell_b = 3$ for *f* shell elements. As for ℓ_a, traditional forms of augmentation usually require $\ell_a = 2\ell_b$ for comparable convergence: `lmf` does it in a unique way that converges much faster than the traditional form and it is usually enough to take $\ell_a = \ell_b + 1$ [31] (Section 3.6).

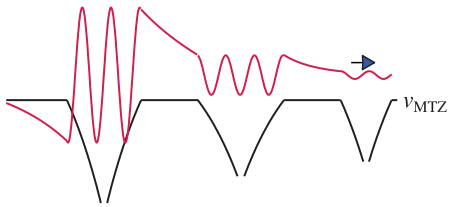

Fig. 3. Schematic of muffin-tin potential (black) and a solution (red) in a three-atom chain.

2.2. Partial waves in the MT spheres

Partial waves in a sphere of radius s about a nucleus must be solved in the four coordinates, $\mathbf{r}$ and energy ε. Provided $\mathbf{r}$ is not too large, $v(\mathbf{r})$ is approximately spherical, $v(\mathbf{r})\approx v(r)$; Even if the potential is not spherical, it is assumed to be for construction of the basis set, as noted earlier. Solutions for a spherical potential are partial waves $\phi_\ell(\varepsilon,r)Y_L(\hat{\mathbf{r}})$, where Y_L are the spherical harmonics. Usually Questaal uses real harmonics $Y_L(\hat{\mathbf{r}})$ instead (see Appendix for definitions).

$\phi_\ell(\varepsilon,r)$ satisfies Schrödinger's equation

$$(-\nabla^2 + v(r) - \varepsilon)\phi_\ell(\varepsilon,r) = 0 \tag{4}$$

We are free to choose the normalisation of ϕ and use

$$\int_0^s \phi^2(\varepsilon,r)r^2dr = 1 \tag{5}$$

One way to think about solutions of Schrödinger's equation is to imagine each nucleus, with some $v_R(r)$ around it, as a scatterer to waves incident on it. Scattering causes shifts in the phase of the outgoing wave. The condition that all the sites scatter coherently, which allows them to sustain each other without any incident wave, gives a solution to Schrödinger's equation. This condition is explicit in the KKR method, and forms the basis for it. Imagine a "muffin-tin" potential—flat in the interstitial but with holes carved out around each nucleus of radius s. The phase shift from a scatterer at R is a property of the shape of $\phi_{R\ell}$ at its MT boundary, $\phi_{R\ell}(\varepsilon,s)$. Complete information about the scattering properties of the sphere, if $v(\mathbf{r}) = v(r)$, can be parameterised in terms of $\phi_{R\ell}(\varepsilon,s)$ and its slope, as we will see.

For a small change in energy of the incident wave, there can be a strong change in the phase shift. In the region near an eigenstate of the free atom the energy dependence is much stronger for the partial wave than it is for the incident waves striking it, so electronic structure methods focus on the scattering properties of the partial waves ϕ_ℓ. This information is typically expressed through the logarithmic derivative function

$$D_\ell(\varepsilon) \equiv D\{\phi_\ell(\varepsilon)\} = \left(\frac{r}{\phi_\ell(\varepsilon,r)}\frac{d\phi_\ell(\varepsilon,r)}{dr}\right)_s \tag{6}$$

Consider the change in $\phi_\ell(\varepsilon,r)$ with ε for a given $v(r)$. As ε increases $\phi_\ell(\varepsilon,r)$ acquires more curvature (Fig. 4). In the interval between $\varepsilon = \varepsilon_0$ where $\phi'_\ell(\varepsilon_0,s) = 0$ so that $D_\ell = 0$, and ε_2 where

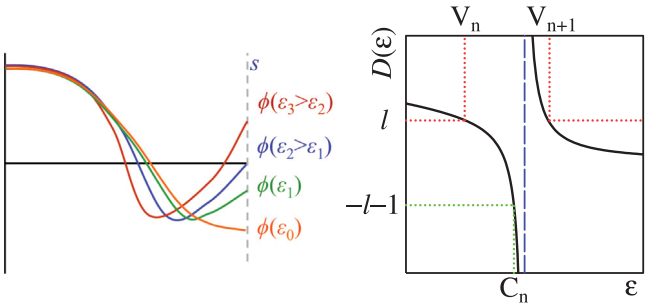

Fig. 4. Variation of $\phi_\ell(\varepsilon, r)$ with ε for an s orbital.

$\phi_\ell(\varepsilon_2, s) = 0$, $\phi'_\ell(\varepsilon, s)$ is positive. D_ℓ thus decreases monotonically, with $D_\ell \to -\infty$ as $\varepsilon \to \varepsilon_2$. At some ε_1 in this region $D_\ell = -\ell - 1$, which is the logarithmic derivative of a Hankel function of energy 0. ε_1 is called the "band centre" C_ℓ for reasons to be made clear in Section 2.5: it is close to an atomic level and in tight-binding theory would correspond to an on-site matrix element. Increasing from ε_2, D_ℓ decreases monotonically from $+\infty$ as shown in Fig. 4, reaching 0 once more, passing through some ε_3 where $D_\ell = +\ell$. This is the logarithmic derivative of a Bessel function of energy 0, and is traditionally called V_ℓ.

Thus D_ℓ is a monotonically decreasing cotangent-like function of ε, with a series of poles. At each pole $\phi_\ell(\varepsilon, r)$ acquires an additional node, incrementing the principal quantum number n. Between poles n is fixed and there are parameters C_ℓ and V_ℓ for each n. The linear method approximates D_ℓ with a simple pole structure, which is accurate over a certain energy window. Similarly pseudopotentials are constructed by requiring the pseudofunction to match D_ℓ of the free atom, in a certain energy window.

For principal quantum number n, ϕ_ℓ has $n - \ell - 1$ nodes and may vary rapidly to be orthogonal to deeper nodes. This poses no difficulty: we use a shifted logarithmic radial mesh, with point i given by

$$r_i = b\{e^{a(i-1)} - 1\}$$

Typically a few hundred points are needed for accurate integration. Core and valence waves use the same mesh.

2.3. Energy derivative of D

Consider the matrix element integrated over a sphere of radius s:

$$0 = \langle \phi_\ell(\varepsilon_\nu)| - \nabla^2 + v - \varepsilon\, |\phi_\ell(\varepsilon)\rangle$$

ε_ν is some fixed energy. Taking into account the boundary condition at s, we obtain

$$\langle \phi_\ell(\varepsilon)|\varepsilon - \varepsilon_\nu|\phi_\ell(\varepsilon_\nu)\rangle = -[D_\ell(\varepsilon) - D_\ell(\varepsilon_\nu)]s\phi_\ell(\varepsilon_\nu, s)\phi_\ell(\varepsilon, s)$$

from Green's second identity. From this we obtain the energy derivative of D_ℓ as [5]

$$\dot{D}_\ell(\varepsilon) \equiv \lim_{\varepsilon\to\varepsilon_\nu} \frac{D_\ell(\varepsilon) - D_\ell(\varepsilon_\nu)}{\varepsilon - \varepsilon_\nu} = \frac{-1}{s\phi_\ell^2(\varepsilon, s)} \tag{7}$$

2.4. Linearisation of energy dependence in the partial waves

An effective way to solve Schrödinger's equation is to linearise the energy-dependence of the partial waves $\phi_\ell(\varepsilon, r)$, as

$$\phi_\ell(\varepsilon, r) \approx \phi_\ell(\varepsilon_\nu, r) + (\varepsilon - \varepsilon_\nu)\dot{\phi}_\ell(\varepsilon_\nu, r) \tag{8}$$

This was Andersen's most important contribution to electronic structure theory [5]: it had a dramatic impact on the entire field. We will make extensive use of the linear approximation here.

In the linear approximation, four parameters ($\phi(s)$, $\dot{\phi}(s)$, and their logarithmic derivatives) completely characterise the scattering properties of a sphere with $v = v(r)$. Only three of them turn out to be independent. To see this, obtain an equation for $\dot{\phi}$ by differentiating Eq. (4) w.r.t. ε:

$$(-\nabla^2 + v - \varepsilon)\dot{\phi}(\varepsilon, r) = \phi(\varepsilon, r) \tag{9}$$

With the normalisation Eq. (5), ϕ and $\dot{\phi}$ are orthogonal

$$\langle \phi(\varepsilon_\nu)\dot{\phi}(\varepsilon_\nu)\rangle = 0 \tag{10}$$

Using the normalisation Eq. (5), we can establish the following relation between $\phi(\varepsilon_\nu, s)$, $\dot{\phi}(\varepsilon_\nu, s)$, $D\{\phi\}$ and $D\{\dot{\phi}\}$:

$$\begin{aligned} 1 &= \langle [\phi(\varepsilon_\nu)]^2\rangle = \langle \phi(\varepsilon_\nu)| - \nabla^2 + v - \varepsilon_\nu|\dot{\phi}(\varepsilon_\nu)\rangle \\ &= \langle \dot{\phi}(\varepsilon_\nu)| - \nabla^2 + v - \varepsilon_\nu|\phi(\varepsilon_\nu)\rangle + W\{\phi, \dot{\phi}\} \\ &= W\{\phi(s), \dot{\phi}(s)\} \\ &= [D\{\phi(\varepsilon_\nu)\} - D\{\dot{\phi}(\varepsilon_\nu)\}]s\phi(\varepsilon_\nu, s)\dot{\phi}(\varepsilon_\nu, s) \end{aligned} \tag{11}$$

The third line follows from Eq. (4), and the second from Green's second identity which adds a surface term when ϕ and $\dot{\phi}$ are interchanged. The Wronskian $W\{a, b\}$ is defined as

$$\begin{aligned} W\{a(s), b(s)\} &\equiv s^2\left[a(s)b'(s) - a'(s)b(s)\right] \\ &= sa(s)b(s)\left[D\{b\} - D\{a\}\right] \end{aligned} \tag{12}$$

for a pair of functions $a(r)$ and $b(r)$ evaluated at point s.

2.5. The traditional LMTO method

For historical reasons Anderson constructed the original LMTO formalism with non-standard definitions for the Hankel and Bessel functions. We follow those definitions in order to be consistent with the historical literature. In this paper they are named $\hat{H}_\ell$ and $\hat{J}_\ell$ and are defined in Appendix B. Here we follow Andersen's development only in the context of $E \le 0$, with $\kappa^2 = -E$. Note that E can be chosen freely and need not be connected to the eigenvalue ε. But for exact solutions in a MT potential, the energy of $\hat{H}_L(E, \mathbf{r})$ and $\hat{J}_\ell(E, \mathbf{r})$ must be chosen so $E = \varepsilon - v_{\mathrm{MTZ}}$.

LMTO and KKR basis sets solve Schrödinger's equation in a muffin-tin potential, which in the interstitial reduces to the Helmholtz equation, and have linear combinations of Hankel functions as solutions that satisfy appropriate boundary conditions. We defer treatment of the boundary conditions to the next section and continue the analysis of partial waves for a single scattering centre for now.

As noted, the scattering properties depend much more strongly on the partial waves than the energy dependence of the envelopes. In keeping with the traditional LMTO method, we assume that the kinetic energy of the envelopes vanishes (the energy is taken to be close to the MT potential), then Schrödinger's equation reduces further to Laplace's equation, whose solutions are Hankel and Bessel functions at $E = 0$, which we denote as $\hat{H}_\ell(\mathbf{r})$ and $\hat{J}_\ell(\mathbf{r})$. In close-packed solids there is reasonable justification for this choice: the spacing between spheres is much smaller than the wavelength of a low-energy solution to the wave equation (one not too far from the Fermi level). $\hat{H}_\ell(\mathbf{r})$ and $\hat{J}_\ell(\mathbf{r})$ are proportional to $r^{-\ell-1}$ and r^ℓ respectively (see Appendix) and so ϕ_ℓ can be continued into the interstitial in the vicinity of s

$$\begin{aligned} \phi_\ell(\varepsilon, r\sim s) = \phi_\ell(\varepsilon, s)\frac{\ell + 1 + D\{\varepsilon\}}{2\ell + 1}\left(\frac{r}{s}\right)^\ell \\ + \phi_\ell(\varepsilon, s)\frac{\ell - D\{\varepsilon\}}{2\ell + 1}\left(\frac{s}{r}\right)^{\ell+1} \end{aligned} \tag{13}$$

The first term is proportional to $\hat{J}_\ell(r)$, the second to $\hat{H}_\ell(r)$. In the remainder of this section we will develop expressions for the general κ case, showing also the $\kappa \to 0$ limit, and finally focus on constructing Hamiltonians and Green's functions with $\kappa = 0$.

2.6. Energy-dependent muffin-tin orbitals

Eq. (13) is not yet a suitable basis because it diverges as $r \to \infty$ because of the J_ℓ term. However, we can construct a family of "muffin-tin" orbitals that are continuous and differentiable

$$\chi_{RL}(\varepsilon, E, \mathbf{r}) = Y_L(\hat{\mathbf{r}}_R) \times \begin{cases} N_{R\ell}(\varepsilon)\,\phi_{R\ell}(\varepsilon, r_R) + P_{R\ell}(\varepsilon)\,\hat{J}_\ell(E, r_R) & r_R < s_R \\ \hat{H}_\ell(E, r_R) & r_R > s_R \end{cases} \tag{14}$$

and that do not diverge for large r_R. $N_{R\ell}$ and $P_{R\ell}$ are coefficients fixed by requiring that the value and slope are continuous at s_R. Thus for $r < s$, $\chi_{RL}(\varepsilon, E, \mathbf{r})$ consists of a linear combination of $\phi_{R\ell}(r)$ and $\hat{J}_{R\ell}(E, r)$ that matches smoothly and differentiably onto $\hat{H}_{R\ell}(E, r)$. Apart from the "contaminating" $\hat{J}_{R\ell}$ term, χ_{RL} is a solution for a single MT potential at sphere R. It vanishes at ε corresponding to the eigenvalue of the MT "atom", which occurs at $\varepsilon = C_\ell$. (Taking $E = 0$, $\varepsilon = C_\ell$ when $D_\ell = -\ell - 1$; see Eq. (13).) In a lattice $\hat{H}_\ell$ must also be augmented at all $R' \neq R$. To form an eigenstate, the contaminating term must be cancelled out by tails from $\chi_{R'L'}$ centred elsewhere. Since any $\hat{H}_{R'L'}(E, r)$ can be expanded as linear combinations of $\hat{J}_{RL}(E, r)$, Eq. (3), it is easy to anticipate how the "tail cancellation condition", which forms the basis of the KKR method (2.6.1), comes about.

The P_ℓ are called "potential functions" and play a central role in constructing eigenfunctions. Expressions for P_ℓ and N_ℓ are developed in Section 2.8. Combined with the linearisation of the partial waves, (Section 2.4), information about P can be encapsulated in a small number of parameters; see Section 2.8.1.

Note the similarity between the χ_{RL} and the partial waves. There is a difference in normalisation, but more importantly the term proportional to $\hat{J}_{RL}$ for $r_R > s$ in Eq. (13) must be taken out of the MTO because it diverges for large r, as noted. Since $\hat{J}_\ell$ is present for $r < s$, χ_{RL} is not a solution of Schrödinger's equation for $r_R < s$. However any linear combination of the χ_{RL}

$$\Psi(\epsilon, \kappa, \mathbf{r}) = \sum_{RL} z_{RL}\chi_{RL}(\epsilon, \kappa, \mathbf{r}) \tag{15}$$

can be taken as a trial solution to Schrödinger's equation. The $z_{R'L'}$ are expansion coefficients, which become the eigenvector if Ψ is an eigenstate. For any $z_{R'L'}$, Ψ solves the interstitial exactly if the potential is flat and $-E$ is chosen to correspond to the kinetic energy in the interstitial, $E = \varepsilon - v_{\mathrm{MTZ}}$, since each χ_{RL} individually satisfies Schrödinger's equation.

2.6.1. Tail cancellation

Inside sphere R there are three contributions to $\Psi(\varepsilon, \mathbf{r})$: partial waves from the "head" function χ_{RL}, the Bessel part of that function, and contributions from the tails of $\chi_{R'L'}$ centred at other sites, which are also Bessel functions, Eq. (3). Thus, all the contributions to Ψ inside some sphere R, in addition to the partial wave, consist of some linear combination of Bessel functions. We can find exact solutions for the MT potential by finding particular linear combinations z_{RL} that cause all the $\hat{J}_{RL}$ inside each augmentation sphere to cancel.

From the definition Eq. (14), Eq. (15), has a one-centre expansion inside sphere R

$$\phi_{RL}(\varepsilon, \mathbf{r})N_{R\ell}(\varepsilon)z_{RL} + \hat{J}_{RL}(E, \mathbf{r})P_{R\ell}(\varepsilon)z_{RL} - \sum_{L'} S_{RL',R'L}(E)\hat{J}_{RL'}(E, \mathbf{r})z_{RL}$$

The one-centre expansion satisfies Schrödinger's equation provided that the second and third terms cancel. This leads to the "tail cancellation" theorem

$$\sum_{RL}\left[P_{RL}(\varepsilon)\delta_{R'L',RL} - S_{R'L',RL}(E)\right]z_{RL} = 0 \tag{16}$$

and is the fundamental equation of KKR theory. For non-trivial solutions ($|z_{RL}| \neq 0$), the determinant of the matrix $P - S$ must be zero. This will only occur for discrete energies $\varepsilon_i(P)$ for which $P - S$ has a zero eigenvalue. The corresponding eigenvector z_{RL} yields an eigenfunction, Eq. (15), which exactly solves Schrödinger's equation for the MT potential in the limit $L \to \infty$ and if E is taken to be the proper kinetic energy in the interstitial, $E = \varepsilon_i - v_{\mathrm{MTZ}}$. In general there will be a spectrum of eigenvalues ε_i that satisfy Eq. (16). The quantity

$$g_{RL,RL'}(\varepsilon) = [P_{RL}\delta_{RL,RL'}(\varepsilon) - S_{RL,RL'}]^{-1} \tag{17}$$

is called the "auxiliary Green's function" and is closely related to the true Green's function G [32] (poles of g and G coincide). We will develop the connection in Section 2.12. In KKR theory g is called the "scattering path operator".

2.7. The atomic spheres approximation

Andersen realised very early that more accurate solutions could be constructed by overlapping the augmentation spheres so that they fill space. There is a trade-off in the error arising from the geometry violation in the region where the spheres overlap and improvement to the basis set by using partial waves in this region rather than envelope functions. The Atomic Spheres approximation, or ASA, is a shorthand for three distinct approximations:

- $v(\mathbf{r})$ is approximated by a superposition of spherically symmetric $v_R(r)$, with a flat potential in the interstitial;
- the MT spheres are enlarged to fill space, so that the interstitial volume is zero. The resulting geometry violations are ignored, except that the interstitial can be accounted for assuming a flat potential (the "combined correction" term). Errors associated with the geometry violation were carefully analysed by Andersen in his NMTO development [33]; and
- the envelope functions are Hankel functions with $\kappa = 0$, augmented by partial waves inside MT spheres. There is no difficulty in working with $\kappa \neq 0$, but $\kappa = 0$ is a good average choice, as noted above. As ASA is an approximate method, little is gained by trying to improve it in this way. Real potentials are not muffin-tins, and the loss of simplicity does not usually compensate for limited gain in precision. The full-potential methods use better envelope functions (Section 3).

2.7.1. Tail cancellation in the ASA

In the ASA the spheres fill space, making the interstitial volume null. By normalising the χ of Eq. (14) as defined there, the eigenvectors z_{RL} of $P - S$ ensure that Ψ is properly normalised if

$$\sum_{RL}|N_{R\ell}z_{RL}|^2 = 1 \tag{18}$$

This is because the Bessels all cancel, and the wave function inside sphere R is purely $\phi_{R\ell}N_{R\ell}z_{RL}$. Normalisation of z_{RL} with the normalisation of ϕ (Eq. (5)) ensures that $\langle\Psi|\Psi\rangle = 1$.

Making the spheres fill space is a better approximation than choosing spheres with touching radii, even with its geometry violation. This is because potentials are not flat and partial waves ϕ_ℓ are better approximations to the eigenfunctions than the $\hat{H}_\ell$. Moreover, it can be shown [34] that the resulting wave function

is equivalent to the exact solution of the Schrödinger equation for $v(\mathbf{r})$ equal to the sum of overlapping spherical potentials $v(\mathbf{r}) = \sum_R v_R(r)$. This means $v(\mathbf{r})$ is deeper along lines connecting atoms with a corresponding reduction of $v(\mathbf{r})$ along lines pointing into voids, corresponding to the accumulation (reduction) of density in the bonds (voids).

Ψ varies in a nonlinear way with ε, so the tail cancellation condition entails a nonlinear problem. Once $P(\varepsilon)$ (more precisely $1/P$) is linearised (Section 2.8), the cancellation condition simplifies to a linear algebraic eigenvalue problem. This provides a framework, through the linear approximation, Eq. (8), for constructing an efficient, energy-independent basis set that yields solutions from the variational principle, without relying on tail cancellation.

2.8. Potential and normalisation functions

The tail cancellation condition Eq. (16), is conveniently constructed through the "potential function" $P_\ell(\varepsilon)$, P_ℓ is closely related to the logarithmic derivative $D_\ell(\varepsilon)$, Eq. (6), which parameterises the partial wave in isolation. P_ℓ depends on both D_ℓ and the boundary conditions, which depend on what we select for the interstitial kinetic energy E. $P_\ell(\varepsilon)$ is an always increasing tangent like function of energy, and in the language of scattering theory, it is proportional to the cotangent of the phase shift.

If the potential were not spherical, a more general tail cancellation theorem would still be possible, but P would need be characterised by additional indices: $P = P_{L,L'}(\varepsilon)$ while for a spherical potential P depends on ℓ only. This is the only case we consider here.

N_ℓ and P_ℓ are fixed by requiring that $\chi_\ell(\varepsilon, \kappa, r)$ be continuous and differentiable at s. Expressions for N_ℓ and P_ℓ are conveniently constructed by recognising that any function $f(r)$ can be expressed as a linear combination of $a(r)$ and $b(r)$ near s (meaning it connects smoothly and differentiably at $r = s$) through the combination

$$f(r) = [W\{f, b\}a(r) - W\{f, a\}b(r)]\, W\{a, b\}^{-1} \tag{19}$$

Thus the matching conditions require N and P to be

$$P_\ell(\varepsilon) = \frac{W\{\hat{H}_\ell, \phi_\ell\}}{W\{\hat{J}_\ell, \phi_\ell\}} \xrightarrow{\kappa\to 0} 2(2\ell+1)\left(\frac{w}{s}\right)^{2\ell+1} \frac{D\{\phi_\ell(\varepsilon)\} + \ell + 1}{D\{\phi_\ell(\varepsilon)\} - \ell} \tag{20}$$

$$N_\ell(\varepsilon) = \frac{W\{\hat{J}_\ell, \hat{H}_\ell\}}{W\{\hat{J}_\ell, \phi_\ell\}} \tag{21}$$

w is an arbitrary length scale, typically set to the average value of s.

With the help of Eq. (7) the energy derivative of P_ℓ is readily shown to be (Eq. A21, Ref. [32])

$$\dot{P}_\ell = \frac{dP_\ell}{d\varepsilon} = \frac{w/2}{[W\{\hat{J}_\ell, \phi_\ell\}]^2} \tag{22}$$

Using the following relation between Hankels and Bessels,

$$W\{\hat{H}_\ell, \hat{J}_\ell\} = w/2$$

it is readily seen that

$$\dot{P}_\ell = \frac{2}{w}\frac{[W\{\hat{H}_\ell, \hat{J}_\ell\}]^2}{[W\{\hat{J}_\ell, \phi_\ell\}]^2} = \frac{2}{w}N_\ell^2$$

and therefore

$$N_\ell = \sqrt{\dot{P}_\ell w/2} \tag{23}$$

2.8.1. Linearisation of P

In this section we consider a single ℓ only and drop the subscript. By linearising the energy dependence of ϕ, Eq. (8), it is possible to parameterise $P(\varepsilon)$ in a simple manner. First, we realise P is explicitly a function of D, and implicitly depends on ε through $D \equiv D\{\phi_\ell(\varepsilon)\}$. Writing Eq. (20) in terms of D we see that

$$P[D] = \frac{\hat{H}(s)}{\hat{J}(s)} \cdot \frac{D - D\{\hat{H}\}}{D - D\{\hat{J}\}} \tag{24}$$

The linearised $\phi(\varepsilon, r)$ (Eq. (8)) may be re-expressed using D in place of ε:

$$\Phi(D, r) \approx \phi(\varepsilon_\nu, r) + \omega[D]\dot{\phi}(\varepsilon_\nu, r) \tag{25}$$

where

$$\omega[D] = -\frac{\phi(\varepsilon_\nu, s)}{\dot{\phi}(\varepsilon_\nu, s)} \cdot \frac{(D - D\{\phi(\varepsilon_\nu)\})}{(D - D\{\dot{\phi}(\varepsilon_\nu)\})}$$

Eqs. (8) and (25) refer to the same object; one is parameterised by ε while the other is parameterised by D or $\omega[D]$. The latter is more convenient because $\omega[D]$ and $P[D]$ both have a simple pole structure. That each have this structure imply that their inverses $D[\omega]$ and $D[P]$ also have a simple pole structure. This further implies that if P is parameterised not by D but instead by $\omega[D]$, $P\{\omega[D]\}$ will also have a simple pole structure in ω, and depends on $\omega[D]$ as:

$$P\{\omega[D]\} = P\{\omega[D\{\dot{\phi}\}]\} \cdot \frac{(\omega[D] - \omega[D\{\hat{H}\}])}{(\omega[D] - \omega[D\{\hat{J}\}])}$$

This relation follows from the fact that $P\{\omega\}$ has a pole structure in ω, that P vanishes when $D = D\{\hat{H}\}$, and $1/P$ vanishes when $D = D\{\hat{J}\}$. The prefactor $P\{\omega[D\{\dot{\phi}(\varepsilon_\nu)\}]\}$ follows from the fact that when $\omega[D] \to \infty$, $D \to D\{\dot{\phi}\}$.

To obtain an explicit form for $P(\varepsilon)$ a relation between ε and ω is required. Matrix elements and overlap for $\Phi[D]$ are readily obtained from Schrödinger's equation, Eqs. (4) and (9). With these equations and normalisation relations Eq. (5) and (10), we can find that

$$\langle \Phi[D'] \mid -\nabla^2 + v - \varepsilon_\nu \mid \Phi[D] \rangle = \omega[D] \tag{26}$$
$$\langle \Phi[D'] \mid \Phi[D] \rangle = 1 + \langle \dot{\phi}^2 \rangle \omega[D']\omega[D]$$

$\varepsilon[D]$ can be obtained from the variational principle

$$\varepsilon[D] = \langle \Phi[D] \mid -\nabla^2 + v \mid \Phi[D] \rangle / \langle \Phi^2[D] \rangle$$
$$= \varepsilon_\nu + \omega[D]\{1 + \omega^2[D]\langle \dot{\phi}^2 \rangle\}^{-1} \tag{27}$$

Linearisation of ϕ or Φ has errors of second order in $\varepsilon - \varepsilon_\nu$, which means the variational estimate for the energy has errors of fourth order. From inspection of Eq. (27) we can deduce that the following linear approximation to ε

$$\widetilde{\varepsilon}[D] = \varepsilon_\nu + \omega[D] \tag{28}$$

has errors of second order in $\varepsilon - \varepsilon_\nu$. Thus, P can be parameterised to second order in ε as

$$P(\varepsilon) \approx \widetilde{P}(\varepsilon) = \frac{1}{\gamma}\frac{\varepsilon - C}{\varepsilon - V}$$

where the "potential parameters" γ, C and V are defined as

$$\gamma = \frac{1}{P[D\{\dot{\phi}\}]} = \frac{W\{\hat{J}, \dot{\phi}\}}{W\{\hat{H}, \dot{\phi}\}} \xrightarrow{\kappa\to 0} \frac{(s/w)^{2\ell+1}}{2(2\ell+1)} \frac{D\{\dot{\phi}\} - \ell}{D\{\dot{\phi}\} + \ell + 1} \tag{29}$$

$$C - \varepsilon_\nu = \omega[D\{\hat{H}\}] = -\frac{W\{\hat{H}, \phi\}}{W\{\hat{H}, \dot{\phi}\}} \xrightarrow{\kappa\to 0} -\frac{\phi(s)}{\dot{\phi}(s)} \frac{(D\{\phi\} + \ell + 1)}{(D\{\dot{\phi}\} + \ell + 1)} \tag{30}$$

$$V - \varepsilon_\nu = \omega[D\{\hat{J}\}] = -\frac{W\{\hat{J}, \phi\}}{W\{\hat{J}, \dot{\phi}\}} \xrightarrow{\kappa\to 0} -\frac{\phi(s)}{\dot{\phi}(s)} \frac{(D\{\phi\} - \ell)}{(D\{\dot{\phi}\} - \ell)} \tag{31}$$

A simple pole structure can be parameterised in several forms. A particularly useful one is

$$\widetilde{P}(\varepsilon) = \frac{(\varepsilon - C)}{\gamma(\varepsilon - C) + \Delta} = \left(\frac{\Delta}{\varepsilon - C} + \gamma\right)^{-1} \tag{32}$$

where $\Delta \equiv (C - V)\gamma$. Δ can be expressed in terms of Wronskians as

$$\sqrt{\Delta} = -\left(\frac{2}{w}\right)^{1/2} W\{\hat{J}, \phi\}$$
$$\xrightarrow{\kappa \to 0} -\left(\frac{w}{2}\right)^{1/2} \frac{\phi(s)}{(2\ell+1)} \frac{s^{\ell+1}}{w^{\ell+1}} [D\{\phi\} - \ell] \tag{33}$$

This last equation defines the sign of $\Delta^{1/2}$.

Eq. (32) is accurate only to second order because of the linear approximation Eq. (28) for $\omega(\varepsilon)$. From the structure of Eq. (27), it is clear that $\omega(\varepsilon)$, and thus $P(\varepsilon)$ can be more accurately parameterised (to third order) by the substitution

$$P(\varepsilon) = \widetilde{P}(\varepsilon') \tag{34}$$

$$\varepsilon' = \varepsilon + (\varepsilon - \varepsilon_\nu)^3 \langle \dot{\phi}^2(\varepsilon_\nu) \rangle \tag{35}$$

The third-order parameterisation requires another parameter, sometimes called the "small parameter"

$$p = \int_0^s \dot{\phi}^2(\varepsilon_\nu, r) r^2 dr = -\frac{\ddot{\phi}(s)}{3\phi(s)} \tag{36}$$

Thus P is parameterised to third order by four independent parameters. It is sometimes convenient in a Green's function context to use P and N (or equivalently $\dot{P}$; see Eq. (23)) in place of C and Δ. Green's functions can be constructed without linearising ϕ; this is the KKR-ASA method. How linearisation of ϕ resolves G in an energy-independent Hamiltonian is described in Section 2.8.3.

2.8.2. Spin orbit coupling as a perturbation

The formalism of the preceding sections can be extended to the Pauli Hamiltonian. If we include the spin–orbit coupling perturbatively, and include the mass–velocity and Darwin terms the Hamiltonian becomes

$$-\nabla^2 + v(r) - \frac{1}{c^2}(\varepsilon - v(r))^2 + \frac{\partial v(r)}{\partial r}\frac{\partial}{\partial r} + \xi(r)\hat{\mathbf{L}} \cdot \hat{\mathbf{S}}$$

where

$$\xi(r) = \frac{2}{c^2}\frac{dv(r)}{rdr}$$

and

$$\hat{\mathbf{L}} \cdot \hat{\mathbf{S}} = \frac{1}{2}\begin{pmatrix} L^z & L^{+-} \\ L^{-+} & -L^z \end{pmatrix}$$

The 2×2 matrix refers to spin space. In orbital space,

$$L^z = \delta_{mm'} m$$
$$L^{+-} = \delta_{m'(m+1)}\sqrt{(\ell + m + 1)(\ell - m)}$$
$$L^{-+} = \delta_{m'(m-1)}\sqrt{(\ell + m)(\ell - m + 1)}$$

$\hat{\mathbf{L}} \cdot \hat{\mathbf{S}}$ mixes spin components; also matrix elements of Eq. (25) and $\hat{\mathbf{L}} \cdot \hat{\mathbf{S}}$ depend on both ℓ and m. We require matrix elements of the $\Phi(D)$, analogous to Eq. (26):

$$\langle \Phi_{\ell' m'}[D'] \mid \xi \hat{\mathbf{L}} \cdot \hat{\mathbf{S}} \mid \Phi_{lm}[D] \rangle = \xi_\ell[D', D]\delta_{ll'}(lm|lm')$$

2.8.3. How the ASA-tail cancellation reduces to a linear algebraic eigenvalue problem

The eigenvalue condition is satisfied when ε is varied so that

$$|P - S| = 0$$

Table 1
Potential parameters in the original LMTO-ASA method. In the tight-binding transformation of this method these symbols become parameterised by α, defining the transformation; see Section 2.10.1.

	Name	Interpretation
C	Band centre	Eigenvalue of MT "atom" and resonance in extended system, Eq. (30)
Δ	Bandwidth	Bandwidth in the absence of hybridisation with other orbitals, Eq. (33)
p	Small parameter	3rd order correction to second-order potential function $\widetilde{P}$, Eq. (32)
V	"bottom"	ε where $D_\ell = +\ell$ (free electrons)
γ	"transformation to orthogonal basis"	See Section 2.9

$P - S$ is a matrix $(P - S)_{RL,R'L'}$ with P diagonal in RL, and S Hermitian. If P is parameterised by Eq. (32)

$$\left|\left(\frac{\Delta}{z - C} + \gamma\right)^{-1} - S\right| = 0$$

Multiply on the right by S^{-1} and the left by P^{-1}

$$\left|\frac{\Delta}{\varepsilon - C} + \gamma - S^{-1}\right| = 0$$

Rearrange, keeping in mind that C, Δ, and γ are real and diagonal in RL

$$\left|-\varepsilon + C + \sqrt{\Delta}\left(S^{-1} - \gamma\right)^{-1}\sqrt{\Delta}\right| = 0$$

This has the form of the linear algebraic eigenvalue problem

$$\widetilde{h}\psi = \varepsilon\psi$$

with

$$\widetilde{h} = C + \sqrt{\Delta}\left(S^{-1} - \gamma\right)^{-1}\sqrt{\Delta} \tag{37}$$

$\widetilde{h}$ is a Hermitian matrix, with eigenvalues corresponding to the zeros in $|\widetilde{P} - S|$. The tilde indicates that $\widetilde{h}$ is obtained from $\widetilde{P}$ and is thus accurate to second order in $\varepsilon - \varepsilon_\nu$ (C and Δ are calculated at ε_ν, Eqs. (30), (33)). Linear MTO's will be constructed in Section 2.13, and $\widetilde{h}$ can be identified with $h^\alpha - \varepsilon_\nu$, Eq. (61), for $\alpha = \gamma$ and if P is parameterised by $\widetilde{P}$.

2.9. "Screening" transformation to short-ranged, tight-binding basis

Formally, Eq. (37) is a "tight-binding" Hamiltonian in the sense that it is a Hamiltonian for a linear combination of atom-centred (augmented) envelopes χ, Eq. (14) taken at some fixed (linearisation) energy ε_ν. The Hamiltonian is long-ranged because $\hat{H}_\ell(r)$ is long ranged (see Eq. (B.14) for $\kappa = 0$) unless $E = -\kappa^2$ is -1 Ry or deeper. But such a basis is not accurate: the optimal E falls somewhere in the middle of the occupied part of the bands, and is roughly $+0.3$ Ry in close-packed systems; $E = 0$ is a compromise.

The idea behind the "screening" or "tight-binding" transformation is to keep $E = -\kappa^2$ near zero but render the Hilbert space of the $\chi_{RL}(\varepsilon, \mathbf{r})$ short range by rotating the basis into an equivalent set $\{\chi_{RL}(\varepsilon, \mathbf{r})\} \rightarrow \{\chi^\alpha_{RL}(\varepsilon, \mathbf{r})\}$, by particular linear combinations which render $\chi^\alpha_{RL}(\varepsilon, \mathbf{r})$ short ranged (or acquire another desirable property, e.g. be orthogonal), see Fig. 5.

Andersen sometimes called the change a "screening transformation" because it is analogous to screening in electrostatics. Note that $\hat{H}_\ell(E = 0, r) \propto 1/r$, has the same form as a charge monopole, with long range behaviour. It becomes short ranged if screened by opposite charges in the neighbourhood. The same applies to higher order multipoles: they can become short ranged in the presence of multipoles of opposite sign.

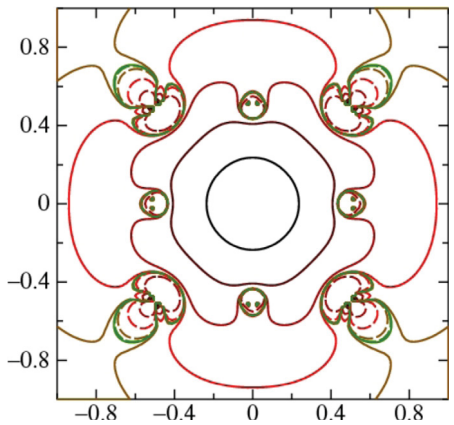

Fig. 5. Contour of a screened s envelope function in a bcc lattice. Each contour represents a reduction in amplitude by a factor of 10. Dashed lines show contours with negative amplitude. At the "hard core" radius (Section 2.10.5) on the head site a screened function has pure ℓ character; at the "hard core" radius on all the tail sites the one-centre expansion of the envelope function vanishes for $\ell < \ell_{\max}$.

The transformation can be accomplished in an elegant manner by admixing to the original ("bare") envelope (renamed from $\hat{H}_{RL}$ to $\hat{H}^0_{RL}$ to distinguish it from the "screened" one) with amounts of $\hat{H}^0_{R'L'}$ in the neighbourhood of R:

$$\hat{H}^\alpha_{RL}(\mathbf{r}) = \sum_{R'L'} \hat{H}^0_{R'L'}(\mathbf{r}) B^\alpha_{R'L';RL} \tag{38}$$

Whatever prescription determines $B^\alpha_{R'L';RL}$, it is evident that the Hilbert space is unchanged, and that $\hat{H}^\alpha_{RL} \to \hat{H}^0_{RL}$ if $B^\alpha_{R'L';RL} \to \delta_{R'L';RL}$.

The one-centre expansion of $\hat{H}^\alpha_{RL}$ in any channel $R'L'$ is some linear combination of Hankel and Bessel functions, because $\hat{H}^0_{RL}$ are Hankels in their head channel $R'L' = RL$, and linear combinations of Bessels in other channels $R'L' \neq RL$ (see Eq. (3)). Thus every $\hat{H}^0_{RL}$ is expanded in $R'L'$ by some particular linear combination of Bessel and Hankel functions

$$J^\alpha_{R'L'}(E,\mathbf{r}) = J^0_{R'L'}(E,\mathbf{r}) - \alpha_{R'L';RL} \hat{H}^0_{R'L'}(E,\mathbf{r}) \tag{39}$$

$B^\alpha_{R'L';RL}$ determines $\alpha_{R'L';RL}$, and vice-versa. α used as a superscript indicates that it determines the screening.

A simple and elegant way to choose the screening transformation is to expand *every* $\hat{H}^0_{RL}$ in channel $R'L' \neq RL$, by the *same* function $J^\alpha_\ell(\kappa, r_{R'}) Y_{L'}(\hat{\mathbf{r}}_{R'})$. Then α need only be specified by two indices, $\alpha_{R'L';RL} \to \alpha_{R'\ell'}$.

To sum up, we specify the transformation through $\alpha_{R'\ell'}$. Envelopes $\hat{H}^\alpha_{RL}$ are expanded at site $R'L' \neq RL$, as linear combinations

$$\hat{H}^\alpha_{RL}(\mathbf{r}) = -\sum_{R'L'} \hat{J}^\alpha_{R'L'}(\mathbf{r}) S^\alpha_{R'L';RL} \tag{40}$$

$$\hat{J}^\alpha_{R'L'}(\mathbf{r}) = [\hat{J}^0_{R'\ell'}(r) - \alpha_{R'\ell'} \hat{H}^0_{R'\ell'}(r)] Y_{R'L'}(\hat{\mathbf{r}}) \tag{41}$$

The structure constants $S^\alpha_{R'L';RL}$ are expansion coefficients that will be determined next, but already it should be evident that $S^\alpha \xrightarrow{\alpha \to 0} S^0$, where S^0 are the structure constants of Eq. (3).

In its own "head" $\hat{H}^\alpha_{RL}$ must have an additional irregular part. By expressing the $B^\alpha_{R'L';RL}$ in Eq. (38) in terms $S^\alpha_{R'L';RL}$ as

$$\hat{H}^\alpha_{RL}(\mathbf{r}) = \sum_{R'L'} \hat{H}^0_{R'L'}(\mathbf{r}) \left(\delta_{R'L';RL} + \alpha_{R'\ell'} S^\alpha_{R'L';RL}\right) \tag{42}$$

we can see indeed that $\hat{H}^\alpha_{RL}(\mathbf{r})$ has the required one-centre expansion, Eqs. (40), (41), provided that S^α obey a Dyson-like equation

$$S^\alpha = S^0 + S^0 \alpha S^\alpha \quad \text{or}$$
$$S^{\alpha\,-1} = S^{0\,-1} - \alpha \tag{43}$$

In practice S^α is calculated from

$$S^\alpha = \alpha^{-1} \left(\alpha^{-1} - S^0\right)^{-1} \alpha^{-1} - \alpha^{-1}$$

It follows immediately from Eq. (43) that if there are two screening representations α and β, the structure constants connecting them are related by

$$S^{\alpha\,-1} + \alpha = S^{\beta\,-1} + \beta \tag{44}$$

2.10. Screened muffin-tin orbitals and potential functions

In this section we develop a screened analogue of the MTO's, Eq. (14) potential functions, Eq. (20), normalisation Eq. (21), and tail cancellation conditions (16). Here we mostly concern ourselves with the ASA with $\kappa = 0$.

2.10.1. Redefinitions of symbols

We have defined a number of quantities in the context of the original MTO basis set, Eq. (14) that will have a corresponding definition in a screened basis set. Several previously defined quantities are now labelled with a superscript 0 to indicate that their definitions correspond to the unscreened $\alpha = 0$ representation: $\hat{H}_{RL} \equiv \hat{H}^0_{RL}$, $\hat{J}_{RL} \equiv \hat{J}^0_{RL}$, $P_{RL} \equiv P^0_{RL}$, $N_{RL} \equiv N^0_{RL}$ and $S_{R'L';RL} \equiv S^0_{R'L';RL}$.

The "potential parameters" C, Δ, and p, Eqs. (30)–(36) and Table 1, can also be relabelled with representation-dependent definitions. It is unfortunately rather confusing, but the original definitions without superscripts Eqs. (30)–(36) correspond not to C^0, Δ^0, and p^0, but to the particular screening representation $\alpha = \gamma$ (dubbed the "γ representation"). To be consistent with the new superscript convention for P^0 and N^0, the appropriate identifications are

$$\phi(\varepsilon_\nu) \equiv \phi^\gamma(\varepsilon_\nu), \quad \dot{\phi}(\varepsilon_\nu) \equiv \dot{\phi}^\gamma(\varepsilon_\nu), \quad \text{and}$$
$$C \equiv C^\gamma, \quad \Delta \equiv \Delta^\gamma, \quad p \equiv p^\gamma \tag{45}$$

Another unfortunate artefact of the evolution in LMTO formalism is that the meaning of many symbols changed over time. γ is called Q^{-1} in Ref. [32]. In the most recent NMTO formalism, S and B have exchanged meanings. Questaal's ASA codes use definitions that most closely resemble the "second generation" LMTO formalism perhaps most clearly expressed in Ref. [35]. Reference 16 of that paper makes correspondences to definitions laid out in earlier papers.

2.10.2. Potential and normalisation functions for screened MTO's

The MTO Eq. (14) is derived by augmenting the envelope $\hat{H}^0_{R\ell}(\varepsilon, r)$ by matching it smoothly onto a linear combination of $\phi_\ell(\varepsilon, r)$ and $\hat{J}^0_{R\ell}(r)$. For the screened case we match $\hat{H}^0$ to $\phi_\ell(\varepsilon, r)$ and $\hat{J}^\alpha_{R\ell}(r)$, the latter defined by Eq. (39): The matching requires

$$P^\alpha_\ell(\varepsilon) = \frac{W\{\hat{H}^0_\ell, \phi_\ell\}}{W\{\hat{J}^\alpha_\ell, \phi_\ell\}} = \frac{P^0_\ell(\varepsilon)}{1 - \alpha_\ell P^0_\ell(\varepsilon)} \tag{46}$$

$$N^\alpha_\ell(\varepsilon) = \frac{W\{\hat{J}^\alpha_\ell, \hat{H}^0_\ell\}}{W\{\hat{J}^\alpha_\ell, \phi_\ell\}} = \sqrt{\dot{P}^\alpha_\ell w/2} \tag{47}$$

Eq. (46) implies

$$[P^\alpha_\ell(\varepsilon)]^{-1} = [P^0_\ell(\varepsilon)]^{-1} - \alpha_\ell \to \frac{\Delta_\ell}{\varepsilon - C_\ell} + \gamma_\ell - \alpha_\ell \tag{48}$$

and is an obvious generalisation of P^0, Eq. (32). There is also the analogue of Eq. (44) for P:

$$[P^\alpha_\ell(\varepsilon)]^{-1} + \alpha_\ell = [P^\beta_\ell(\varepsilon)]^{-1} + \beta_\ell \tag{49}$$

Eq. (48) shows that $\partial[P^\alpha(\varepsilon)]^{-1}/\partial\varepsilon$ is independent of the screening α and

$$\begin{aligned}\left[-\frac{\partial}{\partial\varepsilon}[P^\alpha_\ell(\varepsilon)]^{-1}\right]^{-1/2} &= \frac{P^\alpha_\ell(\varepsilon)}{[\dot{P}^\alpha_\ell(\varepsilon)]^{1/2}} \\ &= -\sqrt{2/w}W\{\hat{H}^0_\ell, \phi_\ell\} \\ &\to \frac{\varepsilon - C_\ell}{\sqrt{\Delta_\ell}}\end{aligned} \tag{50}$$

The last forms of Eqs. (48), (50) apply when P is parameterised by $\widetilde{P}$, Eq. (32).

2.10.3. Screened muffin-tin orbitals

To define the analogue of the MTO, Eq. (14), in a screened representation we write

$$N^\alpha_{R\ell}(\varepsilon)\chi^\alpha_{RL}(\varepsilon, \mathbf{r}) = \begin{cases} N^\alpha_{R\ell}(\varepsilon)\phi_{RL}(\varepsilon, \mathbf{r}) + \sum_{R'L'} \bar{\hat{J}}^\alpha_{R'L'}(\mathbf{r})\,[P^\alpha - S^\alpha]_{R'L';RL} & \text{if } \mathbf{r} \in \{s_R\} \\ \hat{H}^\alpha_{RL}(\kappa, \mathbf{r}) & \text{if } \mathbf{r} \in \text{interstitial} \end{cases} \tag{51}$$

The partial wave $\phi_{RL}(\varepsilon, \mathbf{r})$ is understood to vanish outside its own head sphere, and P is a matrix diagonal in RL: $P^\alpha_{R'L';RL}(\varepsilon) = P^\alpha_{R'L'}(\varepsilon)\,\delta_{R'L';RL}$.

$\bar{\hat{J}}^\alpha_{R'L'}$ is the linear combination of ϕ and $\dot{\phi}$ that matches continuously and differentiably $\hat{J}^\alpha_{R'L'}$ defined in Eq. (39). Eq. (51) uses it instead of $\hat{J}^\alpha$ because when we later construct energy-independent MTO's we can generate basis sets that accurately solve Schrödinger's equation in the augmentation spheres. Since P^α and N^α depend only on values and slopes at the $\{s_R\}$, the substitution has no effect on them.

2.10.4. Tail cancellation in the tight-binding representation

The energy-dependent χ^α_{RL}, Eq. (51) exactly solve the ASA-MT potential because the trial function

$$\Psi(\varepsilon, \mathbf{r}) = \sum_{RL} z^\alpha_{RL}\chi^\alpha_{RL}(\varepsilon, \mathbf{r})$$

that satisfy the set of linear equations

$$\sum_{R'L'}(P^\alpha - S^\alpha)_{RL;R'L'}z^\alpha_{R'L'} = 0 \tag{52}$$

and the normalisation

$$\sum_{RL}|z^\alpha_{RL}|^2 = 1 \tag{53}$$

simplifies to

$$\Psi(\varepsilon, \mathbf{r}) = \sum_{RL} N^\alpha_{R\ell}\phi_{RL}(\varepsilon, \mathbf{r})$$

which is a normalised solution to the SE for $v(\mathbf{r}) = \sum_R v_R(r)$.

In practice the solution is inexact because L summations are truncated. The solution is rapidly convergent in the L-cutoff, however; see Ref. [35] for an analysis.

2.10.5. Hard core radius

α can be physically interpreted as equivalent to specifying a "hard core" radius where the one-centre expansion $\hat{H}^\alpha_{RL}$ vanishes in a sphere centred at R'. This is evident from the one-centre expansion Eq. (40) and the form of $\hat{J}^\alpha_\ell$, Eq. (41). $\hat{J}^\alpha_\ell$ vanishes at the radius where

$$\alpha_{R\ell} = \hat{J}^0_\ell(E, r_R)/\hat{H}^0_\ell(E, r_R)$$

In his more recent developments, Andersen defined the screening in terms of the hard core radius a_ℓ instead of α, because nearly short-ranged basis functions can be obtained for a fixed $a_{R\ell} = 0.7s_R$, independent of κ and ℓ.

It is easy to see how such a transformation can render envelope functions short-ranged. The value of $\hat{H}^\alpha_{R\ell}$ is forced to be zero in a sea of $R'L'$ channels surrounding it. Provided the $a_{R\ell}$ are suitably adjusted, it quickly drives $\hat{H}^\alpha_{R\ell}(r) \to 0$ everywhere for increasing r. If the $a_{R\ell} \to 0$, the screening vanishes and $\hat{H}^\alpha_{R\ell}$ returns to the long-ranged $\hat{H}^0_{R\ell}$; while if the $a_{R\ell}$ becomes comparable to s_R the value on the head must be something like 0 and 1 at the same time (heads and tails meet). The damping is too large and the $\hat{H}^\alpha_{R\ell}(r)$ "rings" with increasing r. For $a_{R\ell} = 0.7r_s$ or thereabouts the ringing is damped and $\hat{H}^\alpha_{R\ell}(r)$ decays exponentially with r, even for $\kappa = 0$.

2.11. MTO's and second order Green's function

Through the eigenvectors z^α_{RL} of Section 2.10.4, we can construct the Green's function. This was done in Appendix A of Ref. [32], where the full Green's function, including the irregular parts, are derived. Here we will adopt a simpler development along the lines of Ref. [5], after linearising the $\chi^\alpha_{RL}(\varepsilon, \mathbf{r})$.

One way to see why $P^\alpha - S^\alpha$ have similar eigenvalues for any α is to note that $[P^\alpha]^{-1} - [S^\alpha]^{-1}$ does not depend on α, since $[P^\alpha]^{-1}$ and $[S^\alpha]^{-1}$ are shifted by the same amount (compare Eqs. (43) and (48)). In scattering theory $[P^0_\ell(\varepsilon)]^{-1}$ is proportional to tangent of the phase shift, and we realise that the transformation $(P^0, S^0) \to (P^\alpha, S^\alpha)$ corresponds merely to a shift of the scattering background. The pole structures of $P^\alpha - S^\alpha$ can depend on α because of the irregular parts: $P^\alpha - S^\alpha$ have the same poles as $[P^\alpha]^{-1} - [S^\alpha]^{-1}$ only where P^α and S^α have no zeros or poles. Some care must be taken when generating the Green's function G.

The MTOs Eq. (51) form a complete Hilbert space for any α, but α can be chosen to satisfy varying physical requirements. To make short-ranged Hamiltonians it has been found empirically that the following universal choice

$$\alpha_s = 0.34857, \qquad \alpha_p = 0.05303, \qquad \alpha_d = 0.010714,$$
$$\alpha_\ell = 0 \text{ for } \ell > 2$$

yields short-ranged basis functions $\chi^\alpha_{RL}(\mathbf{r})$ for $\kappa = 0$, for any reasonably close-packed system.

Another choice is $\alpha = \gamma$. In Section 2.8.3 it was shown how the tail cancellation condition had the same eigenvalues as a fixed Hamiltonian, Eq. (37). We are now equipped make a connection with the χ^γ basis and Eq. (37). Moreover, this connection enables us to construct the second order Green's function. First, Eq. (43) enables us to recognise the quantity $\left(S^{-1} - \gamma\right)^{-1}$ as S^γ. Eq. (37) then has the simple two-centre form $\widetilde{h} = C + \Delta^{1/2}\,S^\gamma\,\Delta^{1/2}$

The Green's function corresponding to some fixed h has the simple form

$$\widetilde{G}(z) = [z - \widetilde{h} - i0^+]^{-1}$$

for complex energy z. It is easy to see that $\widetilde{G}(z)$ can be expressed in the following form:

$$\widetilde{G}(z) = \Delta^{-1/2}\,\widetilde{g}^\gamma\,\Delta^{-1/2} \tag{54}$$

where

$$\widetilde{g}^\gamma = [\widetilde{P}^\gamma - S^\gamma]^{-1} \tag{55}$$

This is the analogue of Eq. (17) in the γ representation.

$\widetilde{P}^\gamma(z)-S^\gamma$ has a direct connection with $z-\widetilde{h}$ because of $\widetilde{P}^\gamma(z)$ takes the simple form $(z-C)/\Delta$. $\widetilde{P}^\gamma - S^\gamma$ is linear in z since S^γ is independent of it, and $\sqrt{\Delta}(\widetilde{P}^\gamma - S^\gamma)\sqrt{\Delta} = z - \widetilde{h}$.

2.11.1. Scattering path operator in other representations

To build $G(\varepsilon)$ from general screening representations β we need to transform the scattering path operator $g^\gamma \rightarrow g^\beta$. This can be accomplished [35] using Eqs. (49) (44). The result is

$$g^\beta = (P^\alpha/P^\beta)g^\alpha(P^\alpha/P^\beta) + (\beta-\alpha)(P^\alpha/P^\beta)$$

These transformations require only Eqs. (49) and (44); they do not depend on parameterisation of P.

Since Eq. (50) is representation-independent, (P^α/P^β) can equally be written in the following forms:

$$\frac{P^\alpha(\varepsilon)}{P^\beta(\varepsilon)} = \left(\frac{\dot{P}^\alpha}{\dot{P}^\beta}\right)^{1/2} = 1+(\alpha-\beta)P^\alpha(\varepsilon) \tag{56}$$

2.12. The ASA Green's function, general representation

As shown in Section 2.11 the relation between the Green's function $G(E)$ and the scattering path operator g is particularly simple when potential functions are parameterised to second order. The relation between the ASA approximation to G and g can be written more generally as follows. In the ASA, every point $\mathbf{r}$ belongs to some sphere R with partial waves $\phi_{RL}(\varepsilon,\mathbf{r})$ so that

$$G_{RR'}(\varepsilon,\mathbf{r},\mathbf{r}') = \sum_{LL'} \phi_{RL}(\varepsilon,\mathbf{r})\, G_{RLR'L'}(\varepsilon)\, \phi_{R'L'}(\varepsilon,\mathbf{r}')$$

where

$$G_{RLR'L'}(\varepsilon) = -\frac{1}{2}\frac{d\ln \dot{P}^\alpha_{RL}(\varepsilon)}{d\varepsilon} + \sqrt{\dot{P}^\alpha_{RL}(\varepsilon)}\, g^\alpha_{RLR'L'}(\varepsilon) \sqrt{\dot{P}^\alpha_{R'L'}(\varepsilon)} \tag{57}$$

$\mathbf{r}$ and $\mathbf{r}'$ (or R and R') are the field and source points, respectively. It was first shown in Ref. [32], for the "bare" representation $\alpha = 0$, and for a screened representation α in Ref. [35]. The first term cancels a pole appearing in the second term, connected to the irregular part of G (which we do not consider here).

$P^\alpha(\varepsilon)$ can be computed by integration of the radial Schrödinger equation for any ε. If this is done, and the structure constants S are taken as energy-dependent, this is the screened KKR method. Questaal's `lmgf` and `lmpg` parameterise $P^\alpha(\varepsilon)$, to second order (Eq. (48)), or to third order (Section 2.13.4) in ε.

Note that G does not depend on choice of α; it can be used as a stringent test of the correctness of the implementation.

2.13. The ASA Hamiltonian: Linearisation of the muffin-tin orbitals

The energy-dependent MTO, Eq. (51), exactly solves the ASA potential for a fixed ε. To make a fixed, energy-independent basis set, we constrain the energy-dependent MTO, Eq. (51), to be independent of ε. As we saw in Sections 2.8.3 and 2.11, there is a simple energy-independent Hamiltonian Eq. (37) that has the same eigenvalue spectrum as the second order $\widetilde{G}$; this is $\chi^\gamma_{RL}(\varepsilon_\nu,\mathbf{r})$.

We can construct an energy-independent basis $\chi^\alpha_{RL}(\mathbf{r}) = \chi^\alpha_{RL}(\varepsilon_\nu,\mathbf{r})$ for any α, by choosing the normalisation $N^\alpha_{R\ell}(\varepsilon)$ in such a way that $\partial\chi^\alpha_{RL}(\varepsilon,\mathbf{r})/\partial\varepsilon = 0$ at $\varepsilon=\varepsilon_\nu$. Thus we require

$$\frac{\partial}{\partial\varepsilon}\left[N^\alpha(\varepsilon)\phi(\varepsilon,r) + \bar{\dot{J}}^\alpha(r)P^\alpha(\varepsilon)\right]_{\varepsilon=\varepsilon_\nu} = 0$$

$$\dot{N}^\alpha(\varepsilon)\phi(\varepsilon,r) + N^\alpha(\varepsilon)\dot{\phi}(\varepsilon,r) + \bar{\dot{J}}^\alpha(r)\dot{P}^\alpha(\varepsilon)_{\varepsilon=\varepsilon_\nu} = 0$$

Define

$$\phi^\alpha(\varepsilon,r) \equiv [N^\alpha(\varepsilon)/N^\alpha(\varepsilon_\nu)]\phi(\varepsilon,r) \tag{58}$$

Then the condition that $\dot{\chi}^\alpha_{RL}(\mathbf{r})$ vanish for all $\mathbf{r}$ becomes

$$\bar{\dot{J}}^\alpha(r) = -[\dot{\phi}^\alpha(\varepsilon,r)N^\alpha(\varepsilon)/\dot{P}^\alpha(\varepsilon)]_{\varepsilon=\varepsilon_\nu} \tag{59}$$

Henceforth, when the energy index is suppressed it means that the energy-dependent function or parameter is to be taken at the linearisation energy ε_ν. If $\bar{J}$ is replaced by Eq. (59), Eq. (51) becomes

$$\chi^\alpha_{RL}(\mathbf{r}) = \begin{cases} \phi_{RL}(\mathbf{r}) + \sum_{R'L'} \dot{\phi}^\alpha_{R'L'}(\mathbf{r}) h^\alpha_{R'L';RL} & \text{if } \mathbf{r}\in\{s_R\} \\ (N^\alpha_{R\ell})^{-1}\hat{H}^\alpha_{RL}(\kappa,\mathbf{r}) & \text{if } \mathbf{r}\in \text{interstitial} \end{cases} \tag{60}$$

$$h^\alpha = -P^\alpha(\dot{P}^\alpha)^{-1} + [\dot{P}^\alpha]^{-1/2} S^\alpha [\dot{P}^\alpha]^{-1/2} \tag{61}$$

The Hilbert space of the $\{\chi^\alpha_{RL}(\mathbf{r})\}$ consists of the pair of functions $\phi_{R'L'}$ and $\dot{\phi}_{R'L'}$ inside all augmentation channels $R'L'$. Changing α merely rotates the Hilbert space, modifying how much $\phi_{R'L'}$ and $\dot{\phi}_{R'L'}$ each χ^α_{RL} contains.

Eq. (60) may be regarded as a Taylor series of $\chi^\alpha_{RL}(\varepsilon,\mathbf{r})$, Eq. (51), to first order in $\varepsilon-\varepsilon_\nu$, with $\dot{\phi}^\alpha(r)\,h^\alpha$ playing the part of $\dot{\chi}^\alpha(r)(\varepsilon-\varepsilon_\nu)$. The eigenvalues of h^α are fact the eigenvalues of $\chi^\alpha_{RL}(\varepsilon,\mathbf{r})$ to first order in $\varepsilon-\varepsilon_\nu$. If P is parameterised to second order, $P(\varepsilon) \rightarrow \widetilde{P}(\varepsilon)$, h^α in the γ representation becomes the second order ASA Hamiltonian, Eq. (37), when $\alpha=\gamma$.

The relation between $N^\alpha(\varepsilon)$ and $P^\alpha(\varepsilon)$ was already established in Eq. (47). To confirm it is consistent with Eq. (59) at ε_ν, revisit the definition of N^α using Eq. (59)

$$N^\alpha(\varepsilon) = \frac{W\{J^\alpha,\hat{H}^0\}}{W\{J^\alpha,\phi\}} = \frac{w/2}{W\{-\dot{\phi}^\alpha N^\alpha(\varepsilon)/\dot{P}^\alpha(\varepsilon),\phi\}}$$
$$= \frac{\dot{P}^\alpha(\varepsilon)w/2}{N^\alpha(\varepsilon)W\{\dot{\phi}^\alpha,\phi\}} = \frac{\dot{P}^\alpha(\varepsilon)w/2}{N^\alpha(\varepsilon)}$$

which confirms Eq. (47).

2.13.1. Potential parameters C^α, Δ^α, and o^α

The LMTO literature suffers from an unfortunate proliferation of symbols, which can be confusing. Nevertheless we introduce yet another group because they offer simple interpretations of what is happening as the basis changes with representation α, and also to make a connection with the Green's function. It is helpful to remember there is a single potential function $P^\alpha(\varepsilon)$, which determines the normalisation $N^\alpha(\varepsilon)$ though Eq. (47).

$P^\alpha(\varepsilon)$ can be parameterised to second order with three independent parameters C, Δ, and γ (Eq. (48)) and to third order with the "small parameter" p (Eq. (36)).

The energy derivative of $\phi^\alpha(\varepsilon,r)$ at ε_ν is

$$\dot{\phi}^\alpha_\ell(\varepsilon_\nu,r) = \dot{\phi}^\alpha_\ell(r) = \dot{\phi}_\ell(r) + o^\alpha_\ell\,\phi(r) \tag{62}$$

$$o^\alpha_\ell \equiv \dot{N}^\alpha_\ell/N^\alpha_\ell \rightarrow \frac{\alpha-\gamma}{(\alpha-\gamma)(C-\varepsilon_\nu)+\Delta} \tag{63}$$

The last form applies when P is parameterised to second order. We have introduced the "overlap" potential parameter o^α_ℓ. It vanishes in the γ representation and consequently $\dot{\phi}^\gamma = \dot{\phi}$. This fact provides a simple interpretation of $\chi^\alpha_{RL}(\mathbf{r})$, Eq. (60) in the γ representation. χ^γ_{RL} acquires pure ϕ character for its own head, and pure $\dot{\phi}$ in spheres where $R'\neq R$. This implies that the $\{\chi^\gamma\}$ basis are orthogonal apart from interstitial contributions (neglected in the ASA) and small terms proportional to p (Eq. (36)). ϕ and $\dot{\phi}$ combine in every sphere in the exact proportion Eq. (8) at each eigenvalue ε_i of h^γ. Thus eigenvalues of h^γ are correct to one order in $\varepsilon_i-\varepsilon_\nu$ higher than $h^{\alpha\neq\gamma}$.

In the LMTO literature two other parameters are introduced to characterise h^α in a suggestive form:

$$h^{\alpha}_{RL;R'L'} = (C^{\alpha} - \varepsilon_v)_{R\ell}\,\delta_{RL;R'L'} + \Delta^{\alpha}_{R\ell} S^{\alpha}_{RL;R'L'} \Delta^{\alpha}_{R'\ell'} \tag{64}$$

where

$$C^{\alpha}_{\ell} - \varepsilon_v \equiv -P^{\alpha}_{\ell}(\dot{P}^{\alpha}_{\ell})^{-1} \rightarrow (C - \varepsilon_v)\left[1 + \frac{(C-\varepsilon_v)(\alpha-\gamma)}{\Delta}\right] \tag{65}$$

$$\sqrt{\Delta^{\alpha}} \equiv [\dot{P}^{\alpha}]^{-1/2} \rightarrow \sqrt{\Delta}\left[1 + \frac{(C-\varepsilon_v)(\alpha-\gamma)}{\Delta}\right] \tag{66}$$

$h^{\alpha} + \varepsilon_v$ becomes $\widetilde{h}$ (Eq. (37)) when $\alpha = \gamma$.

2.13.2. How h^{α} changes with representation

From Eq. (63) implies that $\dot{\phi}^{\alpha}$ transforms as

$$\dot{\phi}^{\alpha} - o^{\alpha}\phi = \dot{\phi}^{\beta} - o^{\beta}\phi$$

Dividing χ^{α} in to ϕ and $\dot{\phi}$ parts, we realise that $o^{\alpha} + [h^{\alpha}]^{-1}$ is independent of representation and therefore

$$o^{\beta} + [h^{\beta}]^{-1} = o^{\alpha} + [h^{\alpha}]^{-1}$$

2.13.3. ASA Hamiltonian and overlap matrix

χ^{α} (Eq. (60)) is an energy-independent basis set and has an eigenvalue spectrum. Within the ASA (Section 2.7) matrix elements of the Hamiltonian and overlap are readily obtained. Using normalisation Eqs. (5) and (36), Schrödinger's equation for partial waves Eq. (4), the parameterisation of $\dot{\phi}^{\alpha}$ Eq. (62), and neglecting the interstitial parts

$$\begin{aligned} H &= \langle \chi^{\alpha} | -\nabla^2 + v | \chi^{\alpha} \rangle \\ &= h^{\alpha}(1 + o^{\alpha}h^{\alpha}) + (1 + h^{\alpha}o^{\alpha})\varepsilon_v(1 + o^{\alpha}h^{\alpha}) + h^{\alpha}\,\varepsilon_v\,p\,h^{\alpha} \\ O &= \langle \chi^{\alpha} | \chi^{\alpha} \rangle \\ &= (1 + h^{\alpha}o^{\alpha})(1 + o^{\alpha}h^{\alpha}) + h^{\alpha}\,p\,h^{\alpha} \end{aligned} \tag{67}$$

2.13.4. Third order Green's function

$\widetilde{G}(z)$ depends on potential parameters C and Δ. These can be replaced with the following:

$$\Delta = [\dot{\widetilde{P}}^{\gamma}(\varepsilon)]^{-1} \quad \text{and} \quad \varepsilon - C = \frac{\widetilde{P}^{\gamma}(\varepsilon)}{\dot{\widetilde{P}}^{\gamma}(\varepsilon)} \tag{68}$$

which follows from Eq. (50) with $\alpha = \gamma$. Then the substitution $\widetilde{P} \rightarrow P$ through the replacement $\varepsilon \rightarrow \varepsilon' = \varepsilon + (\varepsilon - \varepsilon_v)^3 p$, Eqs. (35) and (36). This yields an expression for G to third order—more accurate than the 2nd order $\widetilde{G}$ [32,35]. Questaal codes `lmgf` and `lmpg` permit either second or third order parameterisation.

2.14. Principal layer Green's functions

Questaal has another implementation of ASA-Green's function theory designed mainly for transport. `lmpg` is similar in most respects to the crystal package `lmgf`, except that is written as a principal-layer technique. `lmpg` has a 'special direction', which defines the layer geometry, and for which G is generated in real space. In the other two directions, Bloch sums are taken in the usual way; thus for each q in the parallel directions, the Hamiltonian becomes one-dimensional and is thus amenable to solution in order-N time in the number of layers N.

The first account of this method was presented in Ref. [36], and the formalism is described in detail in Ref. [14], including its implementation of the non-equilibrium case. Here we summarise the basic idea and the main features.

`lmpg` is similar in many respects to `lmgf` except for its management of the layer geometry. The material consists of an active, or embedded region, which is cladded on the left and right by left and right semi-infinite leads.

$$\overbrace{\ldots \text{ PL } -1}^{\text{PLATL}} \;\Big|\; \overbrace{\text{PL } 0 \mid \ldots \mid \text{PL } n{-}1}^{\text{PLAT}} \;\Big|\; \overbrace{\text{PL } n \ldots}^{\text{PLATR}}$$

The end regions are half-crystals with infinitely repeating layers in one direction. All three regions are partitioned into slices, or principal layers (PL), along the 'special direction'. The left- and right-end regions consist of a single PL, denoted -1 and n, which repeat to $\mp\infty$. Thus the trilayer geometry is defined by five lattice vectors: two defining the plane normal to the interface (the potential is periodic in those vectors); one vector `PLAT` for the active region and one each (`PLATL` and `PLATR`) defining the periodically repeating end regions.

Far from the interface the potential is periodic and states are Bloch states. It is assumed that the potential in each end layer is the same as the bulk crystal (apart from a constant shift) and repeats periodically in lattice vectors `PLATL` and (`PLATR`) to $\mp\infty$.

Partitioning into PL is done because $\varepsilon - H = G^{-1}$ is short-ranged. It is requirement that a PL is thick enough so that H only connects adjacent PL. Then H is tridiagonal in the PL representation and the work needed to construct G scales linearly with the number of PL. Moreover it is possible in this framework to construct G for the end regions without using Bloch's theorem.

Principal layers are defined by the user; they should be chosen so that each PL is thick enough so that H connects to only nearest-neighbour PL on either side. (Utility `lmscell` has a facility to partition the active region into PL automatically.)

2.14.1. Green's function for the trilayer

`lmpg` constructs the auxiliary g, and if needed builds G from g by scaling (Section 2.12). In many instances g is sufficient (e.g. to calculate transmission and reflection probabilities [14]), although G is needed to make the charge density. Note there is a g (or G) connecting every layer to every other one; thus g has two layer indices, g_{ij} (i and j refer to PL here).

$g = (P - S)^{-1}$ for the entire trilayer can be constructed in one of two ways. The first is a difference-equation method, described in Ref. [36]. The second is simply to invert $(P - S)$ using sparse matrix techniques. Both methods require as starting points the diagonal element $g^{s,L}_{-1-1}$ for semi-infinite system (consisting of all layers between ∞ and -1, with vacuum for all layers to the right of the L- region) and the corresponding $g^{s,R}_{nn}$ for the R- region.

The sparse-matrix method is simple to describe. Supposing the active region is considered in isolation; denote it as I. Then $g_I^{-1} = (P - S)_I$. The effect of the leads is to modify g_I^{-1} by adding a self-energy to layers 0 and n:

$$\begin{aligned} g^{-1} &= (P - S)_I + \Sigma_0 + \Sigma_n \\ \Sigma_0 &= S_{0,-1}\, g^{s,L}_{-1-1}\, S_{-1,0} \\ \Sigma_n &= S_{n-1,n}\, g^{s,R}_{nn}\, S_{n-1,n} \end{aligned}$$

g^{-1} is inverted by a sparse matrix technique.

`lmpg` implements both the difference-equation and sparse-matrix techniques: both scale linearly with the number of layers, in memory and in time. It has been found empirically that they execute at similar speed for small systems, while the difference-equation method is significantly faster for large systems.

2.14.2. Green's functions for the end regions

To make g, the diagonal surface Green's functions $g^{s,L}_{-1-1}$ and $g^{s,R}_{nn}$ are required. `lmpg` implements two schemes to find them: a "decimation" technique [37] and a special-purpose difference-equation technique applicable for a periodic potential [38].

The latter method requires solution of a quadratic algebraic eigenvalue problem, which yields eigenvalues r: they correspond physically to wave numbers as $r = e^{ika}$. a is the thickness of the PL and k the wave number in the plane normal to the interface. k is in general complex since no boundary conditions are imposed; it is real only for propagating states. Eigenvalues occur in pairs, r_1

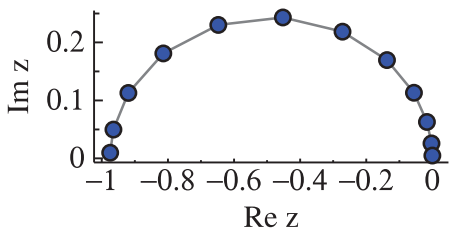

Fig. 6. Representative contour integration for occupied states. Twelve points are taken for the interval $(-1, 0)$. The contour deformation has eccentricity $= 1/2$, and a bunching parameter 1/2, giving more weight to the points near $E_F = 0$.

and r_2, and in the absence of spin–orbit coupling, $r_1 = 1/r_2^*$. There is a boundary condition on the end leads for the trilayer, which excludes states that grow into an end region. The surface Green's function g^s can be constructed from the same eigenvectors that make the "bulk" g [38].

A great advantage of this method is that its solution provides the eigenfunctions of the system; thus the Green's function can be resolved into normal modes. A large drawback is the practical problem of finding a solution to the eigenvalue problem. It can be converted into a linear algebraic eigenvalue equation of twice the rank; however, the resulting secular matrix can be nearly singular (especially if S is short-ranged). Also, the pairs r_1/r_2 can range over a very large excursion, of unity for propagating states and many orders of magnitude for rapidly decaying ones. Capturing them by solving a single eigenvalue problem imposes severe challenges on the eigenvalue solver.

Decimation is recursive and generally efficient; however problems can appear at special values of k energy where the growing and decaying pair r_1 and r_2 become very close to unity. Unfortunately, those "hot spots" are often the physically interesting ones.

At present Questaal's standard distribution does not have a fully satisfactory, all-purpose method to determine g^s, though one has been developed and will be reported in a future work.

2.15. Contour integration over occupied states

Many properties of interest involve integration over the occupied states. In contrast to band methods which give eigenfunctions for the entire energy spectrum at once, Green's functions are solved at a particular energy, and must be numerically integrated on an energy mesh for integrated properties. The spectral function or density of states is related to G as

$$A(\varepsilon) = \pi^{-1}|\mathrm{Im}\, G(\varepsilon)| \tag{69}$$

and in the noninteracting case is comprised of a superposition of δ-functions at the energy levels. Thus $G(\varepsilon)$ has lots of structure on the real axis, which makes integration along it difficult. However, since below the Fermi level G should have no poles in the upper half of the complex plane, the Cauchy theorem can be used to deform the integral from the real axis to a path in the complex plane (Fig. 6). `lmgf` and `lmpg` use an elliptical path, with upper and lower bounds on the real axis respectively at the Fermi level and some energy below the bottom of the band.

A Legendre quadrature is used, but the weights can be staggered to bunch points near E_F where G has lots of structure. Thus five parameters define the mesh: the number of points, the upper and lower bounds, the eccentricity of the ellipse (between 0 for circle and 1 for a line on the real axis) and bunching parameter which also ranges between 0 and 1. The integrand on the contour is smooth, except near the endpoint $z \to E_F$. Good results can be obtained with a modest number of points, typically 12–20.

The exact potential function P^α has no poles in this half plane; nor does the second order parameterisation but spurious poles may appear in the third order parameterisation. These may be avoided by working in the orthogonal representation ($\alpha = \gamma$) and/or by choosing fairly large elliptical eccentricities for the contour.

2.16. Spin–orbit coupling in the Green's function

It has been shown in Section 2.8.2 that spin–orbit coupling (SOC) can be added perturbatively to the Hamiltonian, resulting in the matrix elements containing $\xi_\ell[D', D]$. These matrix elements are added to the right-hand-side of the first line in Eq. (26), while the overlap integrals (second line in Eq. (26)) remain unchanged. In order to construct the Green's function with SOC, the resulting modification of the variational energy in Eq. (27) needs to be reformulated as a perturbation of the potential parameters.

Because the SOC operator is a matrix (see Section 2.8.2), the exact solutions $\phi_{\nu lj\kappa}(\mathbf{r})$ of the radial Pauli equation are linear combinations of spherical waves $|lm\sigma\rangle$ and $|lm'\sigma'\rangle$ with $m+\sigma = m'+\sigma' = j$. However, our perturbative treatment is still based on basis functions with definite spin that are calculated without SOC. The energy dependence, however, is modified by allowing the ω parameter to become a matrix, so that Eq. (25) is replaced by

$$\Phi_{m\sigma}(D_\uparrow, D_\downarrow, \mathbf{r}) = \phi_{m\sigma}(\epsilon_{\nu\sigma}, \mathbf{r}) + \sum_{m'\sigma'} \omega_{m\sigma,m'\sigma'}(D_\uparrow, D_\downarrow)\dot{\phi}_{m'\sigma'}(\epsilon_{\nu\sigma'}, \mathbf{r}) \tag{70}$$

where we dropped the common ℓ superscript because the SOC operator is diagonal in ℓ. The summation in (70) involves at most two terms with $m+\sigma = m'+\sigma'$.

Instead of the simple variational estimate of $\epsilon(D)$ in Eq. (27), we now construct a generalised eigenvalue equation, which leads to

$$\hat{\omega} + \hat{V}_{\mathrm{SO}} = \epsilon\hat{1} - \hat{\epsilon}_\nu + \hat{\omega}^\dagger\hat{p}(\epsilon\hat{1} - \hat{\epsilon}_\nu)\hat{\omega} \tag{71}$$

where $\hat{\omega}$ is the matrix from Eq. (70), both $\hat{\omega}$ and $\hat{V}_{\mathrm{SO}}$ are functions of $D_\uparrow$ and $D_\downarrow$, and $\hat{p}$ and $\hat{\epsilon}_\nu$ are diagonal matrices with elements $\langle\dot{\phi}^2_{\nu\ell\sigma}\rangle$ and $\epsilon_{\nu\sigma}$, respectively. The matrices are assumed to include the full basis set on the given site, i.e., their dimension is $2(2\ell+1)$.

The $\hat{\omega}$ matrix is found by solving Eq. (71). To first order in $\epsilon - \epsilon_\nu$, it gives $\hat{\omega} = \epsilon\hat{1} - \hat{\epsilon}_\nu - \hat{V}_{\mathrm{SO}}$, so that the matrix elements of $\hat{V}_{\mathrm{SO}}$ are effectively added to $\hat{\epsilon}_\nu$. Promoting the potential function $P(\epsilon)$ to a matrix and using the representation Eq. (32) and the definitions Eqs. (29)–(31), we find that the parameters Δ and γ are unaffected by $\hat{V}_{\mathrm{SO}}$ while C is promoted as $C \to \hat{C} = C\hat{1} + \hat{V}_{\mathrm{SO}}$. However, in order to make the definition of $\hat{P}(\epsilon)$ unambiguous, we need to fix the correct order of matrix multiplication. We also need to ensure that the poles of $G(z)$ have unit residues. We use the following definitions:

$$G(\epsilon) = \lambda(\epsilon) + \mu_L(\epsilon)[P(\epsilon) - S]^{-1}\mu_R(\epsilon), \tag{72}$$

$$P(\epsilon) = \left[\gamma + \sqrt{\Delta}(\epsilon - \hat{C})^{-1}\sqrt{\Delta}\right]^{-1} \tag{73}$$

$$\mu_L(\epsilon) = (1 - \hat{\dot{V}}_{\mathrm{SO}})^{1/2}[\Delta + \gamma(\epsilon - \hat{C})]^{-1}\sqrt{\Delta} \tag{74}$$

$$\mu_R(\epsilon) = \sqrt{\Delta}[\Delta + (\epsilon - \hat{C})\gamma]^{-1}(1 - \hat{\dot{V}}_{\mathrm{SO}})^{1/2} \tag{75}$$

$$\lambda = -\frac{1}{2}\mu_R^{-1}\ddot{P}\mu_L^{-1} \tag{76}$$

where $\hat{\dot{V}}_{\mathrm{SO}} = d\hat{V}_{\mathrm{SO}}/d\epsilon$ comes from the energy dependence of the SOC parameters $\xi_\ell[D', D]$. Note that $\dot{P} = \mu_R\mu_L$, and the structure of $G(z)$ guarantees that the poles of $G(z)$ have unit residues. It is straightforward to check that, with energy-independent $\hat{V}_{\mathrm{SO}}$, $G(z)$

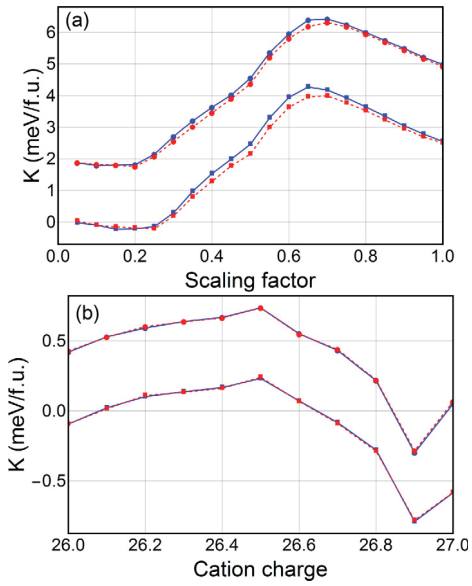

Fig. 7. Benchmarks for magnetocrystalline anisotropy using `lm` (blue symbols and solid curves) and `lmgf` (red symbols and dashed curves). (a) FePt with the magnetic part of the exchange–correlation field scaled by a factor between 0 and 1. (b) $(Fe_{1-x}Co_x)_2B$ alloy in the virtual crystal approximation, with the cation charge varied from 26 to 27. Lower curves with squares: full three-centre `lm` and `lmgf` with third-order potential functions. Upper curves (shifted by 2 meV/f.u. in panel (a) and by 0.5 meV/f.u. in panel (b)): two-centre approximation in `lm` and second-order potential functions in `lmgf`. The charge density in all calculations for the given material is taken from the full self-consistent `lm` calculation (without scaling in the case of FePt).

becomes the resolvent of the second-order Hamiltonian Eq. (37), with added $\hat{V}_{\rm SO}$, as expected.

To third order in $\epsilon-\epsilon_\nu$ and to first order in $\hat{V}_{\rm SO}$, we find from Eq. (71):

$$\begin{aligned}\hat{\omega}^{(3)} &= \epsilon\hat{1}-\hat{\epsilon}_\nu+\hat{p}(\epsilon\hat{1}-\hat{\epsilon}_\nu)^3-\hat{V}^{(3)}_{\rm SO}\\ &= \hat{\epsilon}'-\hat{\epsilon}_\nu-\hat{V}^{(3)}_{\rm SO}\end{aligned} \tag{77}$$

where ϵ' is defined in Eq. (35), and

$$\hat{V}^{(3)}_{\rm SO}=\hat{V}_{\rm SO}+\left\{\hat{V}_{\rm SO},\hat{p}(\epsilon\hat{1}-\hat{\epsilon}_\nu)^2\right\} \tag{78}$$

with $\{\hat{A},\hat{B}\}=\hat{A}\hat{B}+\hat{B}\hat{A}$. Because $\hat{V}_{\rm SO}$ is defined at the fixed values of the logarithmic derivatives, the spin–orbit coupling parameters $\xi_{\ell,\sigma\sigma'}$ in $\hat{V}_{\rm SO}$ are calculated with ϵ replaced by ϵ':

$$\begin{aligned}\xi_{\ell,\sigma\sigma'}(\epsilon) =& \langle\phi_{\ell\sigma}|\xi(r)|\phi_{\ell\sigma'}\rangle\\ &+(\epsilon'_{\ell\sigma}-\epsilon_{\nu\ell\sigma})\langle\dot{\phi}_{\ell\sigma}|\xi(r)|\phi_{\ell\sigma'}\rangle\\ &+(\epsilon'_{\ell\sigma'}-\epsilon_{\nu\ell\sigma'})\langle\phi_{\ell\sigma}|\xi(r)|\dot{\phi}_{\ell\sigma'}\rangle\\ &+(\epsilon'_{\ell\sigma}-\epsilon_{\nu\ell\sigma})(\epsilon'_{\ell\sigma'}-\epsilon_{\nu\ell\sigma'})\langle\dot{\phi}_{\ell\sigma}|\xi(r)|\dot{\phi}_{\ell\sigma'}\rangle\end{aligned} \tag{79}$$

Of course, just as in the non-relativistic case, ϵ is also replaced by ϵ' where it appears explicitly in Eqs. (73)–(75). $\hat{V}_{\rm SO}$ is always calculated as the *exact* energy derivative of $\hat{V}_{\rm SO}$.

Fig. 7 shows the comparison of the magnetocrystalline anisotropy calculated using `lm` and `lmgf` for two benchmark systems. Two cases are displayed: `lmgf` with second-order potential functions compared with the corresponding two-centre approximation in `lm`, and `lmgf` with third-order potential functions compared with the full three-centre `lm` calculation. The agreement in both cases for FePt [panel (a)] is very good, while for the $(Fe_{1-x}Co_x)_2B$ alloy in the virtual crystal approximation it is essentially perfect.

2.17. Fully relativistic LMTO-ASA

We have developed a fully relativistic extension of the LMTO-ASA code within a relativistic generalisation of the density functional formalism [39–43]. In the most general case, one needs to solve the Kohn–Sham Dirac equation

$$H\Psi(\varepsilon,\mathbf{r})=\varepsilon\Psi(\varepsilon,\mathbf{r}) \tag{80}$$

with

$$H=c\alpha\mathbf{p}+(\beta-I_4)mc^2+V(\mathbf{r})I_4+\mu_B\beta\mathbf{B}_{\rm eff}(\mathbf{r})\Sigma \tag{81}$$

where

$$\alpha=\begin{pmatrix}0&\sigma\\ \sigma&0\end{pmatrix},\ \beta=\begin{pmatrix}I_2&0\\ 0&-I_2\end{pmatrix},\ \Sigma=\begin{pmatrix}\sigma&0\\ 0&\sigma\end{pmatrix} \tag{82}$$

Here, σ is the vector of Pauli matrices, $\mathbf{p}$ is the momentum operator, $\mathbf{B}_{\rm eff}(\mathbf{r})$ is an effective spin-dependent potential acting on electrons. It should be noted that this is a simplified form of the relativistic Kohn–Sham equation in which the orbital contribution to the 4-component relativistic current is neglected. This simplification is necessary in order to avoid the significant formulaic and computational complications that arise in the relativistic *current density* formulation of the density functional theory [43].

For a spherically symmetric potential $V(r)=1/2[V_\uparrow(r)+V_\downarrow(r)]$ inside a single MT sphere, the direction of the magnetic field can be assumed to point along the z direction, $\mathbf{B}_{\rm eff}=B(r)\hat{z}$, where $B(r)=1/2[V_\uparrow(r)-V_\downarrow(r)]$. Then, Eq. (81) can be written as

$$H=[c\alpha\mathbf{p}+(\beta-I_4)mc^2+V(r)I_4+\mu_B\beta B(r)\Sigma_z] \tag{83}$$

The solutions of the Kohn–Sham Dirac equation (83) are linear combinations of bispinors:

$$\Psi_\mu(\mathbf{r},\varepsilon)=\sum_\kappa\Psi_{\kappa\mu}(\mathbf{r},\varepsilon) \tag{84}$$

$$\Psi_{\kappa\mu}(\mathbf{r},\varepsilon)=\begin{pmatrix}g_{\kappa\mu}(\varepsilon,r)\Omega_{\kappa\mu}(\hat{r})\\ if_{\kappa\mu}(\varepsilon,r)\Omega_{-\kappa\mu}(\hat{r})\end{pmatrix} \tag{85}$$

Here, $\Omega_{\kappa\mu}(\hat{r})$ are the spin spherical harmonics, μ is the projection of the total angular momentum, and κ is the relativistic quantum number, $\kappa^2=J(J+1)+1/4$. The radial amplitudes $g_{\kappa\mu}(\varepsilon,r)$ and $f_{\kappa\mu}(\varepsilon,r)$ satisfy the following set of coupled differential equations:

$$\begin{aligned}\left[\frac{d}{dr}+\frac{1+\kappa_1}{r}\right]g_{\kappa_1\mu}(\varepsilon,r)&=\left[1+\frac{\varepsilon-V(r)+u'B(r)}{c^2}\right]cf_{\kappa_1\mu}(\varepsilon,r)\\ \left[\frac{d}{dr}+\frac{1-\kappa_1}{r}\right]cf_{\kappa_1\mu}(\varepsilon,r)&=[-(\varepsilon-V(r))-uB(r)]\,cg_{\kappa_1\mu}(\varepsilon,r)\\ &\quad-\sqrt{1-u^2}B(r)g_{\kappa_2\mu}(\varepsilon,r)\\ \left[\frac{d}{dr}+\frac{1+\kappa_2}{r}\right]g_{\kappa_2\mu}(\varepsilon,r)&=\left[1+\frac{\varepsilon-V(r)+u''B(r)}{c^2}\right]cf_{\kappa_2\mu}(\varepsilon,r)\\ \left[\frac{d}{dr}+\frac{1-\kappa_2}{r}\right]cf_{\kappa_2\mu}(\varepsilon,r)&=[-(\varepsilon-V(r))-uB(r)]\,cg_{\kappa_2\mu}(\varepsilon,r)\\ &\quad-\sqrt{1-u^2}B(r)g_{\kappa_1\mu}(\varepsilon,r)\end{aligned} \tag{86}$$

where

$$u=\frac{\mu}{\ell+1/2},\quad u'=\frac{\mu}{\ell-1/2},\quad u''=\frac{\mu}{\ell+3/2} \tag{87}$$

In the general case, in the presence of a magnetic field, one has to solve a system of two infinite sets of mutually coupled differential equations because the magnetic field couples radial amplitudes with different relativistic quantum numbers κ. Specifically, states with κ_1 to those with $\kappa_2=\kappa_1$ and $\kappa_2=-\kappa_1-1$, i.e., states of the same ℓ, but also states with κ_1 to those with

$\kappa_2 = 1 - \kappa_1$, i.e., states with different ℓ's, $\Delta\ell = \pm 2$. To avoid this complication, this coupling is neglected. In this case, the set simplifies into coupled equations for each pair $\ell\mu$. For $|\mu| = \ell + 1/2$, there is no coupling, so similar to the non-relativistic case, there is only regular solution with quantum numbers $\kappa\mu$, while for $|\mu| < \ell + 1/2$, we need to solve the set of four coupled equations (86) for the four unknown radial functions $g_{\kappa_1\mu}$, $f_{\kappa_1\mu}$, $g_{\kappa_2\mu}$, and $f_{\kappa_1\mu}$. The coefficients u, u', and u'' in Eq. (86) result from matrix elements of the type $\langle\kappa\mu|\sigma_z|\kappa'\mu\rangle$ [43].

Physically, the neglected coupling corresponds to a magnetic spin–orbit interaction given by a term $2c^{-2}r^{-1}dB/dr\mathbf{L}\cdot\mathbf{S}$ in the weak relativistic domain [44]. Since this is proportional to the product of two small quantities (c^{-2} and dB/dr) its omission is justified in most cases.

The construction of the MT orbitals and the corresponding boundary conditions proceeds along the same principles as in the non-relativistic case described in the previous sections. The tail cancellation condition is conveniently formulated in terms of relativistic extensions of the potential and normalisation functions:

$$N_{\mathcal{L}}(\varepsilon) = (2\ell+1)\left(\frac{w}{s}\right)^{\ell+1} g_{\mathcal{L}}^{-1}(\varepsilon,s)\left(D_{\mathcal{L}}(\varepsilon,s) - I\ell\right)^{-1} \tag{88}$$

$$P_{\mathcal{L}}(\varepsilon) = 2(2\ell+1)\left(\frac{w}{s}\right)^{2\ell+1} \times \left(D_{\mathcal{L}}(\varepsilon,s) + I_\ell + I\right)\left(D_{\mathcal{L}}(\varepsilon,s) - I\ell\right)^{-1} \tag{89}$$

The logarithmic derivative matrix being given in terms of the small and large components of the radial amplitude:

$$D_{\mathcal{L}}(\varepsilon,r) = scf_{\mathcal{L}}(\varepsilon,s)g_{\mathcal{L}}^{-1}(\varepsilon,s) - \kappa - I \tag{90}$$

$$\kappa = \begin{pmatrix}\kappa_1 & 0\\ 0 & \kappa_2\end{pmatrix}$$

For states with $|\mu| < \ell + 1/2$, $N_{\mathcal{L}}(\varepsilon)$, $P_{\mathcal{L}}(\varepsilon)$, $D_{\mathcal{L}}(\varepsilon)$, $g_{\mathcal{L}}(\varepsilon)$, and $f_{\mathcal{L}}(\varepsilon)$ are 2×2 matrices for each subblock $\mathcal{L} = (\ell,\mu)$ with the general form

$$A_{\mathcal{L}}(\varepsilon) = \begin{pmatrix}A_{\alpha_1\kappa_1}(\varepsilon) & A_{\alpha_1\kappa_2}(\varepsilon)\\ A_{\alpha_2\kappa_1}(\varepsilon) & A_{\alpha_2\kappa_2}(\varepsilon)\end{pmatrix} \tag{91}$$

Note that the indices in this matrix have different physical meaning. Although the values of index α are numerically equal to those of κ, index α stands for different behaviour of the radial function at the origin [42,43] while κ stands for solutions with different quantum states. Therefore, these matrices are not symmetric and do not commute with each other.

The linearisation can be formulated in a matrix form [in the following we will leave out the subblock index $\mathcal{L}$; all matrices presented have the form of Eq. (91)]. Within second-order approximation, the radial amplitudes are expanded around a linearisation energy:

$$g(\varepsilon,r) \approx g_\nu(r) + (\varepsilon - \varepsilon_\nu)\, I\dot{g}_\nu(r) \tag{92}$$

$$f(\varepsilon,r) \approx f_\nu(r) + (\varepsilon - \varepsilon_\nu)\, I\dot{f}_\nu(r) \tag{93}$$

Then, a symmetric matrix form of the linearisation of the logarithmic derivative can be written as

$$(D(\varepsilon) - D_\nu)^{-1} = -\frac{s}{\epsilon} g_\nu g_\nu^T + A \tag{94}$$

where

$$A = -\frac{1}{2}\left[(D_\nu - D_{\dot\nu})^{-1} + \left(D_\nu - D_{\dot\nu}^T\right)^{-1}\right] = -\frac{1}{2}s\left(\dot g_\nu g_\nu^T + g_\nu \dot g_\nu^T\right) \tag{95}$$

By direct substitution of (94) into (89) we obtain the parameterisation of the potential function:

$$P(\varepsilon) = R\,(V-\varepsilon)^{-1}\,R^T + Q \tag{96}$$

or equivalently

$$P^{-1}(\varepsilon) = W\,(C-\varepsilon)^{-1}\,W^T + \gamma \tag{97}$$

where

$$V = \varepsilon_\nu + sg_\nu^T\left[A + (D_\nu - Il)^{-1}\right]^{-1} g_\nu \tag{98}$$

$$R = \sqrt{2s}\left(\frac{w}{s}\right)^{\ell+\frac{1}{2}} (2\ell+1)\left[A\,(D_\nu - Il) + I\right]^{-1} g_\nu \tag{99}$$

$$Q = 2\,(2\ell+1)\left(\frac{w}{s}\right)^{2\ell+1}\left[I + (2\ell+I)\left[(D_\nu - Il) + A^{-1}\right]^{-1}\right] \tag{100}$$

and

$$\gamma = Q^{-1}, \quad W = \gamma R, \quad C = V + R^T\gamma R \tag{101}$$

For an arbitrary representation α:

$$P^{\alpha\,-1}(\varepsilon) = W\,(C-\varepsilon)^{-1}\,W^T + \gamma - \alpha \tag{102}$$

which is the relativistic analogue of Eq. (48). In general, the matrices V, R, Q, C, γ and W are non-diagonal, some are symmetric and some are not, so they do not all commute with each other. This is related to the different physical origin of the matrix indices discussed above. The parameter W is the relativistic analogue of $\sqrt{\Delta}$ introduced earlier [see Eq. (33)], and Eq. (96) is analogous to the non-relativistic Eq. (32). We should note that in the code, we have also included a third-order parameterisation of the potential function similar to the non-relativistic case, Eq. (35).

The physical Green function and Hamiltonian are written in terms of the potential function and scalar relativistic structure constants matrices after a transformation of the former from $\kappa\mu$ to $\ell m m_s$ basis:

$$G^\alpha = -\frac{1}{2}\ddot P^\alpha \dot P^\alpha + \frac{2}{w} N^{\alpha T} g^\alpha N^\alpha \tag{103}$$

where

$$g^\alpha = (P^\alpha - S^\alpha)^{-1} \tag{104}$$

these are the relativistic analogues of Eq. (57) and the scattering path operator.

Within the framework of the TB-LMTO and principal layer approach, the fully relativistic version of the green function for layered geometry is constructed straightforwardly from the site diagonal fully relativistic potential function [43].

2.18. Coherent potential approximation

The coherent potential approximation (CPA) is a Green's function-based method used to describe the electronic structure of disordered substitutional alloys. Questaal's `lmgf` code implements the CPA in the Atomic Spheres Approximation, following the formulation of Refs. [43,45]. Any lattice site i can be occupied by any number of components a with probabilities (concentrations) c_i^a, which must be supplied by the user. These components can have different atomic sphere radii, and each has its own charge density, atomic potential, and diagonal matrix of potential functions $P_i^a(\epsilon)$. Each site with substitutional disorder ("CPA site") is also assigned a coherent potential matrix $\mathcal{P}_i(\epsilon)$ that has the same orbital structure as $P_i^a(\epsilon)$ but is off-diagonal, with the restriction that it must be invariant under its site's point group. The elements of the $\mathcal{P}_i(\epsilon)$ matrix are complex even if ϵ is real, and it is fixed by the CPA self-consistency condition.

The configurational average of the scattering path operator is given by

$$\bar g(\epsilon,\mathbf{k}) = [\mathcal{P}(\epsilon) - S(\mathbf{k})]^{-1} \tag{105}$$

where the site-diagonal matrix $\mathcal{P}(\epsilon)$ absorbs the coherent potential matrices for the CPA sites and the diagonal matrices of the

conventional potential functions for the sites that are occupied deterministically ("non-CPA sites"). The matrix in Eq. (105) is integrated over the Brillouin zone in the usual way, and its site-diagonal blocks $\bar{g}_{ii}$ are extracted.

For non-CPA sites, the full Green's function, density matrix, and the charge density are obtained from $\bar{g}_{ii}$ in the usual way. For each CPA site i, we define the scattering path operator $\bar{g}_i^a$ (separately for each component a), which is the statistical average under the restriction that site i is occupied deterministically by component a, while all other sites in the infinite crystal are occupied statistically, according to their average concentrations. Such quantities are called "conditionally averaged". The charge density for the component a on site i is obtained from the site-diagonal block $\bar{g}_{ii}^a$ on site i of the conditionally averaged $\bar{g}_i^a$. This site-diagonal block can be found from the matrix equation on that site:

$$(\bar{g}_{ii}^a)^{-1} = \bar{g}_{ii}^{-1} + P_i^a - \mathcal{P}_i \tag{106}$$

and the CPA self-consistency condition reads

$$\bar{g}_{ii} = \sum_i c_i^a \bar{g}_{ii}^a \tag{107}$$

Iteration to self-consistency is facilitated [43,45] by introducing the *coherent interactor* matrix Ω_i, for each CPA site, defined through $\bar{g}_{ii} = (\mathcal{P}_i - \Omega_i)^{-1}$. The conditionally averaged $\bar{g}_{ii}^a$ is then $\bar{g}_{ii}^a = (P_i^a - \Omega_i)^{-1}$. The latter two equations can be used to re-express the self-consistency condition (107) in terms of $\mathcal{P}_i$, Ω_i, and P_i^a without an explicit reference to the scattering path operator:

$$(\mathcal{P}_i - \Omega_i)^{-1} = \sum_i c_i^a (P_i^a - \Omega_i)^{-1} \tag{108}$$

The Ω_i matrices (for each required complex energy point) are converged to self-consistency using the following procedure. At the beginning of the CPA iteration, Eq. (108) is used to obtain $\mathcal{P}_i$ for each CPA site from Ω_i, and then $\mathcal{P}_i$ is inserted in Eq. (105). After integration over $\mathbf{k}$, the site-diagonal block $\bar{g}_{ii}$ is extracted for each CPA sites and used to obtain the next approximation for $\Omega_i = \mathcal{P}_i - \bar{g}_{ii}^{-1}$, closing the self-consistency loop. The output Ω_i matrices are linearly mixed with their input values. The mixing parameter can usually be set to 1 for energy points that are not too close to the real axis.

The CPA loop is repeated until the Ω_i matrices converge to the desired tolerance, after which a charge iteration is performed. At the beginning of the calculation, the Ω_i matrices are initialised to zero unless they have already been stored on disk. The CPA loop can sometimes converge to an unphysical symmetry-breaking solution; this can usually be avoided by symmetrising the coherent potentials using the full space group of the crystal.

Fig. 8 illustrates one recent application of Questaal's implementation of the CPA. A number of new, potentially high-impact technologies, Josephson MRAM (JMRAM) in particular, performs write operations by rotating a patterned magnetic bit. These operations consume a significant portion of a system's power. One way to reduce the power consumed by write operations is to substitute materials with smaller saturation magnetisation M_s. The minimum useable M_s is limited by the need to maintain large enough energy barrier to prevent data loss due to thermal fluctuations. Low M_s can greatly reduce power particularly for JMRAM [48], which operates at around 4k. One way to reduce M_s in permalloy ($Ni_{80}Fe_{20}$ alloys, the most commonly used JMRAM material), is to admix Cr into it, as Cr aligns antiferromagnetically to Fe and Ni. Fig. 8 shows some results adapted from a recent joint experimental and theoretical study of $Py_{1-x}Cr_x$, [46]. The CPA calculations of M_s track measured values fairly well. Also shown is the CPA energy band structure for the majority spin. Alloy scattering causes bands to broaden out, and scattering lifetimes can be extracted. A detailed account can be found in Vishina's PhD thesis [47].

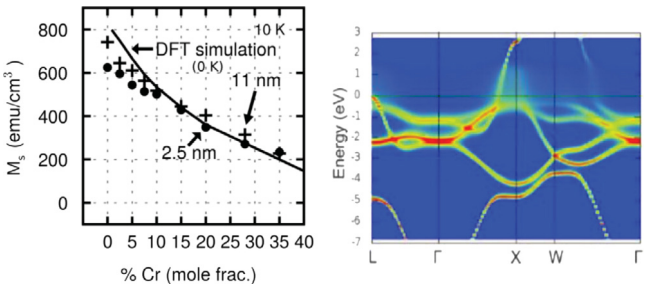

Fig. 8. Left: DFT magnetic moment in $(Ni_{80}Fe_{20})_{1-x}Cr_x$, compared to experiment (films with 11 nm (crosses) and 2.5 nm (dots) thickness) at 10k. Right: Majority-spin CPA spectral function at $x = 20\%$.
Source: Adapted from Ref. [46] and Vishina's PhD thesis [47].

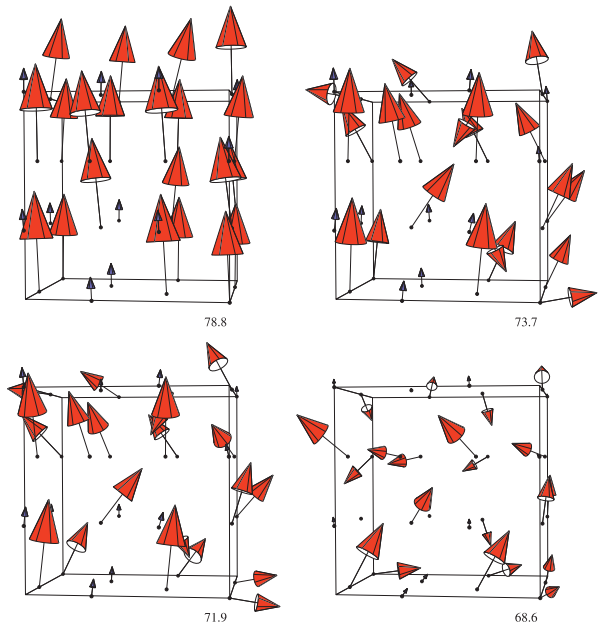

Fig. 9. Self-consistent magnetic spin configurations of the fcc Fe-Ni alloy at the four volumes 78.8, 73.7, 71.9, and 68.6 a.u. Red and blue arrows show magnetic moments on Fe and Ni atoms, respectively.
Source: Taken from Ref. [50].

2.19. Noncollinear magnetism

The ASA codes `lm`, `lmgf` and `lmpg` are fully noncollinear in a rigid-spin framework, meaning that the spin quantisation axis is fixed within a sphere, but each sphere can have its own axis. The formulation is a straightforward generalisation of the nonmagnetic case. Potential and Hamiltonian-like objects become 2×2 matrices in spin space. The structure matrix S is diagonal matrix in this space and independent of spin, while potential functions P and potential parameters of Section 2.8.1, become 2×2 matrices that are diagonal in ℓ but off-diagonal in spin. (Alternatively, a local spin quantisation axis can be defined which makes P diagonal in spin; then S is no longer diagonal.) How Questaal constructs the Hamiltonian in the general noncollinear case, and also for spin spirals, is discussed in Ref. [49].

These codes have implemented spin dynamics [11], integrating the Landau–Lifshitz equation using a solver from A. Bulgac and D. Kusnezov [51], the spin analogue of molecular dynamics and molecular statics. The formulation is described in detail in Ref. [49]. Codes also implement spin statics. Torques needed for both are obtained in DFT from the off-diagonal parts of the spin-density matrix. A classic application is the study of INVAR.

Fe-Ni alloys with a Ni concentration around 35 atomic % (INVAR) exhibit anomalously low, almost zero, thermal expansion over a considerable temperature range. In Ref. [50] it was shown that at 35% composition, Fe-Ni alloys are on the cusp of a collinear–noncollinear transition, and this adds a negative contribution to the Grüneisen parameter. Fig. 9 is reproduced from that paper.

3. Full potential implementation

Questaal's primary code `lmf` is an augmented-wave implementation of DFT without shape approximations. It also handles the one-body part of the quasiparticle self-consistent *GW* approximation, by adding the (quasiparticlised) *GW* self-energy to the DFT part. Closely related codes are `lmfgwd`, a driver supplying input for the *GW*, `lmfdmft`, a driver supplying input for a DMFT solver, and `lmfgws`, a post-processing tool that generates quantities using the one-body part from `lmf` and the dynamical self-energy from *GW* or DMFT. This code uses different definitions for classical functions (see Appendix B), and we shift to those definitions in what follows.

As for the DFT part, Questaal's unique features are:

- it uses a three-component augmentation, following most of the original method of Ref. [31]. The augmentation is reviewed in Section 3.6. Ref. [52] offers a slightly different presentation and shows how it is connected with the PAW method [4];
- it has a more general basis set. Ordinary Hankel functions solve Schrödinger's equation for a MT potential, but real potentials vary smoothly into the interstitial (Fig. 10). The envelope function of a minimal basis set must adapt to this potential. The traditional Questaal basis uses smooth Hankel functions [53], which may be thought of as a convolution of a Gaussian function and a traditional Hankel function; they are developed in Section 3.1. This traditional basis works very well for most systems; however when the system is very open, it is slightly incomplete (Section 3.13). One way to surmount the incompleteness is to combine smooth Hankel functions with plane waves; this is the "Plane-wave Muffin Tin" (PMT) basis [54] (see also Ref. [52]). While it would seem appealing, PMT suffer from two serious drawbacks: first, it tends to become over-complete even with a relatively small number of plane waves, and second, it is not *compact* (minimal, short ranged and as complete as possible for the relevant energy window); and
- Questaal's most recent development is the "Jigsaw Puzzle Orbital" (JPO) basis, which uses information from the augmentation to construct an optimal shape for the envelope functions. To the best of our knowledge, JPO's are the closest practical realisation of compactness. This is particularly important in many-body treatments where the efficacy of a theory hinges critically on compactness. This new basis will be presented more fully elsewhere. In Section 3.12 we show how a transformation to a tight-binding representation can be carried out in a full-potential framework. In the present version the basis set is merely a unitary transformation of the original one.

3.1. Smooth Hankel functions

In his PhD dissertation Michael Methfessel introduced a class of functions $\mathcal{H}_{kL}(E, r_s, \mathbf{r})$. As limiting cases they encompass both ordinary Hankel functions and Gaussian functions (see Eqs. (122) and (124) below). They are explained in detail in Ref. [53]; here we present enough information for development of the Questaal

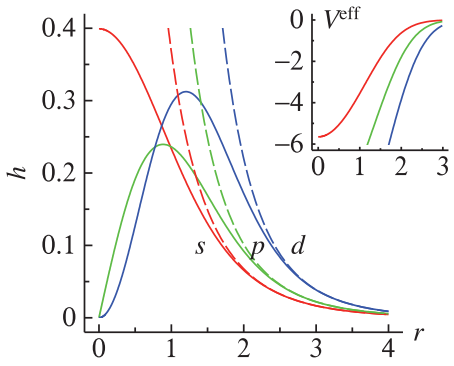

Fig. 10. Radial part of smooth Hankel functions $\mathcal{H}_L$ for s, p, d orbitals (solid lines) and corresponding ordinary Hankel functions H_L (dashed lines). Smoothing radius and energy were chosen to be $r_s = +1$ and $E = -1$, respectively. For $r \gg r_s$, $\mathcal{H}_L \to H_L$, while for $r \ll r_s$, $\mathcal{H}_L \propto r^{\ell}$ and $H_L \propto r^{-\ell-1}$. The H_L satisfy the Helmholtz wave equation (Schrödinger equation for constant potential), while the $\mathcal{H}_L$ satisfy a Schrödinger equation corresponding to a potential $V^{\text{eff}} = -4\pi G_L/\mathcal{H}_L$, as shown in the inset.

basis set. They are all connected to the smooth Hankel function for $\ell = 0$. Its Fourier transform is

$$\widehat{h}_0(E, r_s; q) = -\frac{4\pi}{E - q^2} e^{r_s^2(E-q^2)/4} \tag{109}$$

which is a product of Fourier transforms of ordinary Hankel function for $\ell = 0$ with a Gaussian function of width $2/r_s$.

We use standard definitions of Fourier transforms

$$\widehat{f}(\mathbf{q}) = \int e^{-i\mathbf{q}\cdot\mathbf{r}} f(\mathbf{r})\, d^3r$$

$$f(\mathbf{r}) = \frac{1}{(2\pi)^3} \int e^{+i\mathbf{q}\cdot\mathbf{r}} \widehat{f}(\mathbf{q})\, d^3q \tag{110}$$

Thus in real space, $h_0(E, r_s; r)$ is a convolution of a ordinary Hankel function and a Gaussian function. It has an analytic representation

$$h_0(E, r_s; r) = \frac{1}{2r}\,(u_+ - u_-) \tag{111}$$

$$\dot{h}_0(E, r_s; r) = \frac{1}{4\bar{\kappa}}\,(u_+ + u_-) \tag{112}$$

$$u_\pm = e^{\mp\bar{\kappa}r}\left[1 - \text{erf}\left(\frac{r_s\bar{\kappa}}{2} \mp \frac{r}{r_s}\right)\right] \tag{113}$$

$$E = -\bar{\kappa}^2 \text{ if } E < 0 \tag{114}$$

For small r, h_0 behaves as a Gaussian and it evolves smoothly into an ordinary Hankel when $r \gg r_s$. $\dot{h}_0$ is the energy derivative of h_0.

Questaal's envelope functions are generalised Hankel functions $\mathcal{H}_L(E, r_s; \mathbf{r})$. Their Fourier representation has a closed form

$$\begin{aligned}\widehat{\mathcal{H}}_L(E, r_s; \mathbf{q}) &= \mathcal{Y}_L(-i\mathbf{q})\frac{-4\pi}{E - q^2} e^{r_s^2(E-q^2)/4} \\ &= \mathcal{Y}_L(-i\mathbf{q})\widehat{h}_0(E, r_s; q)\end{aligned} \tag{115}$$

$\mathcal{Y}_L(\mathbf{r})$ is a spherical harmonic polynomial (Appendix A). Following the usual rules of Fourier transforms this function has a real-space representation

$$\mathcal{H}_L(E, r_s; \mathbf{r}) = \mathcal{Y}_L(-\nabla) h_0(E, r_s; r) \tag{116}$$

$\mathcal{Y}_L(\mathbf{r})$, $\mathbf{r} = (x, y, z)$ is a polynomial in (x, y, z), so is meaningful to talk about $\mathcal{Y}_L(-\nabla)$. The extended family $\mathcal{H}_{kL}(\mathbf{r})$ is defined through powers of the Laplacian operator:

$$\widehat{\mathcal{H}}_{kL}(E, r_s; \mathbf{q}) = (-iq)^{2k}\, \widehat{\mathcal{H}}_L(E, r_s; \mathbf{q})$$

$$\mathcal{H}_{kL}(E, r_s; \mathbf{r}) = \nabla^{2k}\mathcal{H}_L(E, r_s; \mathbf{r}) \tag{117}$$

A recursion relation for $\mathcal{H}_L(E, r_s; \mathbf{r})$ for $\ell > 0$ can be derived from properties of Fourier transforms in spherical coordinates. Any function whose Fourier transform factors as

$$\hat{F}(\mathbf{q}) = \hat{f}(q)Y_L(\hat{\mathbf{q}}) \tag{118}$$

has a real-space form

$$F(\mathbf{r}) = f(r)Y_L(\hat{\mathbf{r}}) \tag{119}$$

where f and $\hat{f}$ are related by [53]

$$f(r) = \frac{4\pi i^\ell}{(2\pi)^3}\int_0^\infty q\hat{f}(q)j_\ell(qr)\,qr\,dq$$

$$\hat{f}(q) = 4\pi(-i)^\ell\int_0^\infty rf(r)j_\ell(qr)\,qr\,dr$$

and j_ℓ is the spherical Bessel function. We can therefore express $\mathcal{H}_L$ in a Slater–Koster form

$$\mathcal{H}_L(\mathbf{r}) = h_\ell(r)\,Y_L(\hat{\mathbf{r}}) \equiv \chi_\ell(r)\,\mathcal{Y}_L(\mathbf{r}) \tag{120}$$

$$\chi_\ell(r) = (1/r^\ell)\,h_\ell(r) \tag{121}$$

For envelope functions used in basis sets, e.g. $\mathcal{H}_L(E, r_s; \mathbf{r})$, the radial portion of the Fourier transform does not depend on ℓ; therefore corresponding real-space part depends on ℓ only through the j_ℓ. This is a very useful fact. It is possible to derive a recurrence relation to obtain χ_ℓ for $\ell > 1$; this provides an efficient scheme for calculating them. The recurrence is derived in Ref. [53] (see Eq. 6.21):

$$\chi_{\ell+1} = \frac{2\ell+1}{r^2}\chi_\ell - \frac{E}{r^2}\chi_{\ell-1} - \frac{4\pi}{r^2(\pi r_s^2)^{3/2}}\left(\frac{2}{r_s^2}\right)^{\ell-1} e^{r_s^2E/4-(r/r_s)^2}$$

χ_0 and χ_{-1} are needed to start the recursion. χ_0 is written in Eq. (111) and χ_{-1} can be obtained by combining Eqs. (112), (136) and (121).

By taking limiting cases we can see the connection with familiar functions, and also the significance of parameters E and r_s.

(i) $k = r_s = 0$: $\widehat{H}_0(E, 0; \mathbf{q}) = -4\pi/(E - q^2)\mathcal{Y}_{00}(\mathbf{q})$. This is the Fourier transform of $H_0(E, 0; r) = \mathcal{Y}_{00}(\mathbf{r})\exp(-\bar{\kappa}r)/r$, and is proportional to the $\ell = 0$ spherical Hankel function of the first kind, $h_\ell^{(1)}(z)$. For general L the relation defines Questaal's standard definition of ordinary Hankel functions

$$H_L(E, 0; \mathbf{r}) = H_{0L}(E, 0; \mathbf{r}) = -i^\ell\bar{\kappa}^{\ell+1}h_\ell^{(1)}(i\bar{\kappa}r)\mathrm{Y}_L(\hat{\mathbf{r}}) \tag{122}$$

(ii) $k = 1$ and $E = 0$: $\widehat{\mathcal{H}}_{10}(0, r_s; \mathbf{q}) = -4\pi e^{-r_s^2q^2/4}$. This is the Fourier transform of a Gaussian function of width r_s. For general L we can define the family as

$$G_L(E, r_s; \mathbf{r}) = \mathcal{Y}_L(-\nabla)g(E, r_s; r) \tag{123}$$

$$g(E, r_s; r) = \left(\pi r_s^2\right)^{-3/2} e^{Er_s^2/4}e^{-r^2/r_s^2} \tag{124}$$

Evidently $\widehat{\mathcal{H}}_L(\mathbf{q})$ is proportional to the product of the Fourier transforms of a conventional spherical Hankel function of the first kind, and a Gaussian. By the convolution theorem, $\mathcal{H}_L(\mathbf{r})$ is a convolution of a Hankel function and a Gaussian. For $r \gg r_s$, $\mathcal{H}_L(\mathbf{r})$ behaves as a Hankel function and asymptotically tends to $\mathcal{H}_L(\mathbf{r}) \to r^{-\ell-1}\exp(-\sqrt{-E}r)\mathrm{Y}_L(\hat{\mathbf{r}})$. For $r \ll r_s$ it has structure of a Gaussian; it is therefore analytic and regular at the origin, varying as $r^\ell\mathrm{Y}_L(\hat{\mathbf{r}})$. Thus, the $r^{-\ell-1}$ singularity of the Hankel function is smoothed out, with r_s determining the radius for transition from Gaussian-like to Hankel-like behaviour. Thus, the smoothing radius r_s determines the smoothness of $\mathcal{H}_L$, and also the width of G_L.

By analogy with Eq. (117) we can extend the G_L family with the Laplacian operator:

$$G_{kL}(E, r_s; \mathbf{r}) = \nabla^{2k}\,G_L(E, r_s; \mathbf{r}) \tag{125}$$

$$= \mathcal{Y}_L(-\nabla)\nabla^{2k}g(E, r_s; r) \tag{126}$$

$$= \mathcal{Y}_L(-\nabla)\left(\frac{1}{r}\frac{\partial^2}{\partial r^2}r\cdot\right)^k g(E, r_s; r) \tag{127}$$

$$\widehat{G}_{kL}(E, r_s; \mathbf{q}) = \mathcal{Y}_L(-i\mathbf{q})(-q^2)^k\,e^{r_s^2(E-q^2)/4} \tag{128}$$

Eq. (127) shows that G_{kL} has the structure (polynomial of order k in r^2) $\times$ G_L. Specifically [53]

$$G_{kL}(E, \mathbf{r}) = \frac{2^{k+\ell}(2k+2\ell+1)!!}{r_s^{2k+\ell}(2\ell+1)!!}p_{k\ell}(r_s; r)G_L(E, \mathbf{r})$$

where

$$p_{k\ell}(r_s; r) = \frac{(-1)^k(2\ell+1)!!2^k k!}{r_s^\ell(2k+2\ell+1)!!}L_k^{(\ell+1/2)}(r^2/r_s^2) \tag{129}$$

and

$$G_L = \left(\pi r_s^2\right)^{-3/2} e^{Er_s^2/4}\left(2/r_s^2\right)^\ell e^{-r^2/r_s^2} \tag{130}$$

$L_k^{(\ell+1/2)}(u)$ are generalised Laguerre polynomials [53], which have the following orthogonality relation

$$\int du\,L_k^{(\ell+1/2)}(u)L_{k'}^{(\ell+1/2)}(u)\,e^{-u}u^{(\ell+1/2)} = \Gamma(k+\ell+3/2)\delta_{kk'}$$

and as a consequence the G_{kL} are orthogonal in the following sense:

$$\int G_{kL}G_{k'L'}\,e^{a^2r^2}\,d^3r = \frac{2^{3k+\ell}k!(2k+2\ell+1)!!}{4\pi\left(\pi r_s^2\right)^{3/2}r_s^{4k+2\ell}}\delta_{kk'}\delta_{\ell\ell'}$$

This can be also written as

$$\int G_{kL}P_{k'L'}\,d^3r = \frac{2^{2k}k!(2\ell+1)!!}{4\pi r_s^{2k-\ell}}\delta_{kk'}\delta_{\ell\ell'} \tag{131}$$

where

$$P_{kL}(r_s; \mathbf{r}) = p_{k\ell}(r_s; r)\mathcal{Y}_L(\mathbf{r}) \tag{132}$$

We will need this relation for one-centre expansion of the $\mathcal{H}_L$ around remote sites (Section 3.6) since the simple expansion theorem Eq. (1) does not apply to them.

Comparing the last form Eq. (128) to Eq. (117) and the definition of $\mathcal{H}_{kL}$ Eq. (117), we obtain the useful relations

$$\mathcal{H}_{k+1,L}(E, r_s; \mathbf{r}) + E\mathcal{H}_{kL}(E, r_s; \mathbf{r}) = \left(\nabla^2 + E\right)\mathcal{H}_{kL}(E, r_s; \mathbf{r}) = -4\pi\,G_{kL}(E, r_s; \mathbf{r}) \tag{133}$$

This shows that $\mathcal{H}_{kL}$ is the solution to the Helmholtz operator $-\nabla^2 + E$ in response to a source term smeared out in the form of a Gaussian. A conventional Hankel function is the potential from a point multipole at the origin (see Eq. 6.14 in Ref. [53]): smearing out the singularity makes $\mathcal{H}_L$ regular at the origin, varying as r^ℓ for small r instead of $r^{-\ell-1}$. $\mathcal{H}_{kL}$ is also the solution to the Schrödinger equation for a potential that has an approximately Gaussian dependence on r (Ref. [53], Eq. 6.30).

3.2. Gradients of smooth Hankel functions

Gradients of the $\mathcal{H}_L$ are needed in several contexts, e.g. for forces and for matrix elements of the momentum operator, which enters into the dielectric function and velocity operator. Also, the energy derivative is needed in several contexts, e.g. for integration of some matrix elements (Appendix D).

The form of Eq. (115) suggests that gradient and position operators acting on $\widehat{\mathcal{H}}$ return functions of the same family. This

Table 2
Index m'_i corresponding to coefficient $\overline{C}^{(\pm)}_{i=1..2;\ell,m_\ell;p}$ in Eq. (138). Index K in $C^{(\pm)}_{KL;p}$ corresponding to m'_i is $K = (k \pm 1)^2 + (k \pm 1) + 1 + m'_i$.

i	$p = x$	$p = y$	$p = z$
1	$m_\ell - 1$	$-m_\ell - 1$	m_ℓ
2	$m_\ell + 1$	$-m_\ell + 1$	–

turns out to be the case, and it makes possible matrix elements of these operators with two such functions. Consider the energy derivative first. $\widehat{\mathcal{H}}_L(E, r_s; \mathbf{q})$ is readily differentiated

$$\dot{\widehat{\mathcal{H}}}_L(E, r_s; \mathbf{q}) = \widehat{\mathcal{H}}_L(E, r_s; \mathbf{q}) \left(\frac{1}{E - q^2} - r_s^2/4 \right) \tag{134}$$

In real space, the energy derivative is most easily derived from an integral representation of h_ℓ (Eq. A5 in Ref. [53])

$$h_\ell(r) = \frac{2^{\ell+1} r^\ell}{\sqrt{\pi}} \int_0^{1/r_s} \xi^{2\ell} e^{E/(4\xi^2)} e^{-r^2\xi^2} d\xi \tag{135}$$

It is readily seen that

$$r h_\ell(r) = 2\dot{h}_{\ell+1}(r)$$
$$\left(\frac{d}{dr} - \frac{\ell}{r}\right) h_\ell(r) = h_{\ell+1}(r) \tag{136}$$

Noting that the three components of the real harmonic polynomials $\mathcal{Y}_{1m}(\mathbf{r})$, Eq. (A.6) for $\ell = 1$, are (y, z, x), the gradient and position operators can be obtained from $-\sqrt{4\pi/3}\,\mathcal{Y}_{1p}(-\nabla)$ and $\sqrt{4\pi/3}\,\mathcal{Y}_{1p}(\mathbf{r})$ respectively, with the appropriate permutation of p. Using Eq. (116),

$$\mathcal{Y}_{1p}(\nabla)\mathcal{H}_L = -\mathcal{Y}_{1p}(-\nabla)\mathcal{Y}_L(-\nabla) h_0(E, r_s; r) \tag{137}$$

The operator product $\mathcal{Y}_L(-\nabla)\mathcal{Y}_K(-\nabla)$ can be expanded in the same manner as the product of polynomials $\mathcal{Y}_L(\mathbf{r})\mathcal{Y}_K(\mathbf{r})$ (Appendix A).

More generally, for a function of the form Eq. (119), it is possible to show that

$$\nabla_p f_\ell(r) \mathrm{Y}_L = f^{(+)}_\ell \sum_{i=1,2} \overline{C}^{(+)}_{i;\ell,m_\ell;p} \mathrm{Y}_{\ell+1 m'_i} + f^{(-)}_\ell \sum_{i=1,2} \overline{C}^{(-)}_{i;\ell,m_\ell;p} \mathrm{Y}_{\ell-1 m'_i} \tag{138}$$

where

$$f^{(+)}_\ell = [df_\ell/dr - \frac{\ell}{r} f_\ell]$$
$$f^{(-)}_\ell = [df_\ell/dr + \frac{\ell+1}{r} f_\ell] \tag{139}$$

$\overline{C}^{(\pm)}_{i;\ell,m_\ell;p}$ is a sparse representation of a linear transformation that maps $\mathrm{Y}_{\ell,m}(\mathbf{r})$ to a linear combination of two $\mathrm{Y}_{\ell+1,m'}(\mathbf{r})$ and two $\mathrm{Y}_{\ell-1,m'}(\mathbf{r})$. It is more compact to write each sum in a standard matrix form, as $\sum_K C^{(\pm)}_{KL;p} \mathrm{Y}_K$. $C^{(\pm)}_{KL;p}$ is a family of three rectangular matrices ($p = 1, 2, 3$), which for a particular L, has two nonzero elements corresponding to $m_{i'}$ (Table 2). An alternative definition of $C^{(\pm)}_{KL;p}$ is given in Appendix A.

The position operator $\mathbf{r}_p$ applied to $f_\ell(r)\mathrm{Y}_K$ results in the same form Eq. (138) but with radial functions $f^{(\pm)} \to rf(r)$. Note the close similarity with the gradient operator and in particular $rf_\ell = r^2(f^{(+)}_\ell - f^{(-)}_\ell)/(2\ell + 1)$. This operator appears for optical matrix elements of Hamiltonians with nonlocal potentials. It is more convenient to evaluate it in reciprocal space (Appendix D).

Functions of type Eq. (119) have the Slater–Koster form with another special property, namely that the Fourier transform $\hat{f}(q)$, Eq. (118), is independent of ℓ, from which it follows that $f_\ell(r)$ satisfies the following differential equations (Ref. [53], Eq. 4.7 and 4.20)

$$-f_{\ell+1}(r) = \frac{\partial f_\ell(r)}{\partial r} - \frac{\ell}{r} f_\ell(r)$$

Table 3
Mapping gradient and position operators of smoothed Hankel functions to functions in the same family.

$F(\mathbf{r})$	$f(r)$	$f^{(+)}$	$f^{(-)}$	$rf(r)$
$\mathcal{Y}_L$	r^ℓ	0	$(2\ell+1)r^{\ell-1}$	$r^{\ell+1}$
$r^{-2\ell-1}\mathcal{Y}_L$	$r^{-\ell-1}$	$(-2\ell-1)r^{-\ell-2}$	0	$r^{-\ell}$
$\mathcal{H}_{kL}$	h_{kl}	$-h_{k,\ell+1}$	$-h_{k+1,\ell-1}$	$2\dot{h}_{\ell+1}$
$H_L(E{=}0)$	Same as $\mathcal{Y}_L/r^{2\ell+1}$			
$J_L(E{=}0)$	Same as $\mathcal{Y}_L$			

$$-\nabla^2 f_{\ell-1}(r) = \frac{\partial f_\ell(r)}{\partial r} + \frac{(\ell+1)}{r} f_\ell(r)$$

Thus $f^{(+)}(r) = -f_{\ell+1}(r)$ and $f^{(-)}(r) = -\nabla^2 f_{\ell-1}(r)$.

The $\mathcal{H}_{kL}(\mathbf{r})$ family is of this type, and moreover, $\nabla^2\mathcal{H}_{kL}(\mathbf{r})$ merely maps $\mathcal{H}_{kL}$ to another member of the family (Eq. (117)). Another useful instance is the real harmonic polynomials $\mathcal{Y}_L$ (Eq. (A.6)). Equations in this section, e.g. Eq. (136), yield explicit forms for $f^{(+)}$ and $f^{(-)}$ for a number of functions of interest here, summarised in Table 3.

In general, e.g. for partial waves ϕ_ℓ inside augmentation spheres, $f^{(+)}$ and $f^{(-)}$ must be determined by numerical differentiation. For the pair of functions $\{r^\ell,\ r^{-\ell-1}\}$, $f^{(+)}_\ell$ ($f^{(-)}_\ell$) for the first (second) function vanishes. Any function can thus be matched at some r to this pair (Eq. (12)), which determines the projection of its gradient onto $Y_{\ell\mp1m'}$.

3.3. Two-centre integrals of smoothed Hankels

One extremely useful property of the $\mathcal{H}_{kL}$ is that the product of two of them, centred at different sites $\mathbf{R}_1$ and $\mathbf{R}_2$, can be integrated in closed form. The result is a sum of other $\mathcal{H}_{kL}$, evaluated at the connecting vector $\mathbf{R}_1 - \mathbf{R}_2$. This follows from Parseval's (Plancharel's) identity

$$\int \mathcal{H}^*_1(\mathbf{r} - \mathbf{R}_1)\mathcal{H}_2(\mathbf{r} - \mathbf{R}_2)\, d^3r =$$
$$(2\pi)^{-3} \int \widehat{\mathcal{H}}^*_1(\mathbf{q})\widehat{\mathcal{H}}_2(\mathbf{q}) e^{i\mathbf{q}\cdot(\mathbf{R}_1 - \mathbf{R}_2)}\, d^3q \tag{140}$$

and the fact that $\widehat{\mathcal{H}}^*_{k_1L_1}(\mathbf{q})\widehat{\mathcal{H}}_{k_2L_2}(\mathbf{q})$ can be expressed as a linear combination of other $\widehat{\mathcal{H}}_{kL}(\mathbf{q})$, or their energy derivatives. Derivations of these and related integrals are taken up in Appendix D.

3.4. Smoothed Hankels of positive energy

The smooth Hankel functions defined in (Ref. [53]) for negative energy also apply for positive energy. We demonstrate that here, and show that the difference between the conventional and smooth Hankel functions are real functions.

Ref. [53] defines $\bar{\kappa}$ in contradistinction to usual convention for κ

$$\bar{\kappa}^2 = -\varepsilon \quad \text{with } \bar{\kappa} > 0$$

and restricts $\varepsilon < 0$. According to usual conventions κ is defined as

$$\kappa = \sqrt{\varepsilon}, \quad \mathrm{Im}(\kappa) \geq 0$$

We can define for any energy $\bar{\kappa} = -i\kappa$ and therefore

$$\begin{cases} \bar{\kappa} \text{ real and positive}, & \varepsilon < 0 \\ \kappa \text{ real and positive}, & \varepsilon > 0 \end{cases}$$

Then

$$u_\pm(\varepsilon, r_s; r) = e^{\mp\bar{\kappa} r} \mathrm{erfc}\,(\bar{\kappa} r_s/2 \mp r/r_s)$$
$$= e^{\pm i\kappa r} \mathrm{erfc}\,(-i\kappa r_s/2 \mp r/r_s)$$

The smoothed Hankels for $\ell = 0, -1$ are real

$$h^s_0(r) = (u_+ - u_-)/2r$$
$$h^s_{-1}(r) = (u_+ + u_-)/2\bar{\kappa}$$

as defined in Ref. [53].

To extend the definition to any energy we define $U_\pm$ as:

$$U_\pm = e^{\pm i\kappa r}\,\mathrm{erfc}\,(r/r_s \pm i\kappa r_s/2)$$

The following relations are useful:

$$\mathrm{erfc}(-x^*) = 2 - \mathrm{erfc}^*(x)$$
$$\mathrm{erfc}(x^*) = \mathrm{erfc}^*(x)$$

Then for $\varepsilon < 0$, $i\kappa$ is real and

$$U_+ = 2e^{i\kappa r} - u_+$$
$$U_- = u_-$$

are also real. The difference in unsmoothed and smoothed Hankels is for $\ell = 0, -1$

$$\begin{aligned} h_0 - h^s_0 &= [e^{i\kappa r} - u_+/2 + u_-/2]/r \\ &= [U_+/2 + U_-/2]/r \\ h_{-1} - h^s_{-1} &= [e^{i\kappa r} - u_+/2 - u_-/2]/\bar{\kappa} \\ &= [U_+/2 - U_-/2]/(-i\kappa) \end{aligned}$$

For $\varepsilon > 0$, κ is real and $U_+ = U_-^*$. Then the difference in unsmoothed and smoothed Hankels for $\ell = 0, -1$:

$$h_0 - h^s_0 = [U_+/2 + U_-/2]/r = \mathrm{Re}(U_+)/r$$
$$h_{-1} - h^s_{-1} = [U_+/2 - U_-/2]/(-i\kappa) = -\mathrm{Im}(U_+)/\kappa$$

are real, although h_0 and h_{-1} are complex.

3.5. One-centre expansion

Ordinary Hankel functions have a one-centre expansion in terms of Bessel functions, Eq. (1): the shape of the radial function does not change as the vector $\mathbf{R}$ connecting source and field point changes, only the expansion coefficients S. This is also true for a plane wave. The $\mathcal{H}$ are more complicated: the r-dependence depend on $\mathbf{R}$, tending to Bessel functions only when $|\mathbf{R}| \gg r_s$. This complicates the one-centre expansion of $\mathcal{H}$. They can, however, be expanded in the polynomials P_{kL} (Eq. (132)) using the bi-orthogonality relation Eq. (131). For any smooth function F, its one-centre expansion can be written

$$F(\mathbf{r}) = \sum_{kL} C_{kL} p_{k\ell}(r) y_L(\mathbf{r}) \tag{141}$$

where

$$C_{kL} = \frac{4\pi r_s^{2k+\ell} 2^{2k}}{2^{2k} k!(2\ell+1)!} \int d^3r\, G_{kL}(r_s;\mathbf{r}) F(\mathbf{r})$$

Expressions for integrals G_{kL} with the $\mathcal{H}$ are written in Section 3.3.

Eq. (141) expands $\mathcal{H}_L$, which generalises Eq. (1) which expands H_L (plane waves can similarly be expanded in Bessel functions). Eq. (141) is more cumbersome, and it introduces another cutoff $k_{\max}$. These are drawbacks of this method, but it also brings with it a significant advantage: envelope functions of very different types can be constructed. This adds a considerable amount of flexibility to the method, as smooth Hankels $\mathcal{H}_L$ and plane waves can be combined in a uniform manner; this is the PMT basis [54]. Perhaps more important, basis functions can be tailored to the potential. This enables them to be very accurate whilst staying minimal. This is the JPO basis (Section 3.12).

If r_s becomes small G_{kL} becomes sharply peaked and the polynomial expansion becomes a Taylor expansion about the origin. But by allowing r_s to be a significant fraction of the MT radius s_R, the error gets distributed over the entire sphere and $k_{\max}$ can be set relatively low. Note that r_s used for this expansion need not be (and in practice is not) the same as r_s used to make the envelope functions.

3.6. Three-component augmentation

As noted in the introduction, the pseudopotential method shares many features in common with augmentation. Blöchl's PAW method [4] makes the connection explicit. Questaal starts from an envelope function $F^{(0)}_{RLi}(\mathbf{r})$ that extends everywhere in space, augmented as follows:

$$F_{RLi}(\mathbf{r}) = F^{(0)}_{RLi}(\mathbf{r}) + F^{(1)}_{RLi}(\mathbf{r}) - F^{(2)}_{RLi}(\mathbf{r}) \tag{142}$$

Index RLi labels the envelope function: R and L mark the channel where functions are centred, and the angular momentum, while i marks the kind of envelope function at R. General envelope functions (e.g. plane waves) need not be atom-centred, for such functions we adopt the labelling convention $R = L = 0$. Unlike the ASA, there is typically more than radial envelope function per R and L. (We will sometimes label F with a single Roman index, e.g. using i to implicitly refer to the entire label RLi.), $F^{(1)}$ is some linear combination of partial waves matching value and slope of $F^{(0)}(\mathbf{r})$, and $F^{(2)}(\mathbf{r})$ is the one-centre expansion of $F^{(0)}(\mathbf{r})$. $F^{(0)}$, and $F^{(2)}$ are truncated at some $\ell_{\max}$.

Conventional augmented methods construct matrix elements of a local or semilocal operator X straightforwardly as

$$\langle F^{(0)}_i + F^{(1)}_i - F^{(2)}_i | X | F^{(0)}_j + F^{(1)}_j - F^{(2)}_j \rangle$$

Evaluating the cross terms is unwieldy and for that reason conventional augmentation methods require F_i be expanded to very high L. However, it has been observed independently by several authors, the first being Soler and Williams [55] that the integrand can be approximated in the following three-component form:

$$F^{(0)}_i X F^{(0)}_j + F^{(1)}_i X F^{(1)}_j - F^{(2)}_i X F^{(2)}_j \tag{143}$$

Note the minus sign in the last term. The missing terms may be written as

$$(F^{(0)}_i - F^{(2)}_i) X (F^{(1)}_j - F^{(2)}_j) + (F^{(1)}_i - F^{(2)}_i) X (F^{(0)}_j - F^{(2)}_j)$$

Write the L projection of F as $\mathcal{P}_L F$. By construction $\mathcal{P}_L F^{(0)} = \mathcal{P}_L F^{(2)}$ so if the augmentation is carried out to $L \to \infty$, $F^{(0)}$ and $F^{(2)}$ are identical and the missing terms vanish identically. This form makes it clear that not only does Eq. (143) converge to the exact result in the limit $\ell_{\max} \to \infty$, but also that the error converges much more rapidly in $\ell_{\max}$ than does conventional augmentation.

This occurs for two reasons that work synergistically. The operator X is in practice nearly spherical, coupling parts of F with unequal L only weakly. Consider the one-centre expansion of a particular L of one of these components, and consider for simplicity the unit overlap operator. Near the augmentation boundary all components have the same value and slope by construction, and the (ℓ-projected) $F^{(0)} - F^{(2)}$ or $F^{(1)} - F^{(2)}$ vanishes to second order in $r - s$. Second, the (ℓ-projected) F varies as r^ℓ for large ℓ for all three components of F, because the angular momentum becomes the dominant contribution to the differential operator. Thus for large ℓ the product of two projected functions varies as $r^{2\ell}$ for all components, which is heavily weighted to the outer parts of the sphere, in the region where the projections of differences $F^{(0)} - F^{(2)}$ and $F^{(1)} - F^{(2)}$ are small. For such L projections of $F^{(0)}_i X F^{(0)}_j$ are then a good approximation to the (exact) $F^{(1)}_i X F^{(1)}_j$. By truncating both local components, only the (0) component

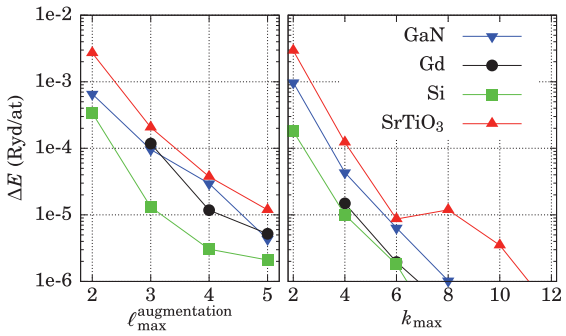

Fig. 11. Convergence of the total energy with respect to augmentation ℓ-cutoff and polynomial degree, Eq. (144).

remains, which as we have seen accurately represents the (exact) basis function, especially near s_R where it is dominant. In other words, the projections of the envelope functions contain projections from $\ell_{\max}+1$ to ∞, and do so with high accuracy starting at a relatively low $\ell_{\max}+1$; indeed convergence is very rapid as shown in Fig. 11 (see also Fig. 3 of Ref. [31]). This explains why the pseudopotential and PAW approximations can be truncated at much lower ℓ than the conventional augmented wave methods. Indeed, the present approach has much in common with PAW, but does not make approximations or involve pseudo-partial waves or projector functions.

3.7. Secular matrix

To construct the secular matrix, we need matrix elements of Hamiltonian and overlap:

$$H_{ij} = \int \chi_i^*(\mathbf{r})[-\nabla^2 + V(\mathbf{r})]\chi_j(\mathbf{r})\, d^3r$$
$$S_{ij} = \int \chi_i^*(\mathbf{r})\chi_j(\mathbf{r})\, d^3r$$

At this stage we need not specify what χ_i is, except to note that it has an envelope part $\chi_i^{(0)}$ that extends everywhere in space, augmented in spheres taking the form Eq. (142). Specifically

$$\chi_i(\mathbf{r}) = \chi_i^0(\mathbf{r}) + \sum_{kL} C_{kL}^{(i)} \left\{ \tilde{P}_{kL}(\mathbf{r}) - P_{kL}(\mathbf{r}) \right\} \tag{144}$$

where $C_{kL}^{(i)}$ are the coefficients (Eq. (141)) that expand the smooth envelope in χ as polynomials $P_{kL}(\mathbf{r})$ (Eq. (132)) and

$$\tilde{P}_{kL}(\mathbf{r}) = \left[A_{kl}\phi_\ell(r) + B_{kl}\dot{\phi}_\ell(r) \right] Y_L(\hat{r}) \equiv \tilde{p}_{kl}(r) Y_L(\hat{r}) \tag{145}$$

$\phi_\ell(r)$ and $\dot{\phi}_\ell(r)$ are the linearised partial waves, Section 2.4. Coefficients A_{kl} and B_{kl} are chosen so that $\tilde{p}_{kl}$ and p_{kl} have the same value and slope at s_R.

3.7.1. Overlap and kinetic energy, and output density

We use Eq. (143) to construct matrix elements of augmented basis functions. Matrix elements of overlap and kinetic energy may be written

$$\int \chi_i^*[-\nabla^2]\chi_j\, d^3r = \int \chi_i^{(0)*}[-\nabla^2]\chi_j^{(0)}\, d^3r + \sum_{kk'L} C_{kL}^{(i)*} \tau_{kk'l} C_{k'L}^{(j)} \tag{146}$$

$$\int \chi_i^* \chi_j\, d^3r = \int \chi_i^{(0)*}\chi_j^{(0)}\, d^3r + \sum_{kk'L} C_{kL}^{(i)*} \sigma_{kk'\ell} C_{k'L}^{(j)} \tag{147}$$

where

$$\tau_{kk'\ell} = \int_S \left\{ \tilde{P}_{kL}[-\nabla^2]\tilde{P}_{k'L} - P_{kL}[-\nabla^2]P_{k'L} \right\} d^3r \tag{148}$$

$$\sigma_{kk'\ell} = \int_{s_R} \left\{ \tilde{P}_{kL}\tilde{P}_{k'L} - P_{kL}P_{k'L} \right\} d^3r \tag{149}$$

The local kinetic energy matrix $\tau_{kk'l}$ is symmetric, even while the individual terms in Eq. (148) are not, because the integral is confined to a sphere. However the surface terms from the true and smooth parts cancel because $\tilde{P}_{kL}$ and P_{kL} match in value and slope there.

Note also that if ϕ_ℓ and $\dot{\phi}_\ell$ are kept frozen, e.g. computed from a free atom as we might do to mimic a PAW or a pseudopotential method, $\tau_{kk'\ell}$ and $\sigma_{kk'\ell}$ are independent of environment. Though we will not pursue it here, it suggests a path to constructing a unique and transferable pseudopotential.

Matrix elements of the potential are more complicated and will be discussed in the next section. For now we take them as given, along with the overlap and kinetic energy. We can diagonalise the secular matrix and obtain eigenfunctions and the output density as a bi-linear combination of basis functions

$$\psi_n(\mathbf{r}) = \sum_i Z_{in}\chi_i(\mathbf{r})$$
$$n^{\text{out}}(\mathbf{r}) = \sum_n w_n |\psi_n(\mathbf{r})|^2 = \sum_{ij} \left\{ \sum_n w_n Z_{in}^* Z_{jn} \right\} \chi_i^*(\mathbf{r})\chi_j(\mathbf{r}) \tag{150}$$

where the sum runs over the occupied eigenstates with w_n the occupation probability. The bi-linear product may be written in a three-component form

$$\chi_i{}^*\chi_j = \chi_i^{0*}\chi_j^0 + \sum_{Rkk'LL'} C_{RkL}^{(i)*} \left\{ \tilde{P}_{RkL}\tilde{P}_{Rk'L'} - P_{RkL}P_{Rk'L'} \right\} C_{Rk'L'}^{(j)} \tag{151}$$

The first term yields a smooth n_0 which extends everywhere, and then two local terms, true and one-centre expansion of n_0 to some L cutoff

$$n^{\text{val}}(\mathbf{r}) = n_0^{\text{val}}(\mathbf{r}) + \sum_{RL} \left\{ n_{1RL}^{\text{val}}(\mathbf{r}) - n_{2RL}^{\text{val}}(\mathbf{r}) \right\} \tag{152}$$

As pointed out before, we expect the higher L components be carried accurately by n_0^{val} even for low $\ell_{\max}$, which provides a simple justification for the pseudopotential and PAW approximations.

3.7.2. Core

The core levels and core density can be computed either scalar-relativistically or fully relativistically from the Dirac equation (Section 2.17). The core density is constructed in one of two ways: *Self-consistent core:* a core partial wave ϕ_ℓ^c is integrated in the true potential, subject to the boundary condition $\phi_c(s_R) = \phi_c'(s_R) = 0$. It was once thought that the advantage of determining the core in the true potential outweighed the inaccuracy originating from the artificial boundary condition at s_R. *Frozen overlapping core approximation:* ϕ_ℓ^c is computed in the free atom and kept frozen—experience shows this to be a better approximation. We develop this case here.

The core density has three components, which add to the valence density:

$$\tilde{n}_0(\mathbf{r}) = \tilde{n}_0^{\text{val}}(\mathbf{r}) + \sum_R g_R^{\text{cn}}(\mathbf{r})$$
$$n_{1R}(\mathbf{r}) = n_{1R}^{\text{val}}(\mathbf{r}) + n_{1R}^{\text{core+nuc}}(\mathbf{r})$$
$$\tilde{n}_{2R}(\mathbf{r}) = \tilde{n}_{2R}^{\text{val}}(\mathbf{r}) + g_R^{\text{cn}}(\mathbf{r}) \tag{153}$$

where

$$g_R^{\rm cn}(\mathbf{r}) = C^G\, G_{L=0}(r_g;\mathbf{r}) + C^H \mathcal{H}_{L=0}(E_c; r_h; \mathbf{r}) \tag{154}$$

The second term is a smooth Hankel function with C^H fixed to fit the spill-out of the core density into the interstitial. It is important that this be taken into account when constructing the total energy and potential. The alternative, to renormalise the core and confine it to the sphere, is a much cruder approximation. This makes it possible for shallow cores to be treated accurately, without needing to include them in the secular matrix (unfortunately at the loss of strict orthogonality between these and the valence states).

The pseudocores $g_R^{\rm cn}$ in $\tilde{n}_0$ and $\tilde{n}_{2R}$ nearly cancel, but they need to be included to construct $V^{\rm eff}$ and the total energy described below. Also the two kinds of $g_R^{\rm cn}$ do not exactly cancel because it extends into the interstitial in the former case, but is truncated in the latter.

Smoothing radius r_g must be small, so the core density is almost completely confined to the sphere. In practice we use the same Gaussians as for $g_R^{\rm val}$ entering into the valence multipole moments (see Eq. (155) below). In that way $g_R^{\rm cn}$ and $g_R^{\rm val}$ can be merged together. The Hankel smoothing radius r_h is chosen independently, but it is also typically small, of order $s_R/2$. The result should depend minimally on the choice of either.

3.7.3. Local orbitals

Inside augmentation spheres, the usual Hilbert space consists of partial waves $\{\phi_\ell,\ \dot{\phi}_\ell\}$. Questaal has the ability to extend this space by one local orbital (LO), to $\{\phi_\ell(\varepsilon_\nu, r),\ \dot{\phi}_\ell(\varepsilon_\nu, r),\ \phi_\ell^z(\varepsilon_z, r)\}$. ϕ_ℓ^z consists of a partial wave evaluated at some energy ε_z far above or far below the linearisation energy ε_ν for ϕ_ℓ. Addition of a LO greatly extends the energy window over which the band structure is valid; see for example Fig. 1, of Ref. [56], where the band structure for Si was compared to LAPW, and to a full (non-linearised) APW. Properties of Questaal's LO are explained in that reference in a *GW* context; here we outline the main features.

LO have one each of the following attributes:

- either core-like or high in energy, with principal quantum number $\mp$ 1 that of valence ϕ_ℓ, respectively;
- either a conventional LO, which consists ϕ_z with some ϕ_ℓ and $\dot{\phi}_\ell$ admixed to make the value and slope vanish at s_R [57], or an extended LO, which consists of ϕ_z with a smooth Hankel function tail that extends continuously and differentiably into the interstitial. The extended LO can be applied only to the core-like case; and
- either a scalar relativistic LO, or a Dirac LO with $\mu = 1/2$. This is not usually important, but it can have consequences for heavy elements; see how it affects LDA and *GW* gaps in Section 3.9. This method was first introduced by Kuneš et al. [58].

High-lying states are not usually important for DFT, because states above E_F do not contribute to the potential. Core states of order $E_F - 2$ Ry and deeper can usually be treated quite adequately as core levels integrated independently of the secular matrix (Section 3.7.2). There are mild exceptions: for high resolution such as needed for the Delta Codes project (Section 3.13), both deep and high-lying LO were used (in the latter case, particularly adding 4*d* LO to the 3*d* transition metals).

GW is more sensitive to states far from E_F than DFT. Including low-lying cores at $E_F - 2$ Ry or even somewhat deeper with LO can modify valence QP levels by ∼0.1 eV. High-lying states are also more important in the *GW* context. ZnO is an extreme case where they were found to shift the gap by several tenths of an eV [59], and many high-lying LOs are required to reach convergence [60].

3.7.4. Effective potential

We seek a three-component form of the total energy functional. Its variational derivative with respect to the density yields an effective potential $V^{\rm eff}$, which should give the first-order variation of the electrostatic and exchange–correlation energy with respect to the trial density. The density has three components, the first interacts with a smooth potential $\tilde{V}_0$, the second with the local true potential V_{1R}, and the third with the local projection of the potential V_{2R}. Correspondingly, the accumulated sums over the first, second, and third terms produce $n_0^{\rm out}$, $n_{1R}^{\rm out}$, and $n_{2R}^{\rm out}$, respectively.

Construction of $V^{\rm eff}$ is complicated by the fact that the electrostatic potential at some point depends on the density everywhere. Therefore it is not possible to construct the potential in a sphere solely from the density inside this sphere, or the smooth potential from $n_0^{\rm val}$ alone. We can solve this difficulty by defining a pseudodensity $\tilde{n^{\rm val}}_0$ which has the same multipole moments inside augmentation sphere R as the true density ${n^{\rm val}}_{1R}$. Suppose the local sphere density $n_{1R}^{\rm val} - n_{2R}^{\rm val}$ has multipole moments q_M, where M = angular momentum. Then define

$$\tilde{n}_0^{\rm val}(\mathbf{r}) = n_0^{\rm val}(\mathbf{r}) + \sum_R g_R^{\rm val}(\mathbf{r})$$
$$g_R^{\rm val}(\mathbf{r}) = \sum_M q_{RM} G_{RM}(\mathbf{r}) \tag{155}$$

where G_{RM} is a Gaussian of angular momentum M centred at R, with unit moment. G_M must be sufficiently localised inside s_R that the potential from the second term at s_R is negligibly different from the potential of a unit point multipole at R (typically $s_R/4$). Results should not depend on r_g, but in practice if it is made too small, numerical integration of the G become inaccurate, while if too large charge is not confined to s_R. Provided it is confined, the compensating Gaussians ensure that the electrostatic potential $V^{\rm es}[\tilde{n}_0^{\rm val}]$ is exact in the interstitial.

The local representation $n_{2R}^{\rm val}$ of the smoothed density must be equally compensated

$$\tilde{n}_{2R}^{\rm val} = n_{2R}^{\rm val} + g_R^{\rm val} \tag{156}$$

The q_{RM} can be decomposed into a sum of partial moments

$$q_{RM} = \sum_{kk'LL'} Q_{Rkk'LL'M}$$
$$Q_{Rkk'LL'M} = \int_{s_R} \left\{ \tilde{P}_{kL}\tilde{P}_{k'L'} - P_{kL}P_{k'L'} \right\} r^m Y_M(\hat{r})\, d^3r \tag{157}$$

By decomposing q_{RM} this way, we can make any possible variation in the local density in the Hilbert space spanned by Eq. (151), and obtain the potential from the electrostatic corresponding to the variation.

Matrix elements of the local potential are determined from variation of the electrostatic energy

$$\pi_{1Rkk'LL'} = \int_{s_R} \tilde{P}_{RkL} V_{1R} \tilde{P}_{Rk'L'}\, d^3r$$
$$\pi_{2Rkk'LL'} = \int_{s_R} P_{RkL} \tilde{V}_{2R} P_{Rk'L'}\, d^3r + \sum_{RM} Q_{Rkk'LL'M} \int_{s_R} \tilde{V}_{2R} G_{RM}\, d^3r \tag{158}$$

The true potential V_{1R} is the response to true partial density $\tilde{P}_{RkL}\tilde{P}_{Rk'L'}$, and $\tilde{V}_{2R}$ the response to the smooth partial density $P_{RkL}P_{Rk'L'} + \sum Q_{Rkk'LL'M} G_{RM}$. These two partial densities have the same multipole moments by construction.

We solve the Poisson equation for each of the three components

$$\nabla^2 \tilde{V}_0^{\rm es} = -8\pi \tilde{n}_0$$
$$\nabla^2 V_{1R}^{\rm es} = -8\pi n_{1R}$$

$$\nabla^2 \tilde{V}^{\rm es}_{2R} = -8\pi \left\{ n_{2R} + \sum_M q_{RM} G_{RM} \right\} \tag{159}$$

All density components must include core (or pseudocore) contributions (Section 3.7.2). Poisson's equation for $\tilde{V}^{\rm es}_0$ is most easily solved by fast Fourier transforming $\tilde{n}_0$ to reciprocal space, scaling each Fourier component n_G by $-8\pi/G^2$ and transforming back. In principle the electrostatic potential can be calculated this way but in practice the compensation Gaussians in $\tilde{n}_0$ can be fairly sharply peaked. Some parts are separated out and evaluated analytically to keep the plane wave cutoff low, as discussed in the next section. $\tilde{V}^{\rm es}_0(\mathbf{r})$ is exact in the interstitial region and extends smoothly through the spheres. $V^{\rm es}_{1R}$ and $\tilde{V}^{\rm es}_{2R}$ can be solved up to an arbitrary homogeneous solution which is determined by the boundary condition at s_R. This adds a term proportional to $r^m Y_M(\mathbf{r})$, which is determined by resolving $\tilde{V}^{\rm es}_0$ into a one-centre expansion at s_R.

Pseudopotentials $\tilde{V}^{\rm es}_0$ and $\tilde{V}^{\rm es}_{2R}$ approximately cancel each other, but not exactly. This is in part because the latter has a finite L cutoff, and in part because $\tilde{V}^{\rm es}_0$ carries the tails of the core densities spilling into the interstitial. These incomplete cancellations are the key to rapid ℓ convergence of this method, and also allow relatively shallow cores not to be included in the secular matrix, with minimal loss in accuracy.

Three component total energy in DFT

The total energy in DFT is comprised of the kinetic energy, electrostatic energy, and exchange–correlation energy

$$E_{\rm tot} = \langle T \rangle + U + E_{\rm xc} \tag{160}$$

Questaal implements two functionals for $E_{\rm tot}$. The traditional Kohn–Sham functional is evaluated at the output density: U and $E_{\rm xc}$ are evaluated from $n_{\rm out}$ generated by the secular matrix, and $\langle T\rangle_{\rm HKS} = \sum_i^{\rm occ} \langle \psi_i | -\nabla^2 | \psi_i \rangle + T_{\rm core}$. $T_{\rm core} = \sum \epsilon_i^{\rm core} - \int n^{\rm core} V^{\rm eff}$ is the kinetic energy of the core. If the core is kept frozen, it can be evaluated from the free atom.

Harris [61] and independently Foulkes and Haydock [62] showed that the kinetic energy can be expressed in terms of $n_{\rm in}$: $U = U[n_{\rm in}]$ and $E_{\rm xc} = E_{\rm xc}[n_{\rm in}]$ and

$$\langle T\rangle_{\rm HF} = \sum_i^{\rm occ} \epsilon_i^{\rm val} - \int n^{\rm val}_{\rm in} V^{\rm eff}_{\rm in} + T_{\rm core} \tag{161}$$

Both $E_{\rm HKS}$ and $E_{\rm HF}$ are calculated, though usually the latter is preferred because it converges more rapidly with deviations from self-consistency. It is nevertheless useful to have both energies, because they are calculated differently in Questaal. U and $E_{\rm xc}$ are computed from different densities, but more, $\langle T\rangle_{\rm HKS}$ and $\langle T\rangle_{\rm HF}$ are calculated in different ways. It is not trivial that at (nominal) self-consistency $E_{\rm HKS} = E_{\rm HF}$. The two always differ slightly; when they differ more than a small amount, it is an indication that some parameter (e.g. plane-wave cutoff) is set too low. Thus, $E_{\rm HKS} - E_{\rm HF}$ is one measure of the convergence of the basis representing the charge density.

In Questaal's three-component framework, $nV^{\rm eff}$ is a sum of three terms bilinear in the three components $\tilde{n}_0\tilde{V}^{\rm es}_0$, $n_{1R}V^{\rm es}_{1R}$ and $\tilde{n}_{2R}\tilde{V}^{\rm es}_{2R}$. $\langle T\rangle$, U and $E_{\rm xc}$ each have independent contributions from three components, for example the contribution to the electrostatic energy from the valence electrons is

$$E^{\rm es} = \frac{1}{2}\int \tilde{n}^{\rm val}_0(\mathbf{r})\tilde{V}^{\rm es}_0(\mathbf{r}) + \sum_R \int_{S_R} \left\{ n^{\rm val}_{1R} V^{\rm es}_{1R}\, d^3r - \tilde{n}^{\rm val}_{2R}\tilde{V}^{\rm es}_{2R} \right\} d^3r \tag{162}$$

The last term cancels most of the first inside the augmentation spheres, except the first retains L components above $\ell_{\rm max}$ to all orders. This is because $\tilde{V}_{2R}$ should be a one-centre expansion of $\tilde{V}_0$ up to $\ell_{\rm max}$, and $\tilde{n}^{\rm val}_{2R}$ should be a one-centre expansion of $\tilde{n}^{\rm val}_0$. Because $V^{\rm es}[\tilde{n}^{\rm val}_0]$ is exact in the interstitial, including at s_R, and it continued inside s_R by adding by $V^{\rm es}_{1R}\, d^3r - \tilde{V}^{\rm es}_{2R}$, the true potential is correct everywhere. Note that the derivative of $E^{\rm es}$ wrt coefficients $C^{(i)}_{Rk'L'}$ in Eq. (151) yield the local potentials Eq. (158), so that the effective one-body Hamiltonian corresponds to the functional derivative of the energy.

The total electrostatic energy must include the core. Write it analogously to Eq. (162) but add the (pseudo) core and nuclear density to $n^{\rm val}$ (Section 3.7.2)

$$U = \tilde{U}_0 + \sum_R \{U_R - \tilde{U}_R\} \tag{163}$$

The first term is the electrostatic energy of the smooth density, evaluated on a uniform mesh. It involves a valence and pseudo-(core+nucleus) term

$$\tilde{U}_0 = \tilde{U}^{\rm val}_0 + \tilde{U}^{\rm c+n}_0$$
$$\tilde{U}^{\rm val}_0 = \frac{1}{2}\int d^3r\, \tilde{n}^{\rm val}_0(\mathbf{r})\, \tilde{V}^{\rm es}(\mathbf{r})$$
$$\tilde{U}^{\rm c+n}_0 = \sum_R (C^G_R - Z_R) V^G_{R0} + C^H_R V^H_R$$

where

$$V^G_{RL} = \int d^3r\, G_{RL} \tilde{V}^{\rm es}[\tilde{n}_0]$$
$$V^H_R = \int d^3r\, \mathcal{H}_{R0} \tilde{V}^{\rm es}[\tilde{n}_0]$$

Besides being needed for $U^{\rm c+n}_0$, V^G_{RL} is used in assembling $\tilde{U}^{\rm val}_0$. $\tilde{U}^{\rm c+n}_0$ and $\tilde{U}^{\rm val}_0$ must be evaluated with some care to avoid numerically integrating sharply peaked functions (core pseudodensity and the multipole contribution to $\tilde{n}_0$ (Eq. (155)). V^G_{RL} and V^H_R are evaluated by splitting the potential into two terms $\tilde{V}^{\rm es}[\tilde{n}_0] = \tilde{V}^{\rm es}[n_0] + \{\tilde{V}^{\rm es}[\tilde{n}_0] - \tilde{V}^{\rm es}[n_0]\}$. Integrals of the second term can be performed analytically (the interested reader is referred to Section XI of Ref. [31]), thus avoiding integrals of sharply peaked functions. All of the other integrals are much smoother and are evaluated on the grid of G vectors.

U_R and $\tilde{U}_R$ are readily integrated in the sphere together with the valence parts in Eq. (162).

The exchange–correlation potential is made analogously with the electrostatic potential:

$$E_{xc} = \int n_0 \epsilon_{\rm xc}[n_0]\, d^3r + \sum_R \int_{s_R} \left\{ n_{1R}\epsilon_{\rm xc}[n_{1R}] - n_{2R}\epsilon_{\rm xc}[n_{2R}] \right\} d^3r \tag{164}$$

It is also divided into valence + core contributions.

We reiterate that differences $\int n_{1R} V^{\rm es}_{1R}\, d^3r - \tilde{n}_{2R}\tilde{V}^{\rm es}_{2R}$ and $\int n_{1R}\epsilon_{\rm xc}(n_{1R}) - n_{2R}\epsilon_{\rm xc}(n_{2R})$ converge much more rapidly with ℓ-cutoff than either term separately, and as a consequence the three-component augmentation is much more efficient than the standard one.

3.7.5. Forces

Our original formulation [31] derived a simple expression for the derivative of $H_{\rm HF}$ when a nucleus changes from $\mathbf{R}$ to $\mathbf{R} + \delta\mathbf{R}$, assuming that the partial waves $\phi_{R\ell}$ shift rigidly with $\delta\mathbf{R}$. (`lmf` has a "frozen ϕ" switch, which if imposed, guarantees $\phi_{R\ell}$ is rigid. In practice we have found that the differences are small.)

The original expression yielded forces that converge only linearly with deviations $\Delta n = n_{\rm out} - n_{\rm in}$ from self-consistency. It is possible to achieve faster convergence in Δn, by adding a correction $\delta V \Delta n$, where δV is determined from a change $\delta n_{\rm in}$ in

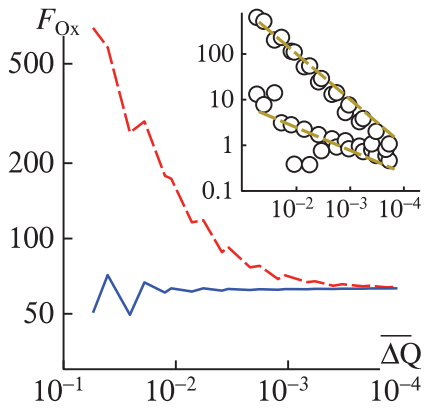

Fig. 12. Harris force F_{Ox} (mRy/au) on O neighbouring Co in Co-doped TiO_2. $\overline{\Delta Q}$ is a measure of the RMS change $n_{out} - n_{in}$ integrated over the unit cell. Solid and dashed lines are Eq. (165) with and without the third term. Inset shows deviation in F_{Ox} from the converged value. Yellow lines show $\overline{\Delta Q}$ and $(\overline{\Delta Q})^2$ as guides to the eye.

n_{in}, but there is no unambiguous way to determine δn_{in}. If the static susceptibility (Section 6) were calculated, a good estimate could be made. But $\chi(\mathbf{r}, \mathbf{r}', 0)$ is a second derivative of H_{HF}: it is far more cumbersome to make than H_{HF} itself, and approximation of χ by models such as the Lindhard function worked less well than hoped. A practicable, and simple alternate ansatz is to assume that δn_{in} is given by a change in the Mattheis density construction (superposition of free-atomic densities), arising from $\delta\mathbf{R}$. This has worked very well in practice: it dramatically improves the convergence of the forces with iterations to self-consistency. Here we omit the derivation but refer to the original paper [31], and also to Appendix B of Ref. [52] for an alternate derivation including the correction term. The final expression for shift of nucleus R from $\mathbf{R} \rightarrow \mathbf{R} + \delta\mathbf{R}$ is

$$\delta E_H = \int \tilde{V}_0 \, \delta g_R^{cn} + \delta^R \sum_i^{occ} \epsilon_i^{val} + \delta V_{in}(n_{out} - n_{in}) \tag{165}$$

$\tilde{V}_0$ is the sum of the smooth electrostatic and exchange–correlation potential. This term describes the force of the smooth density on the Gaussian lumps which represent the core and the nucleus at each site, and is the analogue of the Hellmann–Feynman theorem.

The eigenvalue shifts account for a change in the atom-centred basis set in the shift in **R**. This is the price for taking a tailored basis whose Hilbert space changes when **R** shifts. Here the smooth mesh potential is fixed and augmentation matrices τ, σ, and π (Eqs. (148), (149) and (158)) shift rigidly with the nucleus. Gradients only include redistribution of plane waves owing to the shifts in envelope functions, and in the expansion coefficients $C^{(j)}_{Rk'L'}$ (Eq. (144)).

The last term is the correction noted above. In future, it will probably be replaced by an efficient calculation of χ; meanwhile the Mattheis-shift ansatz works rather well. Fig. 12 shows the force F_{Ox} on an O neighbouring Co substituted for Ti in a 12-atom unit cell of TiO_2, as `lmf` iterates to self-consistency. The deviation ΔF from the self-consistent force is vastly smaller with the correction term (compare solid and dashed lines), and F_{Ox} is reasonable already for the Mattheis construction (first iteration). The inset compares the two kinds of ΔF with iteration to self-consistency to show that without and with the correction, ΔF scales as $\Delta F_{Ox}^{uncorr} \sim \overline{\Delta Q}$ and $\Delta F_{Ox}^{corr} \sim (\overline{\Delta Q})^2$, the latter being the same rate of convergence as the energy itself.

3.8. LDA+U

The LDA+U method, introduced originally by Anisimov et al. [63–65] for open shell d- or f-shell materials including on-site Coulomb and exchange interactions by means of the Hubbard model is implemented in both the `lm` and the `lmf` codes following the rotationally invariant approach described in [66]. Rotational invariance means that the occupations of the different d_m orbitals are specified by a density matrix $\mathbf{n}^\sigma = n^\sigma_{mm'}$ independent of the specific choice of Cartesian axes defining the spherical harmonics. For brevity we refer to the orbitals on which U is applied as the d orbitals although the code is written sufficiently general that U terms can be applied to any nl of choice and even on multiple sets of orbitals.

The on-site Coulomb interactions are added to the LSDA total energy and a double-counting term is subtracted.

$$E^{LDA+U}[\rho^\sigma(\mathbf{r}), \mathbf{n}^\sigma] = E^{LSDA}[\rho^\sigma(\mathbf{r})] + E^U[\mathbf{n}^\sigma] - E_{dc}[\mathbf{n}^\sigma] \tag{166}$$

Several schemes for the double counting have been proposed [63,64,66–68] and are implemented in `lmf` but the prevailing approach is the fully localised limit (FLL) in which the double-counting term and U-terms cancel for the fully localised (atomic) limit in which the density matrix becomes a diagonal matrix with integer occupations. This means that in principle the total energy is already well described in the LSDA in the atomic limit but the orbital energies are not. The task of the LSDA+U approach is to optimise the density matrix when the orbitals on which U is applied are allowed to hybridise with the other orbitals in the system. It hence describes strictly speaking the orbital polarisation in the system. Note that the total energy is then separately a functional of the spin-dependent electron density and the local density matrix on the orbitals for which U terms are added. The density matrix or orbital occupations also influence the spin-charge densities and so the two contributions are not really independent. The U terms in LSDA+U are treated in the static mean-field Hartree–Fock approximation in contrast to the DMFT method where they are treated dynamically. The expressions for the $E^U[\mathbf{n}^\sigma]$ part of the total energy can be found in [66] and are given in terms of density matrices $n^\sigma_{mm'}$ and matrix elements of the Coulomb interaction $\langle m, m''|V_{ee}|m', m'''\rangle$. They fully take into account the rotational aspects of the Y_{lm} of the d-states included. The double counting term, on the other hand, is written in terms of average Coulomb U and exchange J parameters and involves only the trace of the density matrix $n^\sigma = Tr(\mathbf{n}^\sigma)$. [66] U and J are the screened Coulomb and exchange parameters, which can for example, be estimated by a separate self-consistent supercell calculation treating the atoms in which the d-orbital occupation is changed as an impurity problem [69]. or by means of the linear response approach [70]. Most often they are treated empirically and U is adjusted to spectroscopic splittings, for example.

The bare Coulomb interaction matrix elements can be evaluated exactly in terms of the Slater $F^k(ll)$ integrals and combinations of Gaunt coefficients as worked out for instance in Condon and Shortley [71]. The Slater $F^k(ll)$-integrals,

$$F^k(ll) = \int\int r_1^2 dr_1 r_2^2 dr_2 R_\ell(r_1)^2 R_\ell(r_2)^2 (r_<^k / r_>^{k+1}) \tag{167}$$

where $r_<$, $r_>$ are the smaller and larger or r_1 and r_2 could easily be calculated directly in terms of the partial waves inside the spheres. However, this would then not take into account the screening of the Coulomb interaction by the other orbitals in the system. Therefore it is customary to use the relation between the average U and J to the $U_{mm'}$ and $J_{mm'}$

$$U = \frac{1}{(2\ell+1)^2} \sum_{mm'} U_{mm'} = F^0$$

$$U - J = \frac{1}{2\ell(2\ell+1)} \sum_{mm'} (U_{mm'} - J_{mm'}) \tag{168}$$

given by Anisimov et al. [64] and where $U_{mm'} = \langle mm'|V_{ee}|mm'\rangle$, $J_{mm'} = \langle mm'|V_{ee}|m'm\rangle$. Using the tables of Condon and Shortley [71] one finds for d electrons

$$14J = F^2(dd) + F^4(dd) \tag{169}$$

and for f electrons

$$3J = \frac{2}{15}F^2(ff) + \frac{1}{11}F^4(ff) + \frac{1}{858}F^6(ff) \tag{170}$$

For p-orbitals $J = F^2(pp)/5$ and for s-orbitals $J = 0$. Examining the tabulated values of the Slater F^k integrals from Hartree–Fock calculations, one finds that the ratios $F^4(dd)/F^2(dd) \approx 0.625$ and $F^6(ff)/F^2(ff) \approx 0.494$, $F^4(ff)/F^2(ff) \approx 0.668$ are approximately constant. Using these fixed ratios and the relation to the average exchange integral J one then can fix the F^k values. The advantage of doing it this way is that the screening of the J and U can then be taken into account and there are only two empirical parameters. On the other hand one finds in practice that while U is strongly reduced from the atomic bare F^0, the J is approximately unscreened, so one might as well use the unscreened directly calculated F^k for $k \geq 2$. From LDA calculated atomic wave functions, we find $F^4(dd)/F^2(dd) = 0.658 \pm 0.004$ for 3d atoms, $F^4(ff)/F^2(ff) = 0.685 \pm 0.001$, $F^6(ff)/F^2(ff) = 0.518 \pm 0.001$ for rare-earth RE^{3+}. (Until recently, Questaal used a different convention; see note [72].)

By minimising the LDA+U total energy with respect to the density matrix elements, one obtains a non-local potential $V^{\sigma}_{mm'}$ which can also be found in [66]. In the ASA `lm`-version it is straightforward to add this nonlocal potential directly to the Hamiltonian inside the spheres because the orbitals defining the matrix $V^{\sigma}_{mm'}$ are just the single lm channel partial waves inside the sphere. It is a little more complex in `lmf`. Essentially, we now add the operator $|\phi^{\sigma}_{lm}\rangle V^{\sigma}_{mm'}\langle\phi^{\sigma}_{lm'}|$. We add the corresponding augmentation matrix elements both for the ϕ and $\dot{\phi}$ and for local orbitals if these are present for this ℓ channel.

The calculation within LDA+U starts from a set of initial occupation numbers of the m orbitals for each channel ℓ for which U and J parameters are provided. These define the initial density matrix. The calculation then proceeds by making this density matrix self-consistent simultaneously with making the standard spin density $\rho^{\sigma}(\mathbf{r})$ self-consistent. We note that the configuration one converges to may depend on the starting density matrix and hence one should in principle consider different starting points and find the one which minimises the energy or use guidance from physical insight. In rare-earth (RE) based systems one finds for example that the lowest energy corresponds to the configuration which obeys Hund's rules [73].

We have here described only the FLL. The around-mean-field approach [63,67] and a mixture of the two [68] are also implemented and make use of the same ingredients. Recently, Keshavarz et al. [74] argued that adding U to spin-unpolarised LDA, thus LDA+U may have advantages over adding them to LSDA, for example to extract inter-atomic exchange interactions independent of the reference system used (AFM or FM). This is presently not supported in the lm and lmf codes but might be useful to add in the future. Note that in Ref. [64] also a LDA+U rather LSDA+U approach was used.

A simpler version of the LDA+U is described by Dudarev et al. [75]. In that case, only a single parameter, namely $\tilde{U} = U - J$ comes into play. One can use this scheme in the codes by setting $J = 0$ and adjusting the provided $\tilde{U}$ parameter. This means essentially that we neglect the orbital polarisation due to the anisotropic Coulomb interaction exchange terms but keep the Hartree–Fock like feature that empty states feel a different potential from occupied states. In fact, in this case the additional potential is

$$V^{\sigma}_{mm'} = \tilde{U}[\tfrac{1}{2}\delta_{mm'} - n^{\sigma}_{mm'}] \tag{171}$$

Or if the density matrix is actually diagonal $V^{\sigma}_{m} = \tilde{U}[\frac{1}{2} - n^{\sigma}_{m}]$. This means that when a spin–orbital $m\sigma$ on a given site is empty it is shifted up by $\tilde{U}/2$ and when it is filled, it is shifted down by $\tilde{U}/2$. This corresponds to the simplest form of the LDA+U formalism. It allows one for example to include a shift of a fully occupied d band. This is useful for example in Zn-containing systems. In LDA, the binding energy of these orbitals is underestimated. The deeper a band lies below the Fermi level, the higher its downward shift by the self-energy and this can be simply mimicked by the LDA+U method in this form. It has the effect for example of reducing the hybridisation of the Zn-3d with the valence band maximum and slightly increases the gap, affects things like band-offsets and the valence band maximum crystal field and spin–orbit induced fine splitting.

One can even use this approach for s or p electrons and thereby shift up the mostly cation-s like conduction band minimum in a semiconductor to adjust the gap. While this is of course purely empirical and shifting the gaps for the wrong reason, and hence far less accurate than the QSGW approach, it is sometimes useful in defect calculations because a defect level that would otherwise lie as a resonance in the conduction band can now become a defect level in the gap and allow one to properly control its occupation for different charge states of the defect. Using a U on p-orbitals of anions like O or N, this can also have the effect that the empty defect levels pushed out the valence band and localising on a single p orbital are pushed deeper into the gap. The essential function of the LDA+U terms in these applications is to introduce a Hartree–Fock like orbital dependence of the potential. This is important to reduce the self-interaction error of LDA or GGA and allows one in a simpler and much less expensive way to simulate what a hybrid functional would do. If one picks these shifts carefully to adjust the bulk band structure to the QSGW bands, it provides a rather efficient way to study defects with a corrected band gap and specific orbital self-interaction, two of the main errors of LDA plaguing defect calculations. An example of this can be found in Boonchun et al. [76].

3.9. Relativistic effects

The radial solvers all solve by default a scalar Dirac equation, which incorporates the dominant relativistic effects. All the DFT codes (`lmf`, `lm`, `lmgf` and `lmpg`) have the ability to incorporate spin–orbit coupling perturbatively. In the band codes the perturbation is straightforward (See Section 2.8.2); the Green's function scheme is more subtle, particularly with respect to third order parameterisation, but a formulation is possible (Section 2.16). For `lmf`, the $\mathbf{L}\cdot\mathbf{S}$ term is added to the true local potential V_{1R} (see Section 3.7.4). As of this writing, only the spin diagonal part of the output density is kept in the self-consistency cycle.

Additionally the ASA codes `lm`, `lmgf` and `lmpg` have a fully relativistic implementation (Section 2.17). The full-potential code `lmf` does not as yet; however it does have a facility to compute the core levels fully relativistically. It follows a method similar to that developed by Ebert [42].

As of this writing, the QSGW code allows spin–orbit coupling only in a very restricted manner. Typically SOC has only a minor effect on the self-energy Σ: this is because SOC is a modification to the kinetic energy and only affects the potential in a higher order perturbation. We have found very good results by generating QSGW self-energies leaving out SOC all together, and then as a post-processing step include it in the band structure.

Table 4
Spin orbit splitting Δ_0 of the valence band at Γ, and energy bandgaps computed in the LDA and by QS*GW*, with a scalar relativistic (SR) local orbital on the cation and anion p states, and a Dirac $p_{1/2}$ local orbital. In the zincblende semiconductors the two band edges are at Γ, but in PbTe case they are both at L (Fig. 13). n are the principal quantum numbers of the cation and anion local p orbital; $\gamma = [1-(Z/c)^2]^{1/2}$.

Material	n_c n_a	γ_c γ_a	LDA		QS*GW*	
			Δ_0	E_G	Δ_0	E_G
GaAs (SR)	5	0.97	0.34	0.24	0.33	1.77
GaAs ($p_{1/2}$)	5	0.97	0.35	0.24	0.34	1.75
GaAs (expt)			0.34	1.52	0.34	1.52
ZnSe (SR)	5	0.98	0.39	0.95	0.39	3.05
ZnSe ($p_{1/2}$)	5	0.97	0.40	0.95	0.40	3.00
ZnSe (expt)			0.40	2.82	0.40	2.82
CdTe (SR)	6	0.94	0.84	0.26	0.83	1.83
CdTe ($p_{1/2}$)	6	0.93	0.92	0.26	0.88	1.77
CdTe (expt)			0.90	1.61	0.90	1.61
PbTe (SR)	7	0.80	1.12	−.16	1.00	0.40
PbTe ($p_{1/2}$)	6	0.93	1.30	−.57	1.06	0.22
PbTe (expt)				0.19		0.19

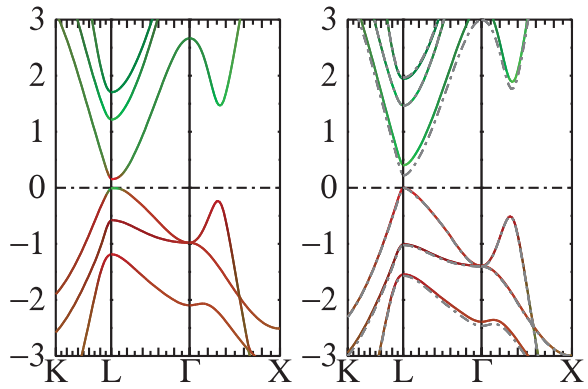

Fig. 13. Energy band structure of PbTe in eV for (left) the LDA and (right) QS*GW* approximations. Colours correspond to projections onto $L_6^{\pm}$ symmetry. L_6^+ is comprised of Pb s and the t_{2g} part of the d orbital, and Te p. L_6^- is comprised of Te s and the t_{2g} part of the d orbital, and Pb p. The dashed lines show the effect of using a Dirac $p_{1/2}$ local orbital instead of the usual scalar relativistic one. Except for the addition of the Dirac $p_{1/2}$ orbital, the basis is similar to that used in Ref. [78]. It is a slight enlargement over the default basis, including g orbitals on Pb and Te and a 6s local orbital on Te. These extra orbitals widen the gap relative to the default basis by ∼0.1 eV.

Additionally, `lmf` has a special mode where the only diagonal parts of $\mathbf{L}\cdot\mathbf{S}$ (see Section 2.8.2) are added to the one-particle Hamiltonian. After diagonalisation the spin off-diagonal parts are used to modify the one-particle levels in a special kind of perturbation theory, as described in the Appendix of Ref. [77]. This has the advantage that the eigenvectors are kept spin-diagonal, which significantly reduces the cost of the *GW* code. Tests show that at the LDA level the modification to the band structure is very similar to that of the full $\mathbf{L}\cdot\mathbf{S}$, and for purposes of determining the effect on Σ it should be quite reliable except in very special cases. Once a modified Σ has been found (as a one-shot correction to QS*GW* computed without SOC), `lmf` can be run with the full $\mathbf{L}\cdot\mathbf{S}$, including the modification of Σ approximately in the manner just described. We have found the difference to be negligible except when elements are very heavy: $CH_3NH_3PbI_3$ is the most extreme case we have found so far: the effect of SO on the band structure through changes in Σ reduced the gap by 0.1 eV [77]

Finally `lmf` has an ability to fold in approximately the effects of the proper Dirac Hamiltonian on the valence states. The difference between the full Dirac and scalar Dirac partial waves can be important for very small r, where the SOC contributions are the largest (see $\xi(r)$ in Section 2.8.2). For small $\mathbf{r}$, the true Dirac wave function varies as r^γ, $\gamma^2 = \kappa^2 - (2Z/c)^2$, whereas in the scalar Dirac case it varies as r^ℓ. As a result, matrix elements $\xi(r)$ are underestimated. The effect is typically very small, but it becomes non-negligible in heavy semiconductors such as CdTe. To ameliorate this error, `lmf` has the ability to use a fully relativistic partial wave in place of a standard local orbital, choosing $\mu = -1/2$. Most important is the $p_{1/2}$ state with $\kappa = 1$.

SO coupling is well described for light semiconductors such as GaAs and ZnSe (see Table 4), as Z increases γ deviates progressively from unity and the discrepancy with experiment increases. It is already significant for CdTe, and in PbTe the effect exceeds 0.1 eV. Replacing the scalar Dirac LO with the Dirac $p_{1/2}$ LO significantly decreases this error in CdTe. In PbTe, the valence band edges are both at L: the two band edges have symmetry L_6^+ and L_6^-, respectively (Fig. 13). The LDA gap is nominally positive, but L_6^+ and L_6^- are inverted very near L as can be seen by the change in colours at the band edges; thus the gap is actually inverted. Anticipating the discussion in Section 4, if a diagonal-only *GW* were carried out on top of this, the L_6^+ and L_6^- would order properly, but the wrong topology would manifest itself as a band crossing near L, as shown for Ge in Fig. 6 of Ref. [56]. QS*GW* orders L_6^+ and L_6^- correctly, somewhat overestimating the gap [78]. Adding a Dirac $p_{1/2}$ local orbital reduces the gap by nearly 0.2 eV. In the LDA the gap at L *widens* as a consequence of the fully relativistic partial wave (Table 4), another reflection of the inversion of L_6^+ and L_6^-.

Table 5
Nature and size of different basis configurations for the case of $SrTiO_3$.

Basis	Character of $\mathcal{H}_L$	$N(\mathcal{H}_L)+N$(LO)
Null	no $\mathcal{H}_L$	6
Hyper	*sp* on O only	18
Small	*spd* on all atoms	51
Large	+2nd *spd* (*spd* on Ti)	81

3.10. The PMT method

In the introduction it was noted that the LAPW and Questaal's generalised LMTO method are essentially the same except for the choice of envelope functions. LMTO's are much more compact, and require a much smaller Hilbert space, but they suffer from problems with basis set completeness. An obvious alternative is to combine the two basis sets; this is the "PMT basis" [54]. This can be accomplished in practice quite neatly, because Questaal's standard $\mathcal{H}_L$ require a one-centre expansion, Eq. (141), more general than simple LMTO's (Eq. (1)). Plane waves can be expanded in polynomials P_{kL} in a similar manner, as can other kinds of envelope functions. Thus the method provides a unified framework to seamlessly mix different kinds of envelope functions. Expressions for total energy, forces, assembling output density, etc, remain unchanged. The procedure is described in more detail in Ref. [54]; here we focus on the primary strengths and weaknesses of the method, using the total energy calculated for $SrTiO_3$ shown in Fig. 14 for discussion. It was redrawn from Ref. [54]. Four basis sets were chosen, ranging from pure LAPW ('null') to a moderately large basis of 75 smooth Hankels $N(\mathcal{H}_L)$ + 6 local orbitals (see Table 5).

Advantages of PMT. As expected, Questaal's purely atom-centred basis is vastly more compact than a pure LAPW basis (Fig. 14). Even a tiny basis of just O *sp* states dramatically improves on the convergence of the LAPW basis. PMT offers a marked advantaged from LAPW perspective: by augmenting that basis by a few functions. From the atom-centred perspective Questaal's standard

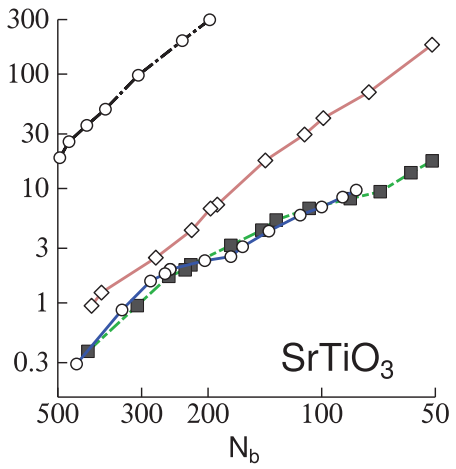

Fig. 14. Deviation ΔE from fully converged total energy/atom, in mRy, *vs* total number of basis functions (PW + $\mathcal{H}_L$ + LO) N_b^Γ at Γ. Black-dashed, peach-solid, green-dashed, and blue-solid correspond respectively to the null, hyper, small, and large basis sets. ΔE is drawn on a log–log scale as a function plane-wave cutoff, denoted as $1/N_b$. N_b is the total number of basis functions PW + $\mathcal{H}_L$ + LO at the Γ point. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

moderately large basis misses about 5 mRy/atom in total energy, showing that it is not complete. Finally the figure shows that adding $\mathcal{H}_L$ continues to dramatically improve the convergence until the standard minimal basis of $\mathcal{H}_L$ is reached (single *spd* function on each atom). Beyond that initial rapid gain, it does not seem to matter much whether the basis is increased by adding PW or more $\mathcal{H}_L$ (compare blue-solid and green-dashed lines).

Drawbacks of PMT. While they are far more efficient than ordinary LAPWs, PMTs show less promise than was initially hoped for.

- Compactness: LAPWs are not "intelligent", i.e. tailored to the potential, and therefore convergence is slow in the basis size. Also their extended range precludes the advantages of short-ranged functions.
- Over-completeness: PMT has difficulties when $\mathcal{H}_L$ and PWs are both sizeable. This is because they span much the same Hilbert space. Workarounds (reducing the Hilbert space by projecting out parts that contribute to a small overlap) have been implemented, but they are inefficient and somewhat finicky—e.g. energy may not vary smoothly with the change in a nuclear coordinate.
- Interpolating Σ_0: in QS*GW* context, the ability to interpolate the QS*GW* self-energy Σ_0 is degraded. A workaround has been made for this problem also (Ref. [79]) by projecting Σ_0 onto the MTO part of the eigenfunctions, interpolating it, and re-embedding. Information is lost in the projection/re-embedding step, resulting in ambiguities. A much better way to address difficulties with interpolation is to construct very short-ranged, compact functions. This is shown clearly in Section 3.12.

In summary, PMT does best when treated as a slight augmentation of either the LAPW limit, or the MTO limit. Adding a few PWs with a low cutoff (2 Ry) was useful in our participation in the Delta Codes project (Section 3.13) particularly for the molecular solids such as N_2 which consist of 90% interstitial, or even more. Questaal's pure MTO basis does not handle those compounds that accurately.

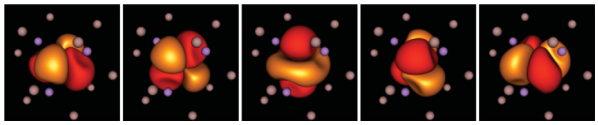

Fig. 15. Screened d envelope functions, xy, yz, $3z^2-1$, xz, and x^2-y^2, for a zincblende lattice. See also Fig. 5.

3.11. Floating orbitals

While the PMT method can make the basis nearly complete, it causes severe difficulties when interpolating the QS*GW* Σ_0 to an arbitrary k. An alternative is to add "empty" sites — points outside augmentation spheres where extra $\mathcal{H}_L$ can be added to the basis. These sites have no augmentation spheres; they only enhance completeness in the interstitial. They are more *ad hoc* than plane waves (Section 3.10) particularly because there is no systematic path to convergence. Nevertheless they can be implemented efficiently; and Questaal has an automatic procedure to locate points to fill voids in the interstitial. In practice we use floating orbitals to converge QS*GW* self-energies with respect to the single-particle basis, in open systems. Their effect at the LDA level is usually not large, unless systems are very open or very accurate total energies are required. But for QS*GW*, where the potential depends on unoccupied states as well as occupied ones, it makes more of a difference, e.g correcting the fundamental gap in Si by of order 0.1 eV. We expect that the need for floating orbitals in such systems will be obviated by the new JPO basis. A precursor to it is discussed next.

3.12. Screened, short-ranged orbitals in the full-potential framework

Building on the success of the "screening" of LMTOs in the ASA (Section 2.9), we have developed an analogue within the full-potential LMTO framework. As in the ASA case, "screening" does not alter the Hilbert space but renders basis functions short ranged. Screening is a precursor to the next-generation JPO basis, which will alter the Hilbert space, rendering it significantly more accurate for the same rank of Hamiltonian. JPOs are still in development; the method will be presented in a future work. They have the following properties:

- they solve the SE with a (nearly) optimal number of basis functions for a given accuracy in the four dimensions ($\mathbf{r}$, E);
- they are very short ranged (see Figs. 5 and 15); and
- they are atom centred with a definite L character on their own MT sphere.

This provides a framework for exploiting the many advantages inherent in a short ranged basis sets. They can be used as projectors, replacing Wannier functions; they can be used to construct minimal Hamiltonians; near-sightedness can be exploited to make very efficient $O(N)$ solvers in DFT and $O(N^3)$ solvers in *GW* and *GW*+BSE. That the basis functions are associated with a definite L character is important, for example, when singling out a correlated subspace.

Screening of the $\mathcal{H}_L$ works in a manner similar to screening traditional H_L because $\mathcal{H}_L$ asymptotically approaches a H_L for large r (Fig. 10). $\mathcal{H}_L - H_L$ decays approximately as a Gaussian of radius r_s: for traditional values of r_s ($r_s \lesssim 2s/3$) it becomes negligible beyond first-neighbour distances. The screened $\mathcal{H}_L^\alpha$ employ the expansion coefficients S^α in the same manner as in the ASA (Section 2.9), so they have essentially the same range. What is sacrificed is the reinterpretation of screening in terms of hard core radius (Section 2.10.5), but this has no effect because the Hilbert space is unchanged. JPOs improve on the screened $\mathcal{H}^\alpha$

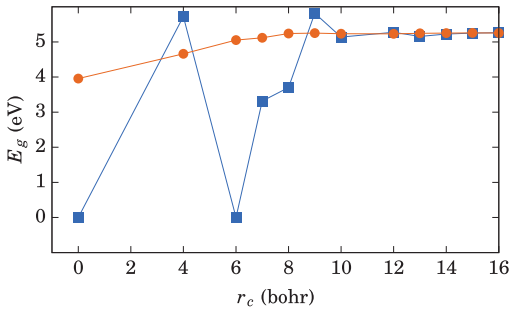

Fig. 16. NiO bandgap in eV as a function of the range truncated self-energy represented in the conventional and screened basis sets. The first point contains onsite terms only, the second adds first neighbours etc. The conventional basis shows erratic behaviour until at least the fourth neighbours are present.

by restoring this condition and exploiting it to make the kinetic energy continuous everywhere.

Screening makes it possible to avoid Ewald summation otherwise necessary in the unscreened FP case (Appendix C). Also, one-centre expansions are made more efficient. By writing $\mathcal{H}^\alpha = \{\mathcal{H}^\alpha - H^\alpha\} + H^\alpha$, the first term which involves $\mathcal{H}$, must be expanded in the more cumbersome polynomials P_{kL} (Eq. (144)). In the screened case the difference becomes short-ranged and can be dealt with efficiently in real space. The latter can use the simpler expansion Eq. (42).

A restriction in transformation is that all functions of a given κ must share the same Hankel energy, while the traditional basis allows arbitrary choice for each orbital. However, this restriction is fairly mild because of the extra flexibility $\mathcal{H}_L$ have through the choice of r_s, which the H_L do not. The r_s strongly affect the kinetic energy near the MT boundary (Fig. 10), and the results are more sensitive to that choice than to the Hankel energy.

One large advantage of the transformation is that $\mathcal{H}^\alpha$ becomes short enough ranged so that matrix elements

$$\Sigma^\alpha_{0;RL,R'L'} = \left\langle \mathcal{H}^\alpha_{RL} | \Sigma_0 | \mathcal{H}^\alpha_{R'L'} \right\rangle$$

are limited by the range of the physical $\Sigma_0(\mathbf{r}, \mathbf{r}')$, and not by the basis set. $\Sigma^\alpha_{0;RL,R'L'}$ turns out not to be very long ranged, as suggested some time ago by Zein et al. [80]. Fig. 16 shows how the bandgap in NiO evolves with range cutoff R_{cut}, defined as follows. $\Sigma^\alpha_{0;RL,R'L'}$ is initially computed without range truncation. Then matrix elements of $\Sigma^\alpha_{0;RL,R'L'}$ are set to zero for $|\mathbf{R} - \mathbf{R}'| > R_{\text{cut}}$, and the band structure is calculated. As Fig. 16 shows, the bandgap E_G converges slowly with R_{cut} in the traditional $\Sigma^{\alpha=0}_{0;RL,R'L'}$ case, but for $\Sigma^\alpha_{0;RL,R'L'}$, E_G is already reasonably converged including first and second neighbours. Note that for $R_{\text{cut}} \to \infty$, the two methods yield the same result, as the screening transformation does not change the Hilbert space.

One powerful feature of `lmf` is its ability to interpolate Σ_0. As noted in Section 4.2, some compromises must be made because the high-energy parts of $\Sigma_0^{nn'}$ do not interpolate well, and elements above an energy threshold (of order 2 Ry) must be set to zero. The screening mitigates this difficulty; it appears that the energy threshold can be taken to ∞. The difference is not large, but in small bandgap cases where the precision of the method is very high, agreement with experiment may be limited by this difference.

The short range of $\Sigma^\alpha_{0;RL,R'L'}$ can be used to much advantage: we realise from the properties of Fourier transforms that the number of k points directly translates into the number of neighbours in real space. Because $\Sigma^\alpha_{0;RL,R'L'}$ is short ranged, it converges faster in the k mesh than $\Sigma^{\alpha=0}_{0;RL,R'L'}$, which saves significantly on computational effort. Also explicitly semilocal algorithms can be constructed, greatly reducing the cost to make Σ_0.

At present the short ranged screened smooth Hankel functions shine the most when used to represent the QSGW self-energy $\Sigma(\omega, q)$ and interpolate it at intermediate q points not directly calculated in the heavy GW step. The interpolation process is significantly simplified and ambiguity is removed because the high energy Σ no longer needs diagonal approximation, there is no need to define energy threshold and there is no loss of information due to discarding the off-diagonal parts. This does add to the computing effort for directly computed QSGW $\Sigma(\omega, q)$ because no states are excluded and matrix sizes increase respectively. To counter that, significantly fewer $\Sigma(q)$ need to be computed exactly, in our experience an equivalent result can be achieved by using only a half or in cases even a quarter of the original BZ sampling in each direction, for cells with few symmetry operations the savings add up quickly considering that the GW walks over pairs of q points (quadrupling the total number of points for low symmetry cells). Performance can be further improved by still utilising the diagonal approximation to $\Sigma(\omega, q)$ because the interpolation is excellent and the high energy part only weakly affects states near E_F. To show the quality of the QSGW $\Sigma(\omega, q)$ expansion in the new, screened basis and the conventional unscreened, we plot the dependence of the NiO bandgap as a function of the truncated self energy in Fig. 16.

At the LDA level, the better spatial localisation of screened basis function paves the way to efficient, real-space assembly of matrices (1 and 2 particle) while maintaining, and in cases improving, accuracy. The screened functions are also significantly less linearly dependent. The implementation is heavily vectorised and at present can handle a couple hundred atoms on a single node, with the scalability properly utilised the numbers can be an order of magnitude larger. For very large systems, it will be possible to construct $O(N)$ methods, but empirically it appears so far that systems of a few hundred atoms are probably still too small for such methods to be advantageous.

To summarise, our current screening transformation offers advantages of short-ranged basis sets. It is not yet optimally compact. However, the JPO basis, to be reported on soon, we believe will be nearly the most compact (optimal convergence for a given rank of Hamiltonian, and short range). We believe these advantages will be very significant, and obviate the need for enhancements such as plane waves (Section 3.10) or floating orbitals (Section 3.11) to obtain high accuracy.

3.13. Delta Codes validation exercise

The Delta Codes project [6] is a continuing effort to mutually validate the different electronic structure codes used in condensed matter and materials modelling. The first comparison has focused on the equations of state calculated for a selection of elemental solids. The relative agreement of different codes, including `lmf`, can be examined on the project site https://molmod.ugent.be/deltacodesdft. Agreement is quantified using an average "Delta value"—the integrated difference between equation of state curves over a specified volume range.

The calculated equations of states (scalar relativistic and PBE-GGA exchange–correlation are specified) agree extremely closely, particularly among the all-electron codes. The pseudopotential projects have also demonstrated precision very similar to the all-electron methods.

The Delta Codes test set has been carried out for `lmf` with an automatic setup, avoiding any system-specific modifications. The main settings defining the calculations are:

Sphere radii. s_R corresponding to touching spheres are used, evaluated at the smallest volume tested.

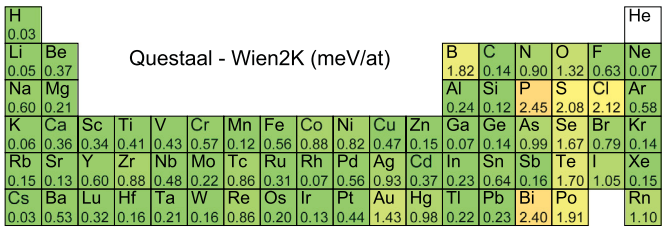

Fig. 17. "Delta Codes" tests: integrated energy difference between the energy-volume curves calculated by `lmf` and `WIEN2K` [81], expressed in meV/atom; the average value is 0.62 meV/atom.

ℓ_{max}. The maximum ℓ for the basis functions is 1 for H and He, 2 for $Z < 18$ and 3 otherwise. In each case, one higher ℓ is included in the partial wave expansion.

Semi-core states. Local orbitals are effective for including semicore states in the valence. We recommend treating states as semicore whenever the leaked charge exceeds $0.002e$, or the eigenvalue is higher than ~ -2 Ry. For transition metals and f-electron systems, the inclusion of high-lying $\ell = 2$ or 3 conventional local orbitals can significantly improve accuracy and these are added by default for these groups of elements.

Molecular cases. For cases where the touching muffin-tin spheres fill less than 30% of the cell volume, an additional LAPW basis has been included. The LMTO basis was not designed for molecular systems — but the shortcomings of the LMTO basis can be easily rectified by the addition of small number of plane-waves.

Basis setup. Basis parameters determined automatically as described below.

3.13.1. Choice of basis parameters

Questaal's full-potential code allows two LMTO basis functions to be used per ℓ. Each of these is defined by two parameters which must be chosen: the Hankel energy and the smoothing radius. Because the size of the LMTO basis is small, the choice of basis parameters can have a significant impact on accuracy. Different schemes for choosing basis parameters have evolved, of which the simplest is to fit the basis functions to the free atom wave functions. This provides one set of basis parameters for each ℓ. The second set is obtained from the first by shifting the Hankel energy (typically by -0.5 to -0.8 Ry). This is a heuristic approach but it is automatic and effective and is the default.

An alternative method chooses smoothing radii directly from the potential. The gradient of the smoothed Hankel functions at typical muffin-tin sizes is sensitive to r_s and variational freedom is obtained by differentiating the two sets of basis functions by r_s instead of the Hankel energy. The smoothing radius describes the point where the basis function begins to deviate from the exponential, Hankel-like tail: this is similar to the behaviour of atomic eigenfunctions at their classical turning points.

Associating the smoothing radius with the classical turning point allows basis functions to be constructed that resemble atomic-like states at energies different to the atomic eigenvalues and the question of choosing suitable r_s is translated into a determining a pair of reference energies—one each for the two basis sets per L. Energies of the valence states are typically in the range 0 to -2 Ry, and choices for the reference energies of $E_1 = -0.5$ and $E_2 = -2.0$ result in accurate basis sets for most materials. In this scheme, the Hankel energy remains a free parameter but similar results are found for Hankel energies in the range -0.1 to -1 Ry.

The choice of reference energies can be automated by expressing them in terms of the atomic potential at the muffin-tin radius. One basis set is chosen with a higher reference energy, giving a more diffuse basis function, the other is setup at a smaller, more negative, energy which yields a smaller r_s and a more tightly bound basis function. Because $v(s_R)$ varies significantly across elements and materials, a proportional scheme is more suitable: e.g., $v(r_s^1) = 2v(s_R)$ and $v(r_s^2) = \frac{1}{2}v(s_R)$; this algorithm is used for automating the DeltaCodes tests.

3.13.2. Comparison with LAPW results

Fig. 17 shows the "Delta values" for Questaal with respect to the LAPW code WIEN2K. The level of agreement is typical (or better) than most all electron methods involved in the validation exercise. In particular, the transition elements are extremely well represented by the smooth Hankel basis. Some of the first period and group VII cases involve molecular problems for which the muffin-tins fill a small fraction of the unit cell: the performance of the LMTO basis in these cases becomes sensitive to r_s and E choices. We do not attempt to choose optimal parameters, instead we include an additional LAPW (see Section 3.10), with cut-off 2 Ry or 4 Ry depending on the packing fraction. Au and group VI are outliers: in both cases the automatic basis setup generates basis functions that are rather too smooth to describe the tightly bound states in these elements. If necessary, improved performance can be obtained, thereby converging V_0 with respect to the basis, by adding plane waves.

In addition to confirming that the smooth Hankel basis is capable of high precision, the testing procedure also demonstrates that the automatic procedure for setting up the basis and setting various numerical parameters is highly reliable.

4. *GW* And QS*GW*

The *GW* approximation (G = Green's function, W = screened Coulomb interaction) is the first-order term in the formally exact diagrammatic expansion around some one-particle Hamiltonian H_0 [82].

Questaal's *GW* formalism is explained in some detail in Ref. [83], and its implementation of quasi-particle self-consistent *GW* (QS*GW*) is explained in Ref. [84]. Here we recapitulate only the main points. From an implementation point of view, *GW* requires two-point functions, such as the susceptibility from which W is made (Section 6). This entails an auxiliary *product basis* of wave function products. Most commonly plane waves are used to implement *GW*. In such a case the same basis for two-particle objects can be used, because a product of plane waves is another plane wave, but for all-electron methods this is not possible. Representation of these objects requires a basis spanning the Hilbert space of wave function products. Four-centre integrals (e.g. Eq. (183)) can be evaluated as integrals of pairs of product basis functions. This was first accomplished by Aryasetiawan [85] in an ASA framework, and extended by us [15] into a full-potential scheme.

In the *ab initio* context, *GW* has traditionally referred to a perturbation around the LDA: i.e. $H_0 = H^{\rm LDA}$ so that $GW = G^{\rm LDA}W^{\rm LDA}$, though in recent years hybrid functionals or LDA+U have become popular. It has become increasingly recognised that results depend rather strongly on the choice of H_0, or equivalently the $V_{\rm xc}$ entering into it, by which we mean the beyond-Hartree parts of the effective potential. Thus the self-energy of H_0 is $\Sigma = V_{\rm xc}$ in the language of many-body perturbation theory. The implications of starting point dependence are profound: the quality of *GW* depends on the quality of the reference. Moreover it is increasingly accepted that $H^{\rm LDA}$ is only a good choice for H_0 for a very restricted materials class.

GW is a perturbation theory, and usually only the lowest order perturbation (diagonal part of Σ in the basis of the reference Hamiltonian) is kept. As a result eigenfunctions are not perturbed and the change in QP level ε_{kn} is

$$\delta\varepsilon_{kn} = Z_{kn}[\langle\Psi_{kn}|\Sigma(\mathbf{r},\mathbf{r}',\varepsilon_{kn}) - V_{\rm xc}^{\rm LDA}(\mathbf{r})|\Psi_{kn}\rangle] \tag{172}$$

Z_{kn} is the quasi-particle (QP) renormalisation factor

$$Z_{kn} = \left[1 - \langle\Psi_{kn}|\frac{\partial}{\partial\varepsilon}\Sigma(\mathbf{r},\mathbf{r}',\varepsilon_{kn})|\Psi_{kn}\rangle\right]^{-1} \tag{173}$$

and accounts for the fact that Σ is evaluated at ε_{kn} rather than at the QP energy $\varepsilon_{kn} + \delta\varepsilon_{kn}$. Eq. (172) is the customary way QP energies are evaluated in *GW* calculations. In Sec. VI of Ref. [56], we show that omitting the Z factor is an approximate way to incorporate self-consistency, and thus should be a better choice than including it. As a practical matter it is also the case that results improve using $Z = 1$.

Ref. [56] notes a serious drawback in the diagonal-only approximation: when levels in the reference Hamiltonian are wrongly ordered as they are, e.g. at the Γ point in Ge, InN, and $CuInSe_2$, and the L point in PbTe [78] the wrong starting topology results in an unphysical band crossing in the *GW* QP levels. (See, for example an illustration for Ge in Fig. 6 of Ref. [56]). Something similar will happen for PbTe (Fig. 13). Moreover, the off-diagonal parts of Σ can significantly modify the eigenfunctions and resulting charge density. This reflects itself in many contexts, e.g. strong renormalisation of the bandgaps in polar insulators such as $TiSe_2$ and CeO_2 [56]. It modifies orbital character of states near the Fermi energy in La_2CuO_4, and significantly affects how the metal–insulator transition comes about, when *GW* is combined with DMFT [86].

4.1. Need for self-consistency

The arbitrariness in the starting point means that there is no unique definition of the *GW* approximation. Errors in the theory (estimated by deviations from experiment) can be fairly scattered for a particular choice of reference, e.g. LDA, making it unclear what the shortcomings of *GW* actually are. Arbitrariness in the starting point can be surmounted by iterating G to self-consistency, that is, by finding a G generated by *GW* that is the same as the G that generates it ($G^{\rm out} = G^{\rm in}$). But it has long been known that full self-consistency can be quite poor in solids [87,88]. A recent re-examination of some semiconductors [89] confirms that the dielectric function (and concomitant QP levels) indeed worsen when G is self-consistent, for reasons explained in Appendix A in Ref. [84]. Fully sc*GW* becomes more problematic in transition metals [90]. Finally, even while sc*GW* is a conserving approximation in the Green's function G, in W it is not: it violates the sum rule [91] and loses its usual physical meaning as a response function. As a result it washes out spectral functions in transition metals [90], often yielding worse results than the LDA.

A simple kind of self-consistency is to update eigenvalues but retain LDA eigenfunctions, as was first done by Aryasetiawan and Gunnarsson [92]. This greatly improves *GW*, especially in systems such as NiO where the LDA starting point is very poor. But it is only adequate in limited circumstances. In $TiSe_2$ it strongly overestimates the bandgap (see below), as it does in NiO [92]. An extreme case is CeO_2 where the position of the Ce 4*f* levels is severely overestimated; see Fig. 3 in Ref. [83]. The off-diagonal elements in Σ, which are needed to modify the eigenfunctions, can be very important.

Questaal employs *quasiparticle* self-consistency [16,83,84]: by construction, the reference H_0 is determined within *GW* as in the fully self-consistent case. But the proper $\Sigma = iGW$ cannot be used because it is energy dependent and non-Hermitian, thus falling outside of an independent particle picture. This causes *GW* to degrade, and higher-order diagrams are needed [87] to restore the quality of *GW*. But by quasiparticlising Σ we can stay within a framework of perturbation around H_0 but choose H_0 in an optimal manner. QS*GW* quasiparticlises $\Sigma(\varepsilon)$ by approximating it by a static Hermitian potential as

$$\Sigma_0 = \frac{1}{2}\sum_{ij}|\psi_i\rangle\left\{\mathrm{Re}[\Sigma(\varepsilon_i)]_{ij} + \mathrm{Re}[\Sigma(\varepsilon_j)]_{ij}\right\}\langle\psi_j| \tag{174}$$

which is the QS*GW* form for a static $V_{\rm xc}$. This process is carried out to self-consistency, until $\Sigma_0^{\rm out} = \Sigma_0^{\rm in}$. Given Σ_0, a new effective Hamiltonian can be made, which generates a new Σ_0:

$$\Sigma_0 \to G_0 \to \Sigma = iG_0W[G_0] \to \Sigma_0 \ldots \tag{175}$$

Σ_0 is nonlocal with off-diagonal components; both of these features are important.

$TiSe_2$ is an interesting case where self-consistency becomes critical. It is a layered diselenide with space group $P\bar{3}m1$. Below $T_c = 200$ K, it undergoes a phase transition to a charge density wave, forming a commensurate $2\times2\times2$ superlattice ($P\bar{3}c1$) of the original structure. It has attracted a great deal of interest in part because the CDW may to be connected to superconductivity.

How the CDW affects the energy band structure is a source of great controversy. It is accepted experimentally that in the $P\bar{3}c1$ phase, $TiSe_2$ is semiconductor with a gap of ~0.15 eV. Above T_c whether intrinsic $TiSe_2$ has a gap is not settled. On the theoretical side, the CDW makes $TiSe_2$ an unusual materials system: DFT predicts a metal in both $P\bar{3}m1$ and $P\bar{3}c1$, as expected, but a recent $G^{\rm LDA}W^{\rm LDA}$ calculation [93] predicts $P\bar{3}m1$ to have a small positive gap. We have revisited this problem and confirmed the findings of Ref. [93] at the $G^{\rm LDA}W^{\rm LDA}$ level. However, at the QS*GW* level, $TiSe_2$ is a *metal* in the ideal $P\bar{3}m1$ phase. This is atypical for insulators: usually self-consistency *widens* the gap, as has long been known (see Fig. 1 in Ref. [83]). The origin can be traced to the modification of the LDA eigenfunctions by off-diagonal elements $\Sigma_{n\neq n'}$, which modify the density $n(r)$. To see approximately the effect of the density change, assume LDA adequately describes $\delta V^{\rm eff}/\delta n$ (this is the inverse susceptibility, the charge analogue of Eq. (180)). For a fixed Σ^0, we can estimate the effect of the change δn renormalising the $V^{\rm eff}$ through the change

$$\delta\Sigma \approx \left\{\Sigma_0 - V_{\rm xc}^{\rm LDA}[n^{\rm LDA}]\right\} + V_{\rm xc}^{\rm LDA}[n^{GW}]$$

This is accomplished in a natural way with the Questaal package. The quantity in curly brackets is generated by *GW* in the first QS*GW* cycle, and `lmf` treats this term as an external perturbation. Running `lmf` to self-consistency allows the system to respond to the potential and screen it, yielding n^{GW}.

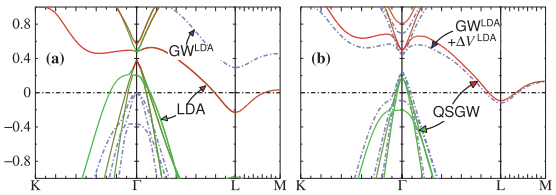

Fig. 18. $TiSe_2$ energy bands in eV for the undistorted $P\bar{3}c1$ structure. (*a*): solid lines are LDA results, with red and green depicting a projection onto Ti and Se orbital character, respectively. Blue dashed line shows shifts calculated in the *GW* approximation based on the LDA. (*b*): blue dashed line shows results from *GW* based on the LDA (same self-energy as in panel (*a*)), with an extra potential $\Delta V^{\rm LDA}$ deriving from a charge density shift computed from the rotation of the LDA eigenvectors. Solid lines are QS*GW* results, with the same colour scheme as in panel (*a*). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

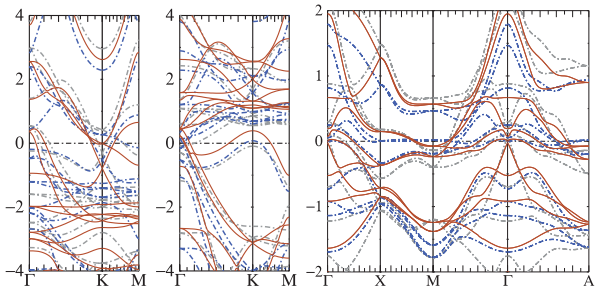

Fig. 19. Energy band structure, in eV, of majority-spin MnAs (left), minority-spin MnAs (middle) and nonmagnetic FeTe (right) in (a) the QS*GW* approximation (orange solid lines), $G^{\mathrm{LDA}}W^{\mathrm{LDA}}$ (dashed blue lines), and the LDA (light dashed grey lines). $G^{\mathrm{LDA}}W^{\mathrm{LDA}}$ was calculated from the LDA, but including the off-diagonal components of Σ with $Z = 1$. It is the 0^{th} iteration of the QS*GW* self-consistency cycle.

The result of this process is shown as a dashed line in Fig. 18(b). It bears a close resemblance to the full QS*GW* result, including the negative gap. We will show elsewhere that $TiSe_2$ is an insulator in the $P\bar{3}c1$ phase only as a consequence of lattice displacements relative to the symmetric $P\bar{3}m1$ phase.

GW is used less often in metals. As in the insulating case, *GW* based on DFT can sometimes work well, as it apparently does for $SrVO_3$ [94]. But the range of applicability is limited, and indeed $G^{\mathrm{LDA}}W^{\mathrm{LDA}}$ can yield catastrophically bad results. Two cases in point are MnAs and FeTe. Fig. 19 shows LDA, $G^{\mathrm{LDA}}W^{\mathrm{LDA}}$, and QS*GW* energy bands near E_F; they are mostly of Mn or Fe 3*d* character. In the $G^{\mathrm{LDA}}W^{\mathrm{LDA}}$ case, E_F must be adjusted to conserve charge. (The shift is large, of order 1 eV.) Consider MnAs first. The majority and minority 3*d* bands lie at about -1.5 eV and $+1$ eV in the LDA. This exchange splitting is underestimated: QS*GW* increases the splitting, putting bands at about $(-2.5, 1.5)$ eV while $G^{\mathrm{LDA}}W^{\mathrm{LDA}}$ does the opposite, reducing the splitting relative to LDA. Also, the Fermi surface topology is poor: the hole pocket at Γ disappears, and the one at K is similar to the LDA. FeTe fares no better. $G^{\mathrm{LDA}}W^{\mathrm{LDA}}$ is somewhere intermediate between LDA and QS*GW*. Most important is the unphysical, dispersionless band at E_F on the X-M line, which yields a nonsensical Fermi surface. In both of these materials, the LDA describes the electronic structure better than $G^{\mathrm{LDA}}W^{\mathrm{LDA}}$.

Ambiguities in the starting point make it difficult to know what errors are intrinsic to the theory and which are accidental. The literature is rife with manifestations of this problem; see e.g. Ref. [95]. Self-consistency (mostly) removes the starting-point dependence so that the errors intrinsic to the *GW* approximation can be best elucidated. Moreover, QS*GW* should yield better RPA total energies on average than those calculated from other starting points, because the path of adiabatic connection is optimally described through QS*GW* [84]. To better illustrate these points Fig. 20 shows the ionisation potential of 3*d* transition metal atoms and the heat of formation of 3*d*-O dimers, computed by the `molgw` code [96]. We focus on these properties because it is known, e.g. from Ref. [97], that the RPA tends to systematically overbind, and the error is connected with short-ranged correlations. The ionisation potential and the dimer formation energy, both of which benefit from partial cancellation of such errors, are much better described.

Note the dramatic difference between choice of a Hartree–Fock or the PBE functional. As for the superiority of QS*GW*, Fig. 20 generally bears this argument out, though there are some surprises, e.g. the heat of formation from RPA@PBE deviates less from CCSD(T) results than QS*GW*. QS*GW* does not describe Cr well, probably because spin fluctuations are important there (Section 4.4), which skews the statistics with this small sample.

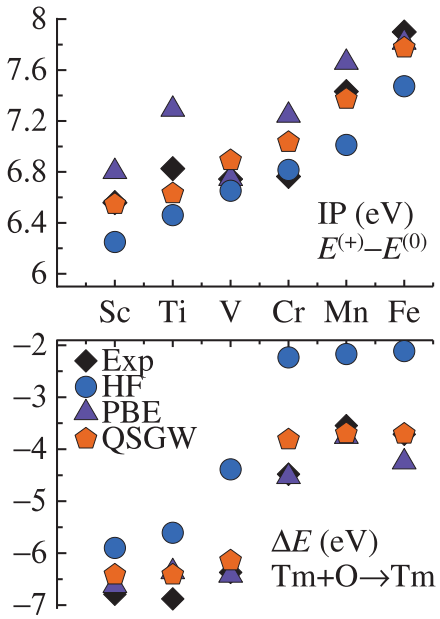

Fig. 20. Ionisation potential, IP, of TM atom and heat of formation ΔE of transition-metal dimers computed within the RPA, using the `molgw` code selecting various choices of starting point. Note the dramatic difference between Hartree–Fock and the PBE starting points. Of the common semi-empirical functionals, HSE06 (not shown) is the closest to QS*GW*. Data were compared to reference CCSD(T) for the ionisation potential and experiment for ΔE.

The discrepancies are small enough that incomplete convergence in the basis set (results in Fig. 20 were obtained with a triple-ζ basis) are of the same order and could account for some of the results. The ionisation potential of simple *sp* atoms was analysed in Ref. [98], where it was computed from RPA total energy differences, compared to the eigenvalue of the *GW* one-particle Hamiltonian. There also it was shown QS*GW* significantly improves on RPA@HF, and also that QS*GW* is slightly biased towards Hartree–Fock, reflecting a slight tendency to underestimate screening. This is consistent with the tendency for QS*GW* to overestimate bandgaps. Koval et al. [99] found somewhat different results for small, second-row molecules.

QS*GW*-RPA is not accurate enough to reach quantum chemical accuracy, $\sim$1 kcal/mol. The addition of ladders should considerably improve on the RPA's known inadequacy in describing short-ranged correlations. It has long been known that ladders dramatically improve the dielectric function $\epsilon(\omega)$ ($\epsilon(\omega)$ and G enter in the coupling constant integration for total energy) and it would be interesting to explore if such low-order diagrams in a QS*GW* framework are sufficient to capture total energy to this accuracy.

While self-consistency does largely surmount starting-point dependence (it was recently shown to be the case for some insulators using Hartree–Fock and LDA as starting points [100]), but it is not strictly true that it does. The QS*GW* H_0 or G_0 has a Hartree–Fock structure, and it can get stuck in metastable valleys in the same way as Hartree–Fock; an example are the dual low-spin and high-spin states found for FeSe (Section 4.4). Also, in strongly correlated systems $\Sigma(\varepsilon)$ can vary rapidly with energy. There can be more than one ϵ_i that forms a stationary point. This happens, for example, in CuO. It is usually an indication that the ground state is not well described by a single Slater determinant. QS*GW* restricts G_0 to a single determinant, and there is no longer an unambiguously optimal choice when there are two or more equally strongly competing ones.

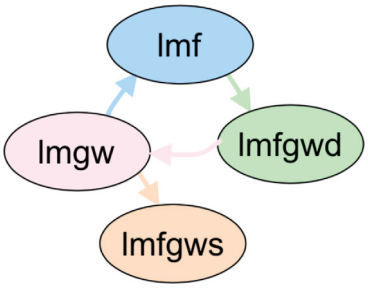

Fig. 21. Questaal's QS*GW* cycle. `lmf` generates a new self-consistent density and noninteracting H_0 for a fixed Σ_0; `lmfgwd` generates information about the eigenfunctions Ψ corresponding to H_0, and `lmgw` receives the output of `lmfgwd` and uses it to make a new Σ_0. `lmgw` can also generate response functions and the fully dynamical self-energy $\Sigma^{nn}(\omega)$. `lmfgws` is a post-processing code that yields observables computed from the interacting G, where $G^{-1} = {G_0}^{-1} + \Sigma^{nn}(\omega) - {\Sigma_0}^{nn}$.

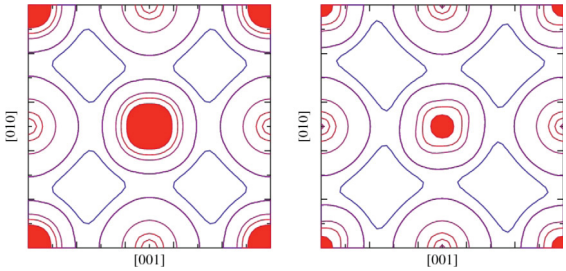

Fig. 22. Self-consistent charge density in antiferromagnetic NiO (left) and CoO (right), in the plane normal to [001], computed by QS*GW*. The Ni (Co) nucleus lies at the centre of the Figure.

There is no unique prescription for quasiparticlisation, but Eq. (174) minimises the difference between the full G and G_0 according a particular choice of norm. Recently it was shown [101] that Eq. (174) minimises the absolute value of the gradient of the Klein functional over all possible spaces of non-interacting Green's functions. There are formal reasons (Z factor cancellation [84] and conserving sum rule [91]) why quasiparticlisation should be better; there is also abundant empirical evidence, that quasiparticle self-consistency performs better than full self-consistency [89,90,102], and the particular construction Eq. (174) is an optimum one.

QS*GW* has been implemented in numerous electronic structure packages, e.g. the `spex` code [103], the `Abinit` code [104], `VASP` [105], recently in the Exciting code [100], and also in molecules codes such as `molgw` [96] and `NWChem` [99]. For solids perhaps the most rigorous implementation is the `spex` code, which can use a local Sternheimer method to rapidly converge the calculation of the dielectric function [106]. As attempts to make efficient schemes that are also well converged, e.g. to calculate the RPA total energy, the local Sternheimer method will likely emerge as being very significant over time.

4.2. Questaal's QSGW implementation

Questaal's implementation of QS*GW* evolved from the `ecalj` package, developed by Kotani and coworkers in the early 2000's out of Aryasetiawan's *GW*-ASA code [13], which in turn was developed from the "Stuttgart" LMTO code. The original `ecalj` package can now be found at https://github.com/tkotani/ecalj/. As of this writing we maintain the *GW* code as a separate branch (Fig. 21). In its present form the *GW* part of Questaal's QS*GW* implementation operates independently, receiving information about eigenfunctions and eigenvalues, and returning response functions or a self-energy. As for the QS*GW* cycle, one cycle occurs in three parts.

Rather than storing Σ_0 on disk, $\Sigma_0 - V_{\rm xc}$ is stored ($V_{\rm xc}$ usually being the LDA potential). Thus `lmf` generates its customary LDA Hamiltonian, and adds $\Sigma_0 - V_{\rm xc}$ to it. In practice this is accomplished by reading $\Sigma_0 - V_{\rm xc}$ on the mesh of k points where it was generated, and inverse Bloch-summing to make $\Sigma_0 - V_{\rm xc}$ in real space. Then $\Sigma_0 - V_{\rm xc}$ can be interpolated to any k by a forward Bloch sum. This is a unique and powerful feature, as it makes it possible to generate QS*GW* eigenfunctions and eigenvalues at any k. There are some subtleties in this step: the interpolation does not work well if all the $\langle n|\Sigma_0^{nn'}|n'\rangle$ are included. It is solved by rotating to the LDA eigenfunctions, and zeroing out $\Sigma_0^{n\neq n'}$ for n or n' whose energy exceeds a threshold, $E_{\rm cut}$; i.e. diagonal-only approximation for high-energy subblocks of Σ_0. $E_{\rm cut}$ is typically $\sim$2 Ry; convergence can be checked by varying $E_{\rm cut}$. The prescription is explained in detail in Sec. IIG of Ref. [84]. The error is connected to the relatively long range of the smooth Hankel envelopes. Preliminary tests using short-ranged, screened Hankels indicate that this truncation is no longer needed.

Typically Σ_0 is a smoother function of k than the kinetic energy. Thus the k mesh on which Σ_0 is made can usually be coarser than the k mesh. Because of its ability to interpolate, `lmf` and `lmgw` operate with independent meshes.

4.3. Successes of QSGW

QS*GW* has the ability to calculate properties for a wide variety of materials classes in a manner that no other single theory can equal. It consistently shows dramatic improvement relative to DFT and extensions such as hybrid functionals or LDA-based *GW*. Atomic ionisation potentials [98], quasiparticle (QP) levels, Dresselhaus coefficients in semiconductors [107–109]; band offsets at the Si/SiO_2 interface [104], magnetic moments and spin wave excitations [110,111]; tunnelling magnetoresistance [112]; impact ionisation [113]; electric field gradients and deformation potentials [114], spectral functions [115], and dielectric response [116]. Its superior ability to yield QP levels and DOS made it possible to accurately determine valence maximum in $SrTiO_3$ [117]. In contrast to $G^{\rm LDA}W^{\rm LDA}$, QS*GW* is uniformly reliable with systematic errors (see Section 4.4). Fundamental gaps are usually in very good agreement with experiments, though they are systematically overestimated (see Fig. 1 in Ref. [83]), even in the strongly correlated M1 phase of VO_2 [118–120], where spin fluctuations are not important. It properly narrows the bandwidth of localised d bands and widens it in wide-gap semiconductors [117] and graphene [121].

QS*GW* can predict properties inaccessible to $G^{\rm LDA}W^{\rm LDA}$, such as magnetic ground state [115] and charge density. Fig. 22 shows the density calculated by QS*GW* in the plane normal to [001] for antiferromagnetic NiO and CoO. Two prominent features are seen: (1) NiO is much more spherical than CoO, and the density contours for CoO are elongated along the [110] line. Both of these findings are consistent with γ ray measurements [122,123].

It was noticed early on [83] that QS*GW* has a systematic tendency to overestimate bandgaps in semiconductors and insulators. To what extent dispersions are well described is much less discussed, in part because there is a limited amount of experimental data reliable to the precision needed. Perhaps the best materials family to serve as a benchmark are the zincblende semiconductors, where critical-point analysis of the dielectric function has been used to accurately measure in most zincblende semiconductors, not only the Γ-Γ (E_0) transition but also the L-L (E_1) and X-X (E_2) transitions, as indicated in Fig. 23, which

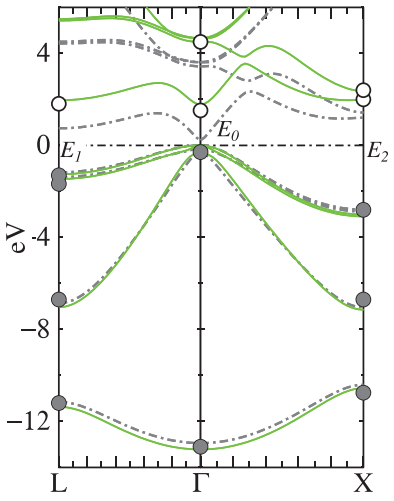

Fig. 23. QSGW energy band structure of GaAs compared with ellipsometry data and photoemission data (open and closed circles). Dashed Grey lines are LDA bands.

presents the QSGW and LDA bands compared with ellipsometry and photoemission data.

Fig. 24 shows that tendency to overestimate E_0 and E_1 is systematic and uniform, but there is less uniformity in E_2. Usually E_0 can be precisely measured; similarly for E_1 because there is a sizeable volume of k near L where valence and conduction bands disperse in a parallel manner. E_2 is more difficult, and values are less certain. That being said, the data suggest E_0-E_2 is, on average, about 0.1 eV below experiment. We have found that this error is, at least in part, due to the necessity of truncating high-lying parts of the off-diagonal Σ_0; see E_{cut}, Section 4.1. It has recently become possible to re-evaluate this because a short-range basis was recently developed (Section 3.12) appears to interpolate Σ_0 allowing $E_{\text{cut}} \to \infty$.

Also reliably measured are the electron effective mass m_e^* (hole masses are much less well known), and for some systems, the nonparabolicity parameter α, characterising deviations from parabolicity near Γ, and defined as $k^2/m^* = \varepsilon(k)[1 - \alpha\varepsilon(k)]$ where ε is the band energy relative to the conduction band minimum. α is an important quantity in several contexts, particularly hot-electron semiconductor electronics. Unfortunately α is difficult to measure, and values reported are usually some scalar average of the second rank tensor, where it has been measured. QSGW (or scaled QSGW, Eq. (176)) seems to predict α to within the experimental resolution; see e.g. Table 6 for GaAs. Fig. 25 compares QSGW and experimental effective masses for various semiconductors. Conduction band masses are well described, although there is a systematic tendency to underestimate it, connected to the tendency to overestimate E_G.

Table 6
GaAs: fundamental gap, effective masses, nonparabolicity α and location of 5/2 and 3/2 Ga 3d core states relative to the valence band maximum. Units are eV. The measurement of α was taken from Ref. [124], which is within 1% of a subsequent measurement [125]. Photoemission data taken from Ref. [126].

	E_G	m^*/m	α	Ga 3d
QSGW	1.78	0.076	−0.71	18.5, 18.0
Eq. (176)	1.47	0.067	−0.82	17.8, 17.3
LDA	0.24	0.020	−4.0	15.2, 14.8
Expt	1.52	0.067	−0.83	19.3, 18.6

Most of this error can be traced to the RPA approximation to the dielectric function. The optical dielectric constant ϵ_∞ (Eq. (191)) in the $\omega \to 0$ limit is uniformly too small by a factor ∼0.8 for many kinds of semiconductors and insulators (Fig. 26). Plasmon peaks in Im$\epsilon(\omega)$ are almost universally blue shifted [127,128] because the RPA omits the attraction between electron–hole pairs in their virtual excitations. A simple Kramers–Kronig analysis shows that as a consequence, $\epsilon_\infty = \text{Re}\,\epsilon(\omega \to 0)$ should be underestimated. Indeed we find that to be the case, but remarkably ϵ_∞ is consistently underestimated, by a nearly universal factor of 0.8. This is what motivated a hybrid of LDA and QSGW functionals [107]

$$\Sigma_0^{\text{scaled}} = 0.8\,\Sigma_0 + 0.2\,V_{\text{xc}}^{\text{LDA}} \tag{176}$$

The reasoning is since W is dominated by the $q \to 0, \omega \to 0$ limit, scaling W by 0.8 justifies Eq. (176). In practice, it seems that the tendency to overestimated bandgaps is almost completely ameliorated in these systems (compare red and blue circles, Fig. 24, as is the tendency to underestimate m_c^*. Exceptions are the small-gap InAs in InSb, where relative errors in bandgaps are still not small even with the 0.8 scaling.

The hybrid scheme does a stellar job in many contexts: it improves on bandgaps in semiconductors (Fig. 24) and predicting the Dresselhaus splitting in sp semiconductors [107]. But the scheme is empirical, and it is limited. It does not properly correct the blue shifts in $\epsilon(\omega)$ or take into account other excitonic effects and fails to bring bandgaps in close agreement with experiment in systems with strong spin fluctuations such as CoO or La_2CuO_4, and deep states such as the Ga 3d shift farther from experiment (see Table 6). Recent work [102,129]

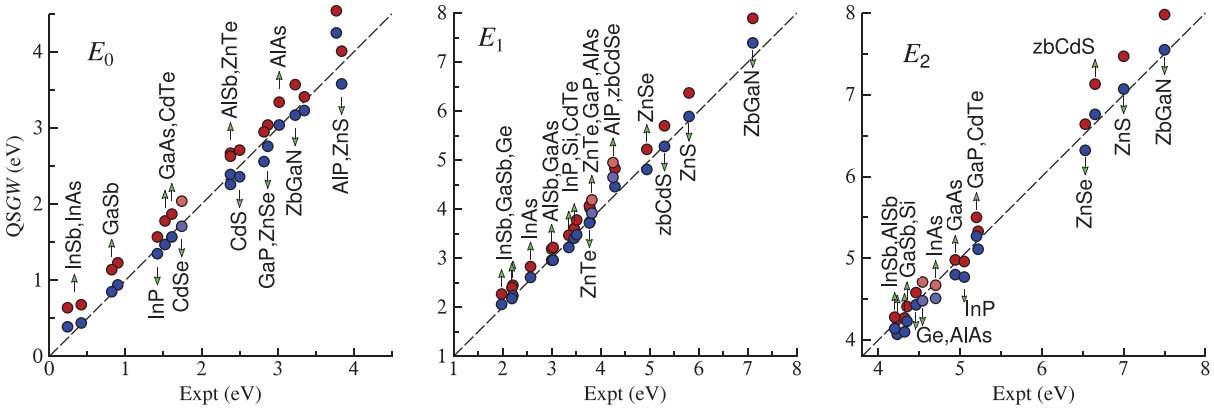

Fig. 24. Transitions E_0, E_1 and E_2 transitions in a zincblende semiconductor compared with critical point measurements. Red and blue circles are QSGW results, and QSGW with Σ_0^{scaled} (Eq. (176)), respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

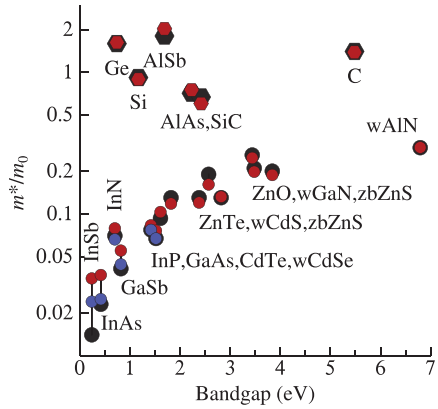

Fig. 25. Conduction band effective masses of selected zincblende compounds compared against experiment. Red and blue circles are QS*GW* results, and QS*GW* with Σ_0^{scaled}; black circles are experimental values.

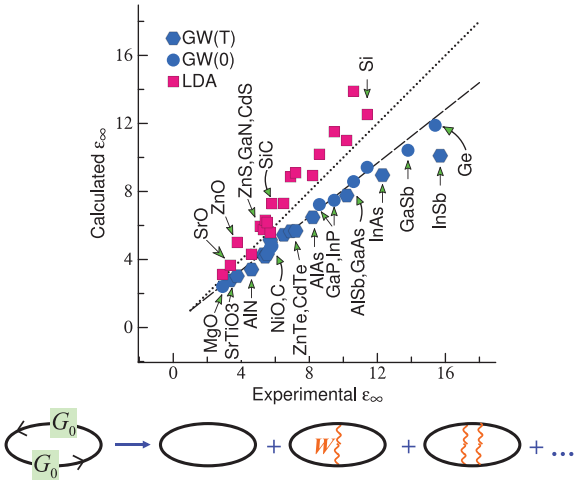

Fig. 26. Dielectric constant ϵ_∞ compared to experiment. Blue are QS*GW* results; red are LDA results. Also shown are the ladder diagrams that modify bubbles in the polarisability, left out in the RPA.

shows that the great majority of the error in charge susceptibility can be accounted for by ladder diagrams. (In the case of polar semiconductors there can also be a significant renormalisation of the gap from the Frölich interaction [130,131]; the electron–phonon interaction has a modest effect in other semiconductors, especially compounds with second row elements [132], diamond being the largest.) Without ladders the low-frequency dielectric constant is – almost universally – about 80% of experiment [133]. As we will show elsewhere [129] the addition of ladders in W seem to dramatically improve on the charge channel even in strongly correlated cases such as CoO. VO_2 is an excellent test bed for optics: it has strong correlations in the monoclinic phase; yet, the conductivity is well described by nonmagnetic QS*GW*, provided ladders are taken into account [119].

Below we present more detailed properties of the band structure obtained from classical QS*GW* on a single system, selecting GaAs because reliable measurements are available and it is representative of results obtained for the entire family of zincblende semiconductors (Table 6). We observe the following:

1. dispersions Γ-L and Γ-X are mostly well described, and vastly better than the LDA (which predicts the Γ-X dispersion to be 1 eV);
2. the valence bands match photoemission data to within experimental resolution;

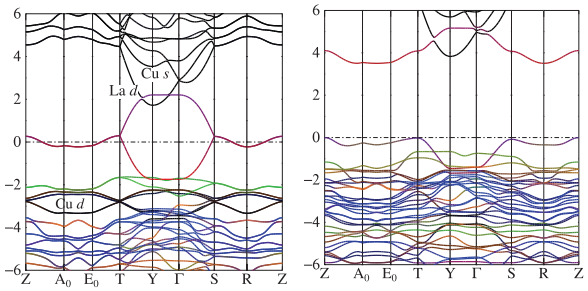

Fig. 27. QS*GW* band structure (in eV) of nonmagnetic La_2CuO_4 (left) and antiferromagnetic $LaCu_2O_4$ (right). Colours reflect the orbital character of the bands as follows: Cu $d_{x^2-y^2}$ (red); Cu d_{3z^2-1} (green), near −2 eV in QS*GW*; O-p_{xy} (blue), between −3 eV and −5 eV in QS*GW*; O-p_z (orange), between −3 eV and −5 eV in QS*GW*. The band near 3.3 eV at Y has significant Cu s character, and it also admixes into the Cu $d_{x^2-y^2}$ near E_F. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3. the Ga 3d (see Table 6) is pushed down relative to the LDA and also relative to $G^{\text{LDA}}W^{\text{LDA}}$. Binding of shallow core-like states on e.g. Cd, Zn, and Cu is underestimates by ∼0.4 eV; for deeper states such as Ga 3d it is larger. It was shown in Ref. [134] that a low-order vertex correction to *GW* accounts for most of this discrepancy, at least for shallow core-like levels;
4. conduction bands are slightly overestimated and the mass slightly too large; and
5. the nonparabolicity parameter α was calculated along the [001], [110], and [111] directions. It varies ∼30% for different directions, but since only scalar quantities are reported, we take a geometrical average. α is very sensitive to m^*, so we can expect it to be underestimated a little; this turns out to be the case. Using Eq. (176) all three quantities align well with experiment.

Other discrepancies become apparent in the rare earths: splitting between occupied and unoccupied 4f levels is too large [135], and multiplet splittings, which are significant for 4f, lie beyond the scope QS*GW*.

As we will show elsewhere [129] most of the systematic errors noted above are largely corrected when ladders are included in W. Kutepov [102] showed that the bandgaps significantly improve, but the improvement extends to a wide range of properties and very diverse kinds of materials systems. See for example, the excellent description of optical conductivity in VO_2 [119].

Metals and local-moment magnetic systems are similarly well described; see for example the excellent description of known properties of Fe in the Fermi liquid regime in Ref. [115], and the electronic structure of NiO and MnO in Ref. [16].

In summary, absent strong spin fluctuations, and a few other mild exceptions, QS*GW* with some low-order extensions (ladders and a low-order vertex for narrow semicore states), describe a wide range of materials throughout the periodic table with uniform accuracy and reliability that cannot be matched by any other method.

4.4. Limitations of QSGW

The greatest failings appear in QS*GW* for systems where spin fluctuations are large. This is not surprising since the main many-body effect in *GW* are plasmons. *GW* has no diagrams in spin beyond the Fock exchange. When spin fluctuations become important, the first effect is to reduce the average moment [136, 137]. As they increase, states begin to lose their coherence: the

Table 7
d_{xz}/d_{yz} and d_{xy} QP levels (meV) near E_F in tetragonal phase of FeSe (reference is the Fermi energy). LDA and QS*GW* data are calculated nonmagnetically; the line marked +DMFT following each is the result with DMFT added [139], as discussed in Section 5. Line SQS is a QS*GW* calculation in a ferrimagnetic SQS structure, as described in the text.

ARPES	Γ		M		Z	
	9	−18	−22	−42	7	34
LDA	109	113	−204	−337	254	141
+DMFT	30	45	−110	−125	42	65
QSGW	41	44	−107	−202	131	56
+DMFT	1	10	−21	−40	10	32
SQS6	60	45	−52	−70	31	68

band picture and QS*GW* both begin to break down. A good example of this is La_2CuO_4 (Fig. 27). Undoped La_2CuO_4 undergoes a paramagnetic to an antiferromagnetic insulator transition upon cooling below 325 K [138]. Fig. 27 shows QS*GW* calculations in the nonmagnetic state, and the antiferromagnetic one. Approximating the paramagnetic state by nonmagnetic state is an inadequate approximation, though it is often done. In this approximation La_2CuO_4 is predicted to be metallic, with a single Cu $d_{x^2-y^2}$ band crossing E_F. In the antiferromagnetic state La_2CuO_4 is found to be an insulator with a 3.5 eV bandgap. Experimentally La_2CuO_4 is an insulator with a gap ∼2 eV.

QS*GW* severely overestimates the optical gap in La_2CuO_4. There is a similar mismatch for CoO: the gap is overestimated by ∼2 eV in these cases, but not in NiO. Particularly telling is how for VO_2 nonmagnetic QS*GW* *does* describe the M1 phase very well when both V–V pairs dimerise [119], but not the M2 phase where only one of them does. (A gap opens up when M2 is calculated antiferromagnetically, but M2 is probably paramagnetic at room temperature.) M1 and M2 differ mainly through the V–V dimerisation, which suppresses spin fluctuations. As we will show elsewhere [129], including ladders considerably improves all of these antiferromagnetic insulators, but discrepancies remain particularly for La_2CuO_4. We believe this to be an artefact of the interaction between spin and charge fluctuations.

FeSe is a heavily studied superconductor with large spin fluctuations: it provides an excellent testbed to compare LDA and QS*GW*. Nonmagnetic calculations for LDA and QS*GW* are shown in Table 7. QS*GW* improves on the LDA, but discrepancies with ARPES are much larger than for, e.g. elemental Fe [115]. That LDA and QS*GW* should be different is readily seen from the k dependence of the Z factor, shown in Fig. 28. Local potentials such as the LDA cannot incorporate such an effect.

Spin fluctuations are important, but QS*GW* does not adequately capture them. They can be incorporated in a mean-field manner by constructing a Special QuasiRandom Structure [140] with, in this case, three spin-up and three spin-down Fe atoms. Two magnetic solutions can be stabilised: a low-moment and high-moment solution. Quasiparticle spectra in the latter case are far removed from experiment, so we consider only the low-moment solution. The moments on each Fe site are different, with a range $|m| = 0.2 \pm 0.15\mu_B$. The addition of magnetic terms shift QP levels closer to ARPES measurements, but a significant discrepancy with experiment remains. In Section 5 these results are revisited where local, high-order diagrams in spin are included.

These two systems show concretely how the shortcomings of *GW* become apparent when spin fluctuations are important. The simplest diagram beyond *GW*, the T matrix, was first considered in an *ab initio* framework by Aryasetiawan and Karlsson [141]; more recently it was implemented in the spex code [142]. If spin fluctuations are not strong, as seems to be the case for Fe and Ni, such an approach seems to describe spin excitations fairly well, though comparisons with experiments are too sparse to draw any strong conclusions yet. For strong spin fluctuations, e.g. in La_2CuO_4 or other unconventional superconductors such as FeSe, it seems probable that many kinds of diagrams are needed, though the effective vertex entering into those diagrams is expected to be mostly local. Dynamical Mean Field Theory is ideally situated to address such cases; it is discussed in Section 5.

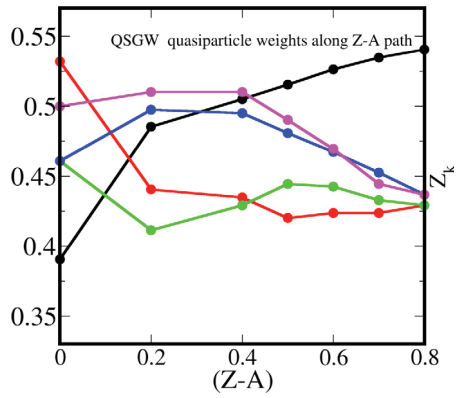

Fig. 28. The orbitally resolved quasiparticle renormalisation (Z) factors for FeSe Fe-3d orbitals along the high-symmetry Z-A line in tetragonal Brillouin zone; d_{xy} (black), d_{yz} (blue), d_{z^2} (green), d_{xz} (magenta), $d_{x^2-y^2}$ (red). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Our view is that QS*GW* with ladders in W are sufficient for reliable description of susceptibilities within the charge channel, especially when ladders are included in the self-consistency cycle [102,129] to make the QS*GW* H_0. In the spin channel DMFT contributes the dominant missing diagrams to the self-energy, and the susceptibilities can be reliably determined from local vertices, connected to k-dependent bubble diagrams.

5. DFT+DMFT and QS*GW*+DMFT

Hartree–Fock and DFT are band structure theories: electrons are treated as though they are independent; in solids electron states are Bloch waves. *GW* includes correlations that go beyond the Bloch picture; still it is a perturbation around the Bloch description. In Landau Fermi-liquid theory, electrons are replaced by 'quasi-particles' which are adiabatically connected to the single-particle Bloch-wave representation of electrons. This, by construction, is a theory for excited states that adds perturbative corrections to the band picture. In Landau adiabatic theory electrons can have an effective mass, which is different from its free mass, and additionally, lifetime broadening effects far from the Fermi energy. One extreme case of this scenario is when the low-energy quasi-particle vanishes (without magnetic ordering) and the atomic-like, high-energy excitation emerges and effective mass at the Fermi energy tends to ∞. In effect, this is an emergent, non-adiabatic, feature where single-particle spectral weight flows to higher energies to conserve the total spectral weight and subsequently leads to suppression of low-energy quasi-particles. Popular rigid-band techniques [143–150] would catch the physics of suppression of quasi-particles at Fermi energy, however, would fail to conserve the spectral weight since a dynamic self-energy ($\Sigma(\omega)$) is absent from those theories.

DMFT was formulated as an effective theory that smoothly interpolates between the Bloch-wave limit and atomic limit. It is formulated as a local impurity embedded self-consistently in a medium or bath, and was introduced in a seminal work by Metzner and Vollhardt [151] in late 80s. Several other works from Mueller-Hartmann [152], Brandt and Mielsh [153], Jarrell et al. [154] and Georges et al. [155] helped to build a strong

foundation for DMFT. The formulation and the implementation became popular through the seminal work by Georges et al. [156] and later emerged as a way to study strong correlations in real systems, supplementing an *ab initio* LDA framework by embedding a model Anderson impurity hamiltonian inside an LDA bath [157].

An implementation of DMFT has three components: the partitioning of the Hamiltonian into impurity + bath, a means to solve the impurity problem, and the bath-impurity self-consistency. The bath is essential because correlated states of interest have both atomic-like and Bloch-like properties with interesting low- and high-energy excitations. Through self-consistency the impurity feeds back on the bath, and ensures that all the symmetries of the entire system are preserved. In Questaal, QS*GW* or DFT can act as the bath. We have interfaces to two local solvers: the hybridisation expansion flavour of continuous time Quantum Monte Carlo (CT-QMC) from `TRIQS` [28] and the CT-QMC written by Kristjan Haule [27].

For the partitioning, some correlated subspace must be identified and this is done using projectors. Usually the subspace is derived from atomic-like *d*- or *f*-states of a transition metal or *f* shell element. For projectors maximally-localised Wannier orbitals have been widely used [158,159]. We have so far adopted the projection into partial waves of a particular ℓ in an augmentation sphere [160]. There are strong reasons to prefer this scheme: the basis set is atom-centred and very localised with a definite ℓ, which is the Hilbert space where the correlations are strong. The less localised the basis, the more it contains weakly correlated components (e.g. O-p states in transition metal oxides). Too much population of the correlated subspace with uncorrelated parts conflicts with the physical interpretation of separation of bath and interacting subspace and can reduce the reliability of the method [161]. Moreover, very localised basis functions have a much weaker energy-dependence. As a consequence, with a very local projector it is possible to construct an effective interaction over a wide energy window, reducing the frequency-dependence of the effective interaction, and essentially recovering the partial waves in augmentation spheres [17]. Partial waves closely resemble atomic orbitals, so solving a purely locally correlated Hamiltonian with the approximation of single-site DMFT (local self energy) is a good approximation.

The thorniest issue is how to construct the effective local Hamiltonian. It requires three related quantities: definition of the local Green's function, an effective Coulomb interaction U, which is partially screened by the bath, and the double counting (DC) correction.

In a fully *ab initio* QS*GW* + DMFT theory, these quantities must be computed from the theory itself, but how to accomplish this is far from a settled issue. Questaal can estimate the Hubbard U and J using the constrained random-phase approximation [162] where the internal transitions between states included in the DMFT subspace are removed. A fully internally consistent theory requires a frequency-dependent $U(\omega)$, but available solvers can only solve the local impurity problem with such a U in the density–density channel. Haule's prescription for excluding states in a large energy window partially mitigates this issue, since U is closer to bare and nearly energy-independent [17]. (It makes another approximation, namely that not all the states excluded from screening are included in the impurity problem; these states are effectively treated only at the bath level.)

5.1. Questaal's implementation of DMFT

Questaal does not have its own DMFT solver. It has an interface to Haule's CTQMC solver [27] and to the `TRIQS` library [28]. Haule's CTQMC solver uses singular value decomposition (SVD) and `TRIQS` uses Legendre polynomial basis as compact representations for single-particle propagators.

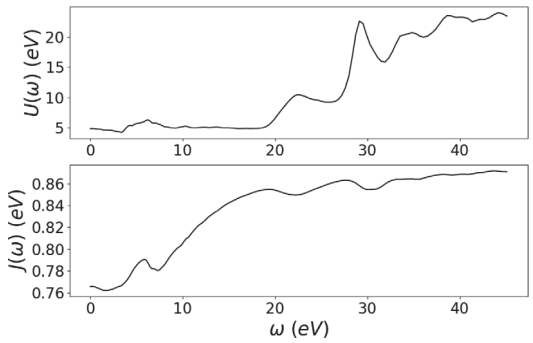

Fig. 29. $U(\omega)$ and $J(\omega)$ computed using our c-RPA scheme for Ni 3d orbitals starting from a QS*GW* bath.

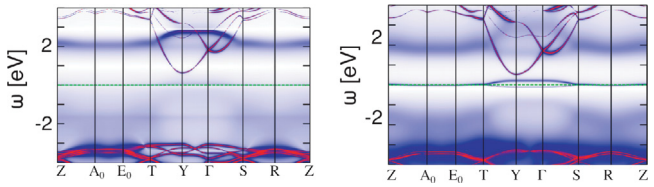

Fig. 30. QS*GW* + DMFT spectral functions for La_2CuO_4 (left) and metallic QS*GW* + DMFT spectral functions for doped ($x = 0.12$) $La_{2-x}Sr_xCuO_4$ (right) which shows the classic three-peaked structure with upper and lower Hubbard bands at around ±2 eV, and an almost dispersionless band at E_F.

A fully *ab initio* implementation is a work in progress, that will be reported on elsewhere. Within Questaal U, J can be calculated within constrained RPA, following Ersoy et al. [163]. It has been tested on several materials; for instance, the Hubbard U and Hund's J for Ni are shown in Fig. 29 which agrees well with Ref. [163]. The Hubbard U for FeSe using QS*GW* gives $U(0) = 3.4$ eV, also close to what has been found in the literature [162]. In practical calculations today, we use a static U and adopt the fully localised limit (FLL) for double counting (Section 3.8).

We have used QS*GW*+DMFT to study single- and two-particle properties of the hole doped cuprate, $La_{2-x}Sr_xCuO_4$. The parent compound at $x = 0$ is a Mott insulator; while QS*GW* calculated non-magnetically leads to a metallic band structure, and an insulator with a wide gap when done antiferromagnetically (Fig. 27). The spins on the Cu in La_2CuO_4 fluctuate strongly and corresponding spin fluctuation diagrams are missing from QS*GW*. DMFT incorporates the local spin fluctuation diagrams with an exact solver (typically CTQMC), for the prescribed local Hamiltonian, and it can in principle, and indeed does, open a Mott gap [17,86] (charge blocking without magnetic ordering). The five Cu-3*d* partial waves in the augmentation spheres comprise the correlated subspace, and the $3d_{x^2-y^2}$ splits off to form the gap (Fig. 30), which is the paramagnetic analogue to the antiferromagnetic QS*GW* band structure (Fig. 27). We solve the correlated impurity Hamiltonian in the Cu-3*d* orbitals using the single-site CT-QMC solver within paramagnetic DMFT. Embedding local Green's functions into the QS*GW* bath, we can compute the non-local Green's functions dressed by the local DMFT self energy. They can be used to make spectral functions and the bubble diagrams entering into response functions (Section 6.3).

On doping, $La_{2-x}Sr_xCuO_4$ undergoes a metal–insulator transition when x is sufficiently large. At high temperatures, the $x = 0.12$ system is known to be metallic. We recover a metallic solution in our QS*GW*+DMFT for $x = 0.12$, using a virtual crystal approximation for x. We find the classical three peak spectral

feature with atomic-like upper and lower Hubbard bands and a low energy incoherent quasi-particle peak at E_F (Fig. 30). This structure is typical to strongly correlated metals in proximity of a localisation–delocalisation transition [156,164]. Comparing with the insulating case at $x = 0$, the spectral features for $x = .12$ has a natural explanation in the fact that spectral weight flows to low energies under doping.

Single-site DMFT incorporates the local spin fluctuation diagrams through a local self energy. It cannot explain Fermi-arc features of weakly hole-doped cuprates and momentum-dependent suppression of electron pockets in de-twinned FeSe. In weakly hole-doped cuprates Fermi arc emerges, most likely, due to spin fluctuations which are longer ranged in nature and the corresponding diagrams are absent from single-site DMFT. Similarly, in FeSe, there are nematic fluctuations (which are inherently nonlocal), and the corresponding diagrams are absent from our theory. To incorporate such diagrams one needs to go beyond the single-site DMFT; cluster DMFT [165] or diagrammatic techniques like Dual Fermions [166].

6. Susceptibilities

Susceptibilities play a central role in many kinds of material properties. In the static limit they reduce to the second derivative of the total energy with respect to the corresponding perturbation. This section presents some general properties of linear susceptibilities, focusing on the transverse spin susceptibility χ^{+-}, which is the magnetic analogue of the better known charge susceptibility. The formalism of the latter is very similar, but it is simpler as it concerns the response to a scalar potential. Also in the spin case, there is no analogue to the lowest order diagram (time-dependent Hartree approximation) in the charge case. To characterise the magnetic susceptibility beyond the noninteracting case requires a vertex such as the T matrix [141].

Here we outline some general features; other sections explain how they are implemented in Questaal. The spin susceptibility relates the induced moment $\delta\mathbf{m}$ to a perturbation $\delta\mathbf{B}$:

$$\delta\mathbf{m}(\mathbf{r},t) = -\int dt' d^3r' \chi(\mathbf{r},\mathbf{r}',t-t')\mathbf{B}(\mathbf{r}',t') \tag{177}$$

χ is a three-component tensor, $\chi^{\alpha\beta}$ with α and β Cartesian coordinates. The charge susceptibility has the same form, only the perturbing potential is the scalar electric potential, which induces a (scalar) charge density. In general, there can be cross-coupling between spin and charge, so that the full χ is a 4×4 matrix connecting spin and magnetisation.

We restrict ourselves to linear perturbations around an equilibrium point, where the unperturbed system is time invariant. χ is a function only of the difference $t - t'$, so its Fourier transform depends on a single frequency ω. If in addition the reference system is in a periodic lattice, its Fourier transform space depends only on the difference in translation vectors, whose Fourier transform has a single $\mathbf{q}$ within the Brillouin zone with $\mathbf{r}$ and $\mathbf{r}'$ limited to a unit cell. Denoting $\hat{\chi}(\mathbf{q},\mathbf{r},\mathbf{r}',\omega)$ as χ Fourier transformed in both space and time, $\hat{\chi}$ is typically what is measured by spectroscopy.

6.1. Spin susceptibility

A perturbation causes the system to generate an internal field $\mathbf{B}_{\rm xc}$, which adds to the total field, $\mathbf{B}_{\rm tot} = \mathbf{B}_{\rm ext} + \mathbf{B}_{\rm xc}$

$$\delta B^\alpha_{\rm tot} = \delta B^\alpha_{\rm ext} + \sum_\beta \frac{\delta B^\alpha_{\rm xc}}{\delta m^\beta}\,\delta m^\beta \tag{178}$$

In collinear spin structures (all spins parallel to z), transverse and longitudinal components are decoupled. Moreover, by rotating from (x,y) to $x^{+-} = x \pm iy$, the three-component χ becomes diagonal with transverse elements χ^{+-}, χ^{-+}, and longitudinal element χ^{zz}. Eq. (177) also applies to the charge channel with the substitution $\mathbf{B}\to V$ and $\mathbf{m}\to n$, so that the full χ becomes a 4×4 matrix, as noted above. In general χ^{zz} is coupled to the charge channel, while χ^{+-} and χ^{-+} are not. Also if the equilibrium system is nonmagnetic, $\chi^{+-} = \chi^{zz}$. In this work we are concerned only with χ^{+-} and the charge susceptibility $\chi^{\rm Q}$, or alternatively the dielectric function.

The "magnetic" kernel $I_{\rm xc} = \partial B_{xc}/\partial m$ is in general a function of $(\mathbf{r},\mathbf{r}',\omega)$, though it must satisfy a sum rule [110]. In many-body perturbation theory, the analogue of I is the most challenging quantity to obtain, but in the adiabatic time-dependent LDA it is static and local, and is written in the transverse case as

$$I^{+-}_{\rm xc}(\mathbf{r}) = \frac{B_{xc}(\mathbf{r})}{m(\mathbf{r}')}\delta(\mathbf{r}-\mathbf{r}') \tag{179}$$

Eq. (177) relates $\mathbf{m}$ to either $\mathbf{B}_{\rm ext}$ or $\mathbf{B}_{\rm tot}$

$$\delta\mathbf{m} = \chi_0\,\delta\mathbf{B}_{\rm tot} = \chi\,\delta\mathbf{B}_{\rm ext} \tag{180}$$

when χ_0 and χ are the bare (noninteracting) and full susceptibilities, respectively. The preceding equations imply the relation

$$[\chi(\mathbf{r},\mathbf{r}',\omega)]^{-1} = [\chi_0(\mathbf{r},\mathbf{r}',\omega)]^{-1} - I_{\rm xc}(\mathbf{r},\mathbf{r}',\omega) \tag{181}$$

The noninteracting susceptibility can be computed by linearising Dyson's equation for the perturbation $\delta\mathbf{B}_{\rm ext}$

$$-\chi_0^{+-}(\mathbf{r}_1,\mathbf{r}_2,\omega) = G_\uparrow(\mathbf{r}_1,\mathbf{r}_2,\omega)\otimes G_\downarrow(\mathbf{r}_1,\mathbf{r}_2,-\omega) \tag{182}$$

$\otimes$ indicates frequency convolution: $f(\omega)\otimes g(\omega) = i/2\pi\int_{-\infty}^{\infty} d\omega\, f(\omega-\omega')g(\omega')$. We use $\downarrow$ for spin 1 and $\uparrow$ for spin 2. Similar equations with permutations of $\uparrow$ and $\downarrow$ may be written for the other components of χ_0. In the Lehmann representation χ_0^{+-} involves the product of four eigenfunctions

$$\begin{aligned}\chi_0^{+-}(\mathbf{q},\mathbf{r}_1,\mathbf{r}_2,\omega) = &\sum_{\mathbf{k}n\downarrow}^{\rm occ}\sum_{\mathbf{k}'n'\uparrow}^{\rm unocc}\frac{\Psi^*_{\mathbf{k}n\downarrow}(\mathbf{r}_1)\Psi_{\mathbf{k}'n'\uparrow}(\mathbf{r}_1)\Psi^*_{\mathbf{k}'n'\uparrow}(\mathbf{r}_2)\Psi_{\mathbf{k}n\downarrow}(\mathbf{r}_2)}{\omega-(\epsilon_{\mathbf{k}'n'\uparrow}-\epsilon_{\mathbf{k}n\downarrow})+i\delta}\\ &+\sum_{\mathbf{k}n\downarrow}^{\rm unocc}\sum_{\mathbf{k}'n'\uparrow}^{\rm occ}\frac{\Psi^*_{\mathbf{k}n\downarrow}(\mathbf{r}_1)\Psi_{\mathbf{k}'n'\uparrow}(\mathbf{r}_1)\Psi^*_{\mathbf{k}'n'\uparrow}(\mathbf{r}_2)\Psi_{\mathbf{k}n\downarrow}(\mathbf{r}_2)}{-\omega-(\epsilon_{\mathbf{k}n\downarrow}-\epsilon_{\mathbf{k}'n'\uparrow})+i\delta}\end{aligned} \tag{183}$$

where $\mathbf{k}' = \mathbf{q}+\mathbf{k}$. We note in passing, that in a proper MBPT formulation of susceptibility, e.g. the T-matrix [141,142], is a four-point quantity: χ is solved by a Bethe–Salpeter equation involving W and a four-point analogue of Eq. (183), which at the end is contracted to two coordinates.

The Questaal code at present does not yet have a perturbative (T matrix) approach for the spin susceptibility, although one was recently reported in a QSGW framework [167]. It does have the ability to include the charge analogue of ladder diagrams to solve a Bethe–Salpeter equation for the dielectric function [116]. DMFT can also be used to build a two-particle Green's function (or susceptibility), including all local diagrams, from which a local four-point vertex can be extracted that replaces I.

The first QSGW implementation of magnetic response functions [110] was designed for local moment systems, by which is meant systems with rigid spins (RSA): when perturbed by a transverse $\delta\mathbf{B}^\perp_{\rm ext}$ $\mathbf{m}$ rotates rigidly without changing shape. (Archetypal examples are Fe, NiO, and MnAs [110].) In such a case the representation of $\chi^{+-}(\mathbf{r},\mathbf{r}',\omega)$ simplifies to

$$m(\mathbf{r})\,\chi^{+-}(\omega)\,m(\mathbf{r}') \tag{184}$$

and I can be completely determined by the sum rule [110], thus avoiding a diagrammatic calculation for it. This is Questaal's present perturbative approach to computing χ.

In the RSA $\chi^{+-}(\mathbf{r},\mathbf{r}',\omega)$ is discretised to a lattice model and can be written as $\chi^{+-}_{\mathbf{R},\mathbf{R}'}(\omega)$. This makes it convenient to construct a Heisenberg model

$$H = \sum_{RR'} J_{RR'}\delta\mathbf{m}_R\delta\mathbf{m}'_R \tag{185}$$

and extract $J_{RR'}$ from χ as

$$J_{RR'} = \frac{\delta^2 E}{\delta\mathbf{m}_R\delta\mathbf{m}_{R'}} = [\chi^{+-}_{\mathbf{R},\mathbf{R}'}(\omega=0)]^{-1} \tag{186}$$

The $\delta\mathbf{m}$ are understood to be transverse to $\mathbf{m}_R$. Thus QS*GW* can be used to determine coefficients $J_{RR'}$ entering into the Heisenberg model. This basic idea, with some enhancements, is what is described in Ref. [110].

6.1.1. Magnetic exchange interactions in the ASA

Section 6 presents general formulations of the spin and charge susceptibility. In the ASA, the static transverse susceptibility is implemented in the `lmgf` code in a formulation essentially similar to the rigid-spin approximation described there, with an additional "long-wave" approximation [168]

$$\chi(\mathbf{q}) \approx I\,\chi^{-1}(\mathbf{q})\,I \tag{187}$$

Using the sum rule, I need not be calculated but inferred from χ_0.

It is possible to compute χ_0 from the full Green's function, Eq. (57). But `lmgf` implements the classic Lichtenstein formula [169], in which exchange parameters $J_{RR'}$ are derived in terms of the auxiliary g and the "magnetic force" theorem. This latter says that the change in total energy upon spin rotation simplifies to the change in eigenvalue sum; it is a special instance of the Hellmanm–Feynman theorem. It was realised sometime later [168] that the Lichtenstein formula is correct only in the $q \to 0$ limit (the long-wave approximation), beginning to deviate at around $k = 0.25{\cdot}2\pi/a$ in elemental 3d magnets. In the notation of this paper, Lichtenstein's formula reads

$$J^{\ell\perp}_{RR'} = \frac{1}{2\pi}\int^{\varepsilon_F} d\varepsilon\ \mathrm{Im}\mathrm{Tr}_L\left[\delta P_{R\ell}\left(g^{\uparrow}_{RLR'L'}\,g^{\downarrow}_{R'L'RL} + g^{\downarrow}_{RLR'L'}\,g^{\uparrow}_{R'L'RL}\right)\delta P_{R'\ell'}\right] \tag{188}$$

Questaal implements it for crystals, and for alloys within the CPA. See Refs. [49,170,171] for a detailed description.

6.1.2. Comparing QSGW and LDA spin response functions

The *GW* code has an implementation of spin response functions within the rigid spin approximation as described Section 6. This avoids direct calculation of the vertex, and is very simple, but is restricted to local-moment systems. Ref. [110] applied this approach to NiO, MnO and MnAs, with excellent results.

Questaal has the ability to compute spin waves (SWs) directly, or to extract parameters J entering into the Heisenberg model, Eq. (185). One application – $Mn_xGa_{1-x}As$ alloys – are particularly instructive because they continue to attract interest as a spintronics material. Here we will compare the (ASA) Lichtenstein formula against full-potential, rigid-spin result using both LDA and QS*GW* Hamiltonians. Computing properties for random alloys of any x is feasible within the ASA, via either the Lichtenstein formula or relaxing large structures using spin statics, Section 2.19. But a similar study is not feasible today in QS*GW*, so instead we consider four-atom Special QuasiRandom Structures [140] to simulate $Mn_{0.25}Ga_{0.75}As$ and $Mn_{0.75}Ga_{0.25}As$, and also the $x = 1$ case. Mn's large local moment of around $4\mu_B$ makes the approximations in Questaal's QS*GW* magnetic response good ones.

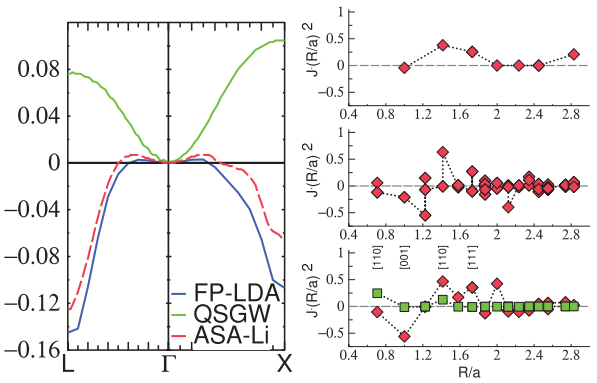

Fig. 31. Left: spin waves in Zb-MnAs (eV) calculated in the *GW* code with QS*GW* potential (green) and LDA potential (blue), and also in the ASA-LDA with the Lichtenstein formula, Eq. (188). Right: Red diamonds show Heisenberg parameters J Eq. (185) calculated in the ASA with Lichtenstein formula for a $MnGa_3As_4$ SQS structure (top), $MnGa_2As_2$ SQS structure (middle), and Zb-MnAs (bottom), as a function of neighbour distance. In the SQS structures, there can be inequivalent neighbours with the same connecting vector. LDA results calculated from the *GW* machinery look similar. Green squares: same parameters in ZB-MnAs using as QS*GW* potential. J is in mRy. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

DFT predicts $Mn_xGa_{1-x}As$ to become a spin glass when $x \gtrsim 0.35$, though T_c depends on what fraction of Mn go into interstitial sites, and how the Mn are ordered on the Ga sublattice. One application of the Questaal code was to show how Mn in alloys that favour ordering on the (201) orientation can optimise T_c [172]. In the pure Zb-MnAs case, $x = 1$, LDA predicts to be strongly antiferromagnetic, with negative spin wave frequencies everywhere in the Brillouin zone (left panel, Fig. 31). The figures show SWs calculated in the ASA using the Lichtenstein formula, Eq. (188). Its long-wave approximation should be accurate for small q, but it underestimates the strength of J for large q [168]. This is apparent in Fig. 31, comparing the LDA and Lichtenstein formulas.

This finding seems to contradict a measurement on a quantum dot of Zb-MnAs (Ref. [173]), which predicts it to be ferromagnetic. Indeed, QS*GW* calculations of the same system show that spin wave frequencies are everywhere positive (left panel, Fig. 31). T_c based on Heisenberg parameters extracted magnon peaks in the SW spectrum is positive, of order 600k (somewhat larger than what Okabayashi observed).

Exchange parameters J were extracted for three compositions, in LDA and for the pure MnAs case, in QS*GW* (right panel, Fig. 31). At 25% composition J is still positive, but the nearest-neighbour J is already very small. At 50%, J is negative on average, and a spin glass is predicted. In pure Zb-MnAs, J oscillate in sign with neighbour distance, but the nearest neighbours (particularly along [001]) are strongly antiferromagnetic, leading to negative frequencies in the SW spectrum. QS*GW* shows a marked contrast. Only the NN is important, with $J > 0$. QS*GW* and LDA are so different because the LDA underestimates the exchange splitting between spin-up and spin-down 3d states, by $\sim$1 eV. This pushes the minority d too close to E_F, and gives rise to long-range, antiferromagnetic interactions.

6.2. Optical response functions in Questaal

The linear optical response is described by the imaginary part of the dielectric response function $\epsilon_2(\omega)$ or equivalently

the real part of the frequency dependent conductivity $\sigma_1(\omega) = -i\omega\epsilon_2(\omega)/4\pi$ at frequencies in the ultraviolet–visible (UV–VIS) range. From it one can obtain the real part $\epsilon_1(\omega)$ by the Kramers–Kronig relation and hence other relevant functions such as the complex index of refraction $\tilde{n}(\omega) = \sqrt{\epsilon_1(\omega) + i\epsilon_2(\omega)} = n(\omega) + i\kappa(\omega)$ with $\kappa(\omega)$ the extinction coefficient and $n(\omega)$ the real part of the index of refraction. The reflectivity $R(\omega)$ is then obtained via the Fresnel equations. For instance, the normal incidence reflectivity

$$R(\omega) = \left|\frac{\tilde{n}(\omega) - 1}{\tilde{n}(\omega) + 1}\right|^2 \tag{189}$$

is a directly measurable quantity. The absorption coefficient

$$\alpha(\omega) = \frac{2\omega\kappa(\omega)}{c} = \frac{\omega\epsilon_2(\omega)}{n(\omega)c} \tag{190}$$

is measurable in transmission and spectroscopic ellipsometry allows one to measure directly $\epsilon_1(\omega)$ and $\epsilon_2(\omega)$.

There are several levels of theory at which one can obtain the optical dielectric function supported in the Questaal package. Typically we need the macroscopic dielectric function:

$$\epsilon_M(\omega) = [\epsilon^{-1}_{G=G'=0}(\mathbf{q} \to 0, \omega)]^{-1} \tag{191}$$

This is essentially obtained as a byproduct of the *GW* method (see Section 4). One way to obtain this limit is to take a small but finite $\mathbf{q}$ and use `lmgw` to obtain the plane wave matrix elements of the inverse dielectric function. The direction of the finite $\mathbf{q}$ then defines the specific component of $[\epsilon_2(\omega)]_{\alpha\alpha}$ which is actually a tensor. If one simply calculates instead

$$\epsilon_{G=G'=0}(\mathbf{q} \to 0, \omega) \tag{192}$$

one obtains the value without local-field corrections.

Alternatively, however, one can take the limit $\mathbf{q} \to 0$ analytically and arrive at the Adler–Wiser formula in the RPA,

$$[\epsilon_2(\omega)]_{\alpha\beta} = \frac{8\pi^2 e^2}{\Omega\omega^2} \sum_n \sum_{n'} \sum_{\mathbf{k}\in \mathrm{BZ}} f_{n\mathbf{k}}(1 - f_{n'\mathbf{k}}) \times \langle\psi_{n\mathbf{k}}|v_\alpha|\psi_{n'\mathbf{k}}\rangle\langle\psi_{n'\mathbf{k}}|v_\beta|\psi_{n\mathbf{k}}\rangle\delta(\omega - \varepsilon_{n'\mathbf{k}} + \varepsilon_{n\mathbf{k}}) \tag{193}$$

Here Ω is the unit cell volume and $f_{n\mathbf{k}}$ are the occupation numbers, which in principle are given by the Fermi function at finite temperature but are in practice taken to be 0 or 1 for empty and occupied states respectively. The v_α are the components of the velocity operator $\mathbf{v} = \dot{\mathbf{r}} = (i/\hbar)[H, \mathbf{r}]$. For systems with at least orthorhombic symmetry the $\epsilon_2(\omega)$ tensor is diagonal in the Cartesian components. This well-known equation in the independent particle and long-wavelength form gives the $\epsilon_2(\omega)$ in terms of matrix-elements of the velocity operator and vertical interband transitions and can also be obtained from the Kubo formula for the optical conductivity. Within the all-electron methods (no non-local pseudopotentials) and if *GW*-self-energies are not included, the velocity matrix elements are equivalent to the momentum matrix elements $\mathbf{v} = \mathbf{p}/m$ and can be obtained straightforwardly from the ∇ operator matrix elements within the muffin-tin spheres and using a Fourier transform for the smooth part of the wave functions in the full-potential implementation. The matrix elements of the ∇ operator between spherical harmonics times radial functions inside the muffin-tin spheres is obtained using the well-known gradient formula [174]; more generally, see Section 3.2 and Appendix D.

When the non-local self-energy operator or its Hermitian version are included, however, $(im/\hbar)[H, \mathbf{r}]$ is no longer equal to the momentum operator. To correct this, the approximation proposed by Levine and Allan [175,176] can be used, which consists in rescaling the matrix elements by a factor $(\varepsilon_{n'\mathbf{k}} - \varepsilon_{n\mathbf{k}})/(\varepsilon^{LDA}_{n'\mathbf{k}} - \varepsilon^{LDA}_{n\mathbf{k}})$. This is exact when the LDA and *GW* eigenfunctions are the same, and it works well in weakly correlated systems. However, it breaks down when correlations become strong, as in NiO. Questaal also has an approximate form for the proper non-local contribution to the velocity operator. As of this writing, it takes into account only intercell contributions. This approximation apparently does not work well in strongly correlated systems, and results with full matrix element will be reported in due course [129].

The long-wave length approximation without local field (or excitonic) effects of $\varepsilon_2(\omega)$ can thus be obtained both in the ASA `lm` and the full-potential `lmf` codes and, when reading in the QS*GW* self-energy, can use the corrected quasiparticle energies rather than the Kohn–Sham eigenvalues. While approximate, it has the advantage that the $\mathbf{q} \to 0$ limit is taken analytically and that one can decompose the optical response in contributions from each occupied-empty band pair. By furthermore plotting vertical interband transition energies for any given pair, one can analyse where the bands are parallel and give their largest contribution to the joint density of states. This provides a way to relate the critical points or van-Hove singularities in the optical response to particular interband transitions. `lmf` and `lm` use this to enable decomposition into band-pair contributions if desired; it can also resolve contributions to ϵ by k.

The code also allows one to calculate dipole optical matrix elements between core states and valence or conduction band states and this can be used to model X-ray absorption (XAS) and emission (XES) spectroscopies. Furthermore, one can calculate Resonant X-ray Emission Spectroscopy (RXES) also known as Resonant Inelastic X-ray Scattering (RIXS) using the Kramers–Heisenberg equations [177]. The interesting point about this spectroscopy is that it allows to probe transitions between valence and conduction band states of the same angular momentum on a given site. In fact, let us say one considers as X-ray edge the K-edge or 1*s* core level, then the RXES probes transitions between valence bands and conduction bands of *p* angular momentum on the site of core hole that are in resonance with the difference of the absorbed and subsequently coherently emitted X-ray photon transitions from these conduction and valence states to the same core-hole. In optical measurements in the VIS-UV region, on the other hand, one probes only ℓ to $\ell \pm 1$ dipole allowed transitions. Using *p*-levels as core hole RIXS thus probes $d - d$ transitions, which often may be strongly influenced by many-body effects. The Kramers–Heisenberg formula assumes the wavevector of the X-ray can be neglected so that only direct transitions at the same $\mathbf{k}$-point are allowed and within this approximation it allows one in principle to extract band-dispersions in a manner complementary to ARPES.

It was established 20 years ago that RPA approximation to the dielectric function can be dramatically improved by adding ladder diagrams via a Bethe–Salpeter equation [127,128] (BSE) in simple *sp* semiconductors. Traditionally the method is used by substituting *GW* eigenvalues for the DFT ones obtained from one-shot *GW* (Section 4). For simple *sp* semiconductors where LDA and *GW* eigenfunctions are similar, this works well. It does not, however, when correlations become strong, e.g. in Cu_2O, where it was shown that $\epsilon(\omega)$ calculated from the BSE starting from a QS*GW* reference described the measured spectrum quite well [178]. Recently Cunningham et al. studied optics from BSE using QS*GW* as reference for a number of systems, and found very good agreement with experiment even in correlated systems such as NiO [116] and the monoclinic phase of VO_2 [119]. However, when spin fluctuations are large and QS*GW* does not describe the underlying band structure well, as in La_2CuO_4 (see Section 4.4), the description begins to break down. Recently Cunningham has added ladders to *W* in the QS*GW* self-consistency

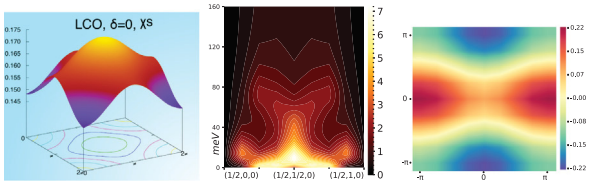

Fig. 32. (left) Real part of $\chi(q,\omega=0)$ for La_2CuO_4 showing a dominant peak at $\mathbf{q}=(\pi,\pi)$. (middle) Imaginary part of $\chi(q,\omega)$ (arbitrary units) for $La_{1.88}Sr_{0.12}CuO_4$ showing that spin fluctuations get gapped at (π,π) below 145 K. (right) Nodal superconducting gap structure for the hole-doped La_2CuO_4 showing $d_{x^2-y^2}$ gap symmetry. The colour bar shows the negative (in blue) and positive (in red) superconducting gap magnitudes in arbitrary units passing through node where the gap closes (in yellow).

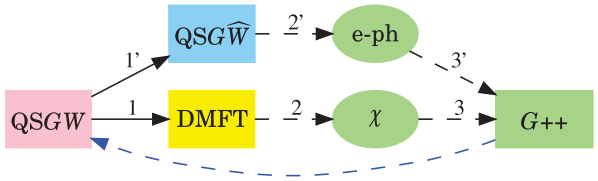

Fig. 33. Questaal strategy for solving the many-electron problem.

cycle. This seems to dramatically improve on the description of $\epsilon(\omega)$, even in an extreme case such as La_2CuO_4. These capabilities are embedded in the standard Questaal distribution.

In the ASA code `lm` second harmonic generation coefficients can also be calculated but this portion of the code has not recently been updated to become applicable in the full-potential `lmf` implementation. The calculation of NLO coefficients is a bit more complex even within the long-wavelength independent particle approximation because of the need to include both intra- and inter-band transitions and disentangle them in a careful manner. The way to obtain divergence free expression was described in a series of papers by John Sipe and coworkers [179–181]. The version implemented in the code [182] uses Aversa's length gauge formalism. The intra-band matrix elements $\mathbf{r}_i$ and inter-band $\mathbf{r}_e$ in this formalism are

$$\langle n\mathbf{k}|\mathbf{r}_i|m\mathbf{k}'\rangle = \delta_{nm}[\delta(\mathbf{k}-\mathbf{k}')\xi_{nn} + i\nabla_{\mathbf{k}}\delta(\mathbf{k}-\mathbf{k}')]$$
$$\langle n\mathbf{k}|\mathbf{r}_e|m\mathbf{k}'\rangle = (1-\delta_{nm})\delta(\mathbf{k}-\mathbf{k}')\xi_{nm} \quad (194)$$

with

$$\xi_{nm} \equiv \frac{(2\pi)^3 i}{\Omega}\int_{\Omega} d^3r\, u^*_{n\mathbf{k}}(\mathbf{r})\nabla_{\mathbf{k}}u_{m\mathbf{k}}(\mathbf{r}) \quad (195)$$

in which Ω is the volume of the unit cell and $u_{n\mathbf{k}}(\mathbf{r})$ is the periodic part of the Bloch function. The ξ_{nn} can be recognised to be a Berry connection. The manipulation of the matrix elements of $\mathbf{r}_i$ in commutators is non-trivial but described in detail in Aversa and Sipe [179]. An update of these parts of the code to incorporate them in `lmf` and to be made compatible with QS*GW* band structures as input is intended.

See also Section 6.3.

6.3. Response functions within DMFT

With a converged self-energy, the CTQMC can sample the two-particle Green's function [183] to obtain local spin and charge vertices. Finally, the non-local Bethe–Salpeter equations (BSE) in spin, charge and superconducting channels [183] can be solved from a local vertex and a non-local polarisation bubble in the respective channels. This allows us to compute the corresponding real [86] and imaginary part of susceptibilities [86,184] in those channels. In Fig. 32 we show the real part of the spin susceptibilities in $x=0$ $La_{2-x}Sr_xCuO_4$, adapted from Ref. [86]. There is a peak at $\mathbf{q}=(\pi,\pi)$ suggesting dominant Néel anti-ferromagnetic spin fluctuations (Fig. 32).

We compute the imaginary part of the spin susceptibilities along the line $(h,k,l)=(0,1/2,0)-(1,1/2,0)$ and find that the peak at (1/2, 1/2, 0) gets spin gapped (vanishing weight at $\omega=0$) when $x=0.12$ below a certain temperature, consistent with experimental observations [185]. Additionally, solving the BSE in the p-p superconducting channel we recover a superconducting gap function with $d_{x^2-y^2}$ symmetry (right panel, Fig. 32). Our QS*GW*+DMFT implementation can be successfully applied to a wide range of systems; weakly correlated metals [115], strongly correlated metals [184] and correlated Mott insulators [86].

7. Towards a high-fidelity solution of the many-body problem

Solving the many-electron problem with high-fidelity is a formidably difficult task. Our strategy to accomplish this relies on the following premises:

(i) many-body solutions are best framed around a non-interacting starting point, and we believe that QS*GW* is the best choice among them, by construction;
(ii) charge fluctuations governed by long-range interactions, but they can be treated accurately with low-order perturbation theory; and
(iii) spin fluctuations are governed mostly by single-site effective Hamiltonian (or action). The short-range interactions can be too strong to handle perturbatively, as can be seen by the non-analytic behaviour of one- and two-particle quantities, and their high degree of sensitivity to small perturbations, e.g. change in temperature. Nonlocal contributions are much weaker and can be treated in perturbation theory.

Charge and spin fluctuations have very different characteristic energy scales: plasmon frequencies are typically of order 5 eV while typical magnon frequencies are of order 100 meV or less. As a consequence they mutually interact weakly so that the self-energy is predominated by independent contributions from spin and charge channels. This is the premise of the Bose Factor Ansatz which has been found to be effective in a DMFT framework [186].

We have framed a hierarchical strategy around these premises (Fig. 33). At the lowest level is QS*GW*, which depends only minimally on DFT. We think this is of central importance to frame a consistently reliable theory, for reasons we have pointed out in Section 4.1.

Referring to the figure, QS*GW* is adequate for many purposes. When not, there are two routes: a perturbative, nonlocal path $(1',2',3')$: low-order diagrams are added to W to make $(W\rightarrow\widehat{W})$. To date we have added ladders to the charge channel; in progress is a project to add the electron–phonon interaction perturbatively [187], effectively adding another bosonic contribution to W. Also possible, but not yet accomplished in Questaal, is to add low order spin fluctuation diagrams such as the T matrix. All of them make a better G (G^{++} in the Figure). The second path begins as nonperturbative, local path (1). From a local interaction nonlocal susceptibilities can be constructed (2). We have shown a few instances of this in Section 6.3, and the results are remarkably good in cases we have studied so far, e.g. Ref. [184]. Finally, the susceptibilities can be used to add a new diagrammatic contribution (3) to G (G++). The simplest addition is Dual Fermions [166].

Perhaps more satisfactory would be to have a single, unified approach, for example the Diagrammatic Monte Carlo (DiagMC)

method [188]. But it is formidably difficult to make this method work practically in real systems, because of the enormous time and memory costs that realistic (as opposed to model) Hamiltonians require. We think that the JPO basis alluded to in Section 3.12 is perhaps the most promising framework to realise DiagMC for realistic Hamiltonians. Even in such a case an optimal approach would likely closely resemble Fig. 33, but with some cross-linking between paths (1) and (1′), using a diagrammatic Monte Carlo solver only to include nonlocal diagrams beyond the RPA.

8. Software aspects

The software package has a long history and has featured a high level of modularisation from its early days. The different parts interoperate through shared interfaces and file data formats.

The implementation is mainly in procedural Fortran 77. Fixes and improvements as well as new feature code uses modern Fortran features (f2008+) for convenience as well as improved reliability and interoperability. Despite f77's shortcomings and questionable reputation, sound software engineering practices have been observed, there is (1) extensive code documentation in an uniform format, (2) almost no shared mutable state (i.e. common blocks or module variables); (3) action through side effect is avoided and data is passed through arguments, (4) unnecessary temporary data copies are avoided, (5) data structures are organised with performance in mind and most heavy number crunching is outsourced to performance libraries implementing the BLAS, LAPACK, FFTW3 APIs, (6) the code is written in an uniform style with consistent flow patterns. This practice has been extended through the use of modern version tracking and continuous integration pipelines incorporating regression, coverage and performance testing, and minimal style policy enforcement.

8.1. Release policy

We use the popular distributed version tracking system `git`. The public online hosted repository is hooked to continuous integration pipelines executing the steps above on each push event. New features are developed in separate feature branches, if pushed to the public online repository, these are visible to all users. When judged safe and reasonably useable, feature branches are merged to the main development branch ("lm"). After extensive use, the development branch is merged to the release branch ("master") and a new release is tagged with a numerical version in the format vmajor.minor.patchlevel. If/when issues are discovered and fixed the patchlevel is incremented in a new tag and the new commits are merged back to the development branch and from there to the feature branches. This approach reduces the maintenance burden significantly, however it does mean that once the minor version is incremented there are no more patches offered to the older versions and users are strongly encouraged to update to the latest version available. Since new developments are effectively done by interested users, and there is as of yet no contractual support offered, we feel this is a justified arrangement and in the long term offers users new features and improved performance with effectively close to no risk. The major version is only incremented when a very significant, possibly incompatible rewrite of core components or the input system has been performed.

New code and changes to existing code require: (1) coding standards pass; (2) regression tests suite pass (approximately 350 tests as of this writing); and (3) test coverage of at least 90% of code.

Table 8
Main executables in the Questaal suite.

Full-potential	
`blm`	Automatic input file generator
`lmf`	Main band code
`lmfa`	Free atom solver
`lmfgwd`	Interface to *GW*
`lmfdmft`	Interface to DMFT
Atomic sphere approximation	
`lm`	ASA band code
`lmgf`	ASA crystal Green's function code
`lmpg`	ASA principal layer Green's function
`lmstr`	ASA structure constants
`tbe`	Empirical tight-binding
Post-processing and utilities	
`fplot`	A graphics package
`pldos`	A DOS postprocessor
`plbnds`	A bands postprocessor
`lmchk`	Checks structure-related quantities
`lmdos`	Assembles partial DOS
`lmfgws`	Dynamical self-energy postprocessing
`lmscell`	Supercell maker
`mcx`	A matrix calculator
Editors	
`lmf|lm --rsedit`	Restart file
`lm --popted`	Optics
`lmf --wsig~edit`	Static Σ_0
`lmfgws --sfuned`	Dynamic Σ
`lmf --chimedit`	Magnetic susceptibility
`lmscell --stack`	Superlattice

8.2. Parallelisation

The programmes of the package support multilevel parallelisation multiprocessing through the message passing interface (MPI) and simultaneous multithreading through performance libraries and (when necessary) manual OpenMP directives. The full-potential program uses multiprocessing mostly during the Brillouin zone sampling, outside of the q-loop the MPI processes cooperate on the local potential generation, only multithreading is available within the processing of each q-point. The empirical polarisable ion tight binding program offers more flexibility in this regard by allowing simultaneous assignment of groups of processes to each q-point as well as spreading different blocks of q-points over groups of processes [189]. A somewhat similar approach has been recently developed by Martin Lüders for the *GW* code. While it is available in a feature branch it is not yet considered production ready and has not been merged to the main branch. The GW code is fairly multithreaded and for systems of at least 15–20 atoms performs well on many-core wide vector architectures like Intel's Xeon Phi family of processors (tested on x200, Knight's Landing). A small but performance-critical part was also ported to Nvidia's CUDA parallel platform and shows promising performance on hardware with good double precision floating point capabilities (for example Titan, P100, V100 compute cards). The polarisable ion tight binding program makes more extensive use of CUDA; it is described in [189].

To ease parallel IO, classic fortran binary files are being moved to HDF5 based data files. This also improves interoperability with various postprocessing environments because the HDF5 libraries offer bindings to a variety of popular programming languages.

8.3. Usability

An ordinary text command line is all one needs to be able to use the programmes from the package. The main executables (see Table 8) support a number of flags, notable among which is "`--input`", it causes the full input understood by the executable to

```
# HCP Gd
VERS FP=7 LM=7

% const pwmode=0 pwemax=0
HAM
 AUTOBAS[MTO=12 LOC=1]
 XCFUN=0 101 130 # libxc's PBE-GGA
 NSPIN=2
 GMAX=10.4
 PWMODE={pwmode} PWEMAX={pwemax}

% const nk=-1000
BZ
 NKABC={nk}

ITER
 NIT=30 CONV=1e-5 MIX=B6,b=0.75

STRUC
 NSPEC=1 NBAS=2
 ALAT=6.9484
 PLAT= sqrt(3/4) -1/2  0
       0          1    0
       0          0    1.6329932

SPEC
 ATOM=Gd Z=64 R=3.474 LMX=3 LMXA=6
         MMOM=0 0 0 7 # mom by l
         PZ=0 0 0 5.4 # 5f high LO

SITE
 ATOM=Gd XPOS=1/3 2/3 1/4
 ATOM=Gd XPOS=2/3 1/3 3/4
```

Fig. 34. Basic input file: a full-potential GGA calculation for HCP Gd with $7\mu_B$ $l = 3$ starting moment. A preprocessor directive (`nk`) is used to specify the k-sampling, requesting 1000 points in the full Brillouin zone: the generated grid (12,12,7) has components corresponding to the relative BZ vector lengths. `nk`, like `pwemax`—which flags the inclusion of LAPW basis functions—can be changed from the command line. The LMTO basis is to be setup automatically, caused by the `AUTOBAS` token in the `HAM` (Hamiltonian) section.

be printed together with default values and short documentation for each token. Another very valuable and possibly unique feature is the ability to override almost any input file value through the command line through the small embedded preprocessor which renders the input transparent internally (Fig. 34). A third is the suite of special-purpose editors (see Table). There is extensive documentation online together with many tutorials and ready made example script snippets. Users are encouraged to report issues and offer ideas through the online ticketing system and contribute improvements or fixes through pull requests or inline patches.

8.4. Selected publications from Questaal

In this section we point to original papers where new capabilities were developed within Questaal. Some the concepts are not original with Questaal, though many are:

- an early DFT calculation of the Schottky barrier height at a metal/semiconductor contact, showing the importance of screening at the interface [190];
- an early calculation of alloy phase diagrams combining statistical theory and DFT [191];
- a systematic technique for deriving force theorems within DFT [9];
- the first formulation of adiabatic spin dynamics within DFT [11] and its application to explain the Invar effect in permalloy [192];
- a formulation of self-consistent empirical tight-binding theory [29];
- a formulation of Electron Energy Loss Spectroscopy in DFT [193];
- the original description of Questaal's full-potential method, Section 3 [31];
- the original description of Questaal's all-electron *GW* approximation [194], and first calculation of RPA total energy in a solid [195];
- solution of the Boltzmann transport equation combined with DFT [196];
- the original formulation of Quasiparticle Self-Consistent *GW* [16], demonstration of its wide applicability [83], and a detailed description [84];
- Questaal's implementation of nonequilibrium electron transport in nanosystems [14], and significant applications to magnetic transport [22,24,25,197,198];
- a fusion of genetic algorithms and exchange interactions to predict and optimise critical temperatures in multi-component systems [172];
- Dresselhaus terms calculated *ab initio* with high fidelity in zincblende semiconductors [107,109,199];
- first *GW* description of 4*f* systems, showing the tendency to overestimate splitting between occupied and unoccupied *f* states [135];
- spin wave theory within QS*GW* [110];
- prediction of Impact Ionisation rates with QS*GW* [113];
- the PMT method of Section 3.10 [54];
- a formulation of tunnelling transport within QS*GW* [112];
- effect of spin–orbit coupling on the QS*GW* self-energy [77];
- approximate description of transient absorption spectroscopy within QS*GW* [200];
- Questaal's first application of QS*GW*+DMFT, showing the need to go beyond *GW* when spin fluctuations are important [115];
- addition of ladder diagrams to dielectric function [116];
- Fröhlich contribution to renormalisation of QS*GW* energy bands [131]; and
- first application of QS*GW*+DMFT+BSE for susceptibilities, Section 6.3 [86].

8.5. Distribution and licensing

Historically the code has been distributed as a set of source tarballs, in addition registered users can git-clone our online repository. The code is distributed under the terms of the General Public license version 3. Contributions under a compatible license are welcome.

9. Conclusions

Questaal is a descendant of one of the early all-electron methods developed in Stuttgart. It has gradually evolved to its present form, but as no paper has been written for any of its stages the present work is intended to summarise much of this evolution. Our aim was to present the many expressions from classic works, combined with previously unpublished recent developments, in a unified way to show the connections between the parts.

We made an attempt to show Questaal's strengths as well as its limits, within varying levels of theory. We think Questaal provides a very promising path to efficiently solve the electronic structure problem with high fidelity. It is unique in its potential to span such a wide range of properties and materials, and with varying levels of approximation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Dimitar Pashov: Conceptualization, Methodology, Software, Validation, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization. **Swagata Acharya:** Methodology, Software, Validation, Investigation, Writing - original draft, Writing - review & editing, Visualization. **Walter R.L. Lambrecht:** Conceptualization, Methodology, Software, Validation, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization. **Jerome Jackson:** Validation, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization. **Kirill D. Belashchenko:** Methodology, Software, Validation, Investigation, Writing - original draft, Writing - review & editing, Visualization. **Athanasios Chantis:** Methodology, Software, Validation, Investigation, Writing - original draft, Writing - review & editing, Visualization. **Francois Jamet:** Methodology, Software, Validation, Investigation, Writing - original draft, Writing - review & editing, Visualization. **Mark van Schilfgaarde:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing - original draft, Writing - review & editing, Visualization, Supervision, Project administration, Funding acquisition.

Acknowledgements

Many people have contributed to the Questaal project. Much of the mathematics and basic algorithms in the `lmf` implementation were largely the work of Michael Methfessel. *GW* was formulated and implemented largely by Ferdi Aryasetiawan and Takao Kotani, and QS*GW* by Sergey Faleev. Martin Lüders wrote a parallel version of this code and was involved in some validation aspects. Derek Stewart wrote the original code for Landauer–Buttiker transport, and Faleev extended it to the nonequilibrium case. Pooya Azarhoosh improved on the optics and performed TAS studies. Alena Vishina redesigned the fully relativistic Dirac code, making it self-consistent. Herve Ness has redesigned significant parts of the layer Green's function technique.

Brian Cunningham and Myrta Grüning added ladder diagrams to *W*, and Savio Laricchia is close to completing a QS*GW*-based approach to the electron–phonon interaction and related quantities. Both of these additions will have significant impact in the future. Also, many thanks to Kristjan Haule and Gabi Kotliar for helping us get started in DMFT, and to Ole Andersen for being the founder and inspiration for so many of the ideas behind Questaal.

Finally, many thanks to Andy Millis and the Simons Foundation, who enriched Questaal in fundamental ways, and to EPRSC which made it possible to become a community code (CCP9 flagship project: EP/M011631/1).

We gratefully acknowledge PRACE for awarding us access to SuperMUC at GCS@LRZ, Germany, the UK Materials and Molecular Modelling Hub for computational resources, which is partially funded by EPSRC, United Kingdom (EP/P020194/1) and STFC Scientific Computing Department's SCARF cluster (www.scarf.rl.ac.uk). J.J. acknowledges support under the CCP9 project "Computational Electronic Structure of Condensed Matter", part of the Computational Science Centre for Research Communities (CoSeC).

WRLL acknowledges support from the US Department of Energy-Basic Sciences under grant No. DE-SC0008933. K.D.B. acknowledges support from the National Science Foundation through Grant No. DMR-1609776.

Appendix

An augmentation radius is labelled by symbol s. Subscript R is used to denote a site-dependent quantity, such as the augmentation radius s_R. When R appears in a subscript of a function of $\mathbf{r}$ such as $H_{RL}(\mathbf{r})$, it implies that $\mathbf{r}$ is relative to $\mathbf{R}$, i.e. $\mathbf{r}-\mathbf{R}$. When a symbol refers to angular momentum, an upper case letter such as L, refers to both angular quantum number and magnetic quantum number parts, ℓ and m.

Appendix A. Definition of real and spherical harmonics

Where spherical harmonics are used, Questaal uses the same definitions as Jackson [201]:

$$Y_{\ell m}(\theta,\phi)=\left[\frac{(2\ell+1)(\ell-m)!}{4\pi(\ell+m)!}\right]^{\frac{1}{2}}P_\ell^m(\cos\theta)e^{im\phi} \tag{A.1}$$

$$P_\ell^m(x)=(-1)^m\frac{(1-x^2)^{m/2}}{2^l l!}\frac{d^{\ell+m}}{dx^{\ell+m}}(x^2-1)^\ell \tag{A.2}$$

The $-m$ and $+m$ functions are related by [see Jackson (3.51) and (3.53)]

$$P_\ell^{-m}(x)=(-1)^m\frac{(\ell-m)!}{(\ell+m)!}P_\ell^m(x) \tag{A.3}$$

$$Y_{\ell,-m}(\hat{\mathbf{r}})=(-1)^m Y^*_{\ell m}(\hat{\mathbf{r}}) \tag{A.4}$$

Questaal mostly uses real harmonics Y_{lm}, which are related to the spherical harmonics as

$$\begin{aligned}\mathrm{Y}_{\ell 0}(\hat{\mathbf{r}})&=Y_{\ell 0}(\hat{\mathbf{r}})\\ \mathrm{Y}_{\ell m}(\hat{\mathbf{r}})&=\frac{1}{\sqrt{2}}[(-1)^m Y_{\ell m}(\hat{\mathbf{r}})+Y_{\ell,-m}(\hat{\mathbf{r}})]\\ \mathrm{Y}_{\ell,-m}(\hat{\mathbf{r}})&=\frac{1}{\sqrt{2}i}[(-1)^m Y_{\ell m}(\hat{\mathbf{r}})-Y_{\ell,-m}(\hat{\mathbf{r}})]\end{aligned} \tag{A.5}$$

It also uses *real harmonic polynomials*, [note: substituted $l\to\ell$]

$$\mathcal{Y}_{\ell m}(\mathbf{r})=r^\ell \mathrm{Y}_{\ell m}(\hat{\mathbf{r}}), \tag{A.6}$$

which are real polynomials in x, y, and z.

The product of two spherical harmonic polynomials can be expanded as a linear combination of these functions in the same manner as ordinary ones

$$\mathrm{Y}_K(\hat{\mathbf{r}})\mathrm{Y}_L(\hat{\mathbf{r}})=\sum_M C_{KLM}\mathrm{Y}_M(\hat{\mathbf{r}}) \tag{A.7}$$

$$\mathcal{Y}_K(\mathbf{r})\mathcal{Y}_L(\mathbf{r})=\sum_M C_{KLM}\,r^{k+\ell-m}\,\mathcal{Y}_M(\mathbf{r}) \tag{A.8}$$

The Gaunt coefficients C_{KLM} are nonzero only when $k+\ell-m$ is an even integer, so the r.h.s. is also a polynomial in (x,y,z).

Note also, from Eq. (A.6) it follows that if α is purely imaginary

$$\mathcal{Y}_{\ell m}(\alpha\mathbf{r})=(-1)^l\,\mathcal{Y}^*_{\ell m}(\alpha\mathbf{r}) \tag{A.9}$$

It also follows from Eq. (A.6) and the discussion around Eq. (138) that

$$\begin{aligned}\nabla_p\mathcal{Y}_L(\mathbf{r})&=(2\ell+1)\sum_K C^-_{KL;p}\,\mathcal{Y}_K(\mathbf{r})\\ -\nabla_p(r^{-2l-1}\mathcal{Y}_L(\mathbf{r}))&=(2\ell+1)\sum_K C^+_{KL;p}\,\mathcal{Y}_K(\mathbf{r})\end{aligned} \tag{A.10}$$

for $p=1,2,3$. $r^{-2l-1}\mathcal{Y}_L(\mathbf{r})$ and $\mathcal{Y}_L(\mathbf{r})$ are ordinary Hankel and Bessel functions of energy zero. In the top (bottom) line, $k=\ell+1$ ($k=\ell-1$). Eqs. (A.10) may be taken as a definition of $C^{(\pm)}_{KL;p}$. Similarly

$$\mathbf{r}_p\mathcal{Y}_L(\mathbf{r})=\sum_K\left\{C^{(+)}_{KL;p}+C^{(-)}_{KL;p}r^2\right\}\mathcal{Y}_K(\mathbf{r}) \tag{A.11}$$

$C^{(-)}$ is the transpose of $C^{(+)}$:

$$C^{(+)}_{KL;p} = C^{(-)}_{LK;p} \tag{A.12}$$

but we keep them separate to distinguish terms that take $Y_{\ell,m}$ into $Y_{\ell+1,m'}$ and those that map into $Y_{\ell-1,m'}$.

Appendix B. Definition of Hankel and Bessel functions

To distinguish radial parts of spherical functions from solid versions, we denote spherical parts with subscript ℓ, and solid functions with subscript L, which refers to both the ℓ and m parts.

The Questaal codes generally follow Methfessel's definitions for Hankel and Bessel functions. Writing $k = \sqrt{E}$ with $\mathrm{Im}\, k \geq 0$, they are related to standard spherical Neumann functions n_ℓ and Bessel functions j_ℓ as follows:

$$\begin{aligned} H_L &= \mathcal{Y}_L(-\nabla)\, h_0(r) \\ &= k^{\ell+1} n_\ell(kr) \mathrm{Y}_L, && E > 0 \\ &= (ik)^{\ell+1} n_\ell(kr) \mathrm{Y}_L, && E < 0 \\ J_L &= k^{-\ell} j_\ell(kr) \mathrm{Y}_L, && E > 0 \\ &= (ik)^{-\ell} j_\ell(kr) \mathrm{Y}_L, && E < 0 \end{aligned}$$

where

$$h_0 = \mathrm{Re}\, e^{ikr}/r \quad \text{and} \quad j_0 = \sin(kr)/kr \tag{B.13}$$

H_L and J_L are real for any real energy, and structure constants Eq. (1) have the property $S_{KL}(\mathbf{R}) = S^*_{LK}(-\mathbf{R})$. They can generated from the $\ell = 0$ functions using the operator $\mathcal{Y}_L(-\nabla)$; see, e.g. Eq. (116). The latter are useful for derivations and also practically, for instance, to express the gradient of ∇H_L as a linear combination of other H_L. Eq. (122) shows the explicit relation to the usual spherical Hankel function of the first kind, $h^{(1)}_\ell(z)$. Spherical harmonics Y_L can be substituted for the Y_L in Eq. (B.13).

For historical reasons, Andersen made a different set of definitions, which are convenient for the ASA when $E = 0$. In this paper we distinguish them from Questaal's standard definitions Eq. (B.13) with a caret. For $E \leq 0$ and $\kappa^2 = -E$,

$$\begin{aligned} \hat{H}_\ell(\kappa, r) &= -[(2\ell-1)!!(\kappa w)^{-\ell}]^{-1}(i\kappa w) h_\ell(\kappa r) \\ &\xrightarrow{\kappa\to 0} (w/r)^{\ell+1} \\ \hat{J}_\ell(\kappa, r) &= (-1/2)[(2\ell-1)!!(\kappa w)^{-\ell}] j_\ell(\kappa r) \\ &\xrightarrow{\kappa\to 0} [2(2\ell+1)]^{-1}(r/w)^\ell \end{aligned} \tag{B.14}$$

w is a length scale which can be chosen arbitrarily (usually taken to be some average of the MT radii).

Appendix C. Calculations in real and reciprocal space

All of the formalism developed so far (with the exception of APWs) can apply to either real or reciprocal space. At present the Questaal codes work in reciprocal space, with the exception of `lmpg` (Section 2.14), which uses real space for the principal layer direction. The k mesh is always a uniform mesh. Along any axis the user can specify whether the mesh has a point coinciding with $k = 0$ or points that symmetrically straddle it.

When screened MTO's are used, Bloch sums are formed directly from the short ranged orbitals. The standard `lmf` basis is not sufficiently short ranged, and Bloch sums are constructed with an Ewald technique. A Bloch sum of $\mathcal{H}_L$ is written as a sum over lattice vectors $\mathbf{R}$:

$$\begin{aligned} \mathcal{H}^{\mathbf{k}}_L(\varepsilon, \gamma; \mathbf{r}) &= \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} \mathcal{H}_L(\varepsilon, \gamma; \mathbf{r} - \mathbf{R}) \\ &= e^{i\mathbf{k}\cdot\mathbf{r}} \sum_{\mathbf{R}} e^{-i\mathbf{k}\cdot(\mathbf{r}-\mathbf{R})} \mathcal{H}_L(\varepsilon, \gamma; \mathbf{r} - \mathbf{R}) \end{aligned}$$

The latter sum is periodic in $\mathbf{R}$ and can be written as a Fourier series in the reciprocal lattice vectors $\mathbf{G}$

$$\sum_{\mathbf{R}} e^{-i\mathbf{k}\cdot(\mathbf{r}-\mathbf{R})} \mathcal{H}_L(\varepsilon, \gamma; \mathbf{r} - \mathbf{R}) = \frac{1}{V} \sum_{\mathbf{G}} e^{i\mathbf{G}\cdot\mathbf{r}} \widehat{\mathcal{H}}_L(\varepsilon, \gamma; \mathbf{k} + \mathbf{G})$$

so that the Bloch-summed $\mathcal{H}_L$ is

$$\mathcal{H}^{\mathbf{k}}_L(\varepsilon, \gamma; \mathbf{r}) = \frac{1}{V} \sum_{\mathbf{G}} \widehat{\mathcal{H}}_L(\varepsilon, \gamma; \mathbf{k} + \mathbf{G}) e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}} \tag{C.15}$$

V is the volume of the unit cell.

Appendix D. Matrix elements of Smooth Hankel Envelope Functions

This Appendix derives analytic expressions for matrix elements of the overlap, laplacian, gradient and position operators for a pair of functions $\mathcal{H}_{k_2 L_2}(\varepsilon_2, r_{s_2}; \mathbf{r} - \mathbf{R}_2)$ and $\mathcal{H}_{k_1 L_1}(\varepsilon_1, r_{s_1}; \mathbf{r} - \mathbf{R}_1)$, of the form Eq. (117). The strategy will be to use Parseval's identity, Eq. (140), and evaluate the matrix elements of functions in reciprocal space. Analytic solutions can be obtained because products of functions $\widehat{\mathcal{H}}^*_2 \widehat{\mathcal{H}}_1$ can be contracted to sums of functions of the same class, whose Fourier transform is known.

To begin with, consider matrix elements of the overlap and Laplacian operators. Since $\nabla^2 \mathcal{H}_{kL} = \mathcal{H}_{k+1L}$, matrix elements of the Laplacian and overlap are different members of the same class. Using Eqs. (115), (117) and (A.8), products of two $\widehat{\mathcal{H}}_{kL}(\mathbf{q})$ read as follows. To shorten the formulas the smoothing radius will be expressed in terms of $\gamma = r_s^2/4$.

$$\begin{aligned} \widehat{\mathcal{H}}^*_1 \widehat{\mathcal{H}}_2 &= \left[\widehat{h}_0(\varepsilon_1, \gamma_1, q) \widehat{h}_0(\varepsilon_2, \gamma_2, q)\right] \\ &\times \left\{(-q^2)^{k_1+k_2} (-1)^{\ell_1}\, \mathcal{Y}_{L_1}(-i\mathbf{q})\, \mathcal{Y}_{L_2}(-i\mathbf{q})\right\} e^{i\mathbf{q}\cdot\mathbf{R}_1} e^{-i\mathbf{q}\cdot\mathbf{R}_2} \end{aligned} \tag{D.16}$$

The quantity in curly braces can be expanded into a linear combination of $\mathcal{Y}_L$ using Eq. (A.8),

$$\begin{aligned} &(-q^2)^{k_1+k_2} (-1)^{\ell_1}\, \mathcal{Y}_{L_1}(-i\mathbf{q})\, \mathcal{Y}_{L_2}(-i\mathbf{q}) \\ &\quad = (-1)^{\ell_1} \sum_M C_{L_1 L_2 M}\, (-q^2)^{k_1+k_2+(\ell_1+\ell_2-m)/2}\, \mathcal{Y}_M(-i\mathbf{q}) \end{aligned} \tag{D.17}$$

and the radial part in square brackets is similarly expanded

$$\begin{aligned} \widehat{h}_0(\varepsilon_1, \gamma_1, q) \widehat{h}_0(\varepsilon_2, \gamma_2, q) &= \frac{(4\pi)^2 e^{\gamma_1(\varepsilon_1 - q^2)} e^{\gamma_2(\varepsilon_2 - q^2)}}{(\varepsilon_1 - q^2)(\varepsilon_2 - q^2)} \\ &= \frac{(4\pi)^2}{\varepsilon_1 - \varepsilon_2} \left[\frac{e^{\gamma_1(\varepsilon_1-\varepsilon_2)} e^{(\gamma_1+\gamma_2)(\varepsilon_2-q^2)}}{\varepsilon_2 - q^2} - \frac{e^{\gamma_2(\varepsilon_2-\varepsilon_1)} e^{(\gamma_2+\gamma_1)(\varepsilon_1-q^2)}}{\varepsilon_1 - q^2} \right] \\ &= \frac{4\pi}{\varepsilon_1 - \varepsilon_2} \left[e^{\gamma_2(\varepsilon_2-\varepsilon_1)} \widehat{h}_0(\varepsilon_1, \gamma, q) - e^{\gamma_1(\varepsilon_1-\varepsilon_2)} \widehat{h}_0(\varepsilon_2, \gamma, q) \right] \end{aligned} \tag{D.18}$$

$$\xrightarrow{\varepsilon_2, \varepsilon_1 \to \varepsilon} \frac{(4\pi)^2}{(\varepsilon_2 - q^2)^2} e^{(\gamma_1+\gamma_2)(\varepsilon - q^2)} \tag{D.19}$$

$$\gamma = \gamma_1 + \gamma_2. \tag{D.20}$$

When the factors in Eq. (D.16) are combined, the product is seen to be the Fourier transform of a linear combination of $\mathcal{H}_{kM}$ with smoothing radius given by $\gamma_1 + \gamma_2$, at the connecting vector

$\mathbf{R}_1 - \mathbf{R}_2$. The overlap matrix can then be written

$$\begin{aligned}\left\langle \mathcal{H}_{k_1L_1}\mathcal{H}_{k_2L_2}\right\rangle &= \int \mathcal{H}^*_{k_1L_1}(\varepsilon_1,\gamma_1;\mathbf{r}-\mathbf{R}_1)\mathcal{H}_{k_2L_2}(\varepsilon_2,\gamma_2;\mathbf{r}-\mathbf{R}_2)\,d^3r\\ &= 4\pi(-1)^{\ell_1}\sum_M C_{L_1L_2M}\\ &\times\left\{\frac{e^{-\gamma_1(\varepsilon_2-\varepsilon_1)}}{\varepsilon_2-\varepsilon_1}\mathcal{H}_{k'M}(\varepsilon_2,\gamma;\mathbf{R}) - \frac{e^{\gamma_2(\varepsilon_2-\varepsilon_1)}}{\varepsilon_2-\varepsilon_1}\mathcal{H}_{k'M}(\varepsilon_1,\gamma;\mathbf{R})\right\}\\ k' &= k_1+k_2+(\ell_1+\ell_2-m)/2,\quad \gamma=\gamma_1+\gamma_2,\quad \mathbf{R}=\mathbf{R}_1-\mathbf{R}_2\end{aligned} \tag{D.21}$$

The case $\varepsilon_1 = \varepsilon_2 = \varepsilon$ must be handled specially. In that limit, the radial part simplifies to Eq. (D.19). This function is not contained in the $\mathcal{H}_{kL}$ pantheon, and a new function must be added

$$\begin{aligned}\widehat{\mathcal{W}}_{kL}(\mathbf{q}) &= \widehat{w}_0(\varepsilon,r_s;q)(-iq)^{2k}y_L(-i\mathbf{q})\\ \widehat{w}_0(\varepsilon,\gamma;q) &= \frac{4\pi}{(\varepsilon-q^2)^2}e^{\gamma(\varepsilon-q^2)} = \frac{-1}{(\varepsilon-q^2)}\widehat{h}_0(\varepsilon,\gamma;q)\end{aligned} \tag{D.22}$$

$\widehat{\mathcal{W}}_{kL}$ is closely related to the energy derivative of $\widehat{\mathcal{H}}_{kL}$. Differentiate Eq. (117), with respect to energy:

$$\begin{aligned}&\dot{\widehat{\mathcal{H}}}_{kL} \equiv \tfrac{\partial\widehat{\mathcal{H}}_{kL}}{\partial\varepsilon} = \widehat{\mathcal{W}}_{kL} + \gamma\widehat{\mathcal{H}}_{kL}\\ &\text{which are generated from radial functions}\\ &\dot{\widehat{h}}_0(\varepsilon,\gamma;q) = \widehat{w}_0(\varepsilon,\gamma;q) + \gamma\,\widehat{h}_0(\varepsilon,\gamma;q)\end{aligned} \tag{D.23}$$

This provides a practical scheme to obtain $\mathcal{W}_{0L}$ in real space since analytic forms for $\mathcal{H}_{0L}$ and $\dot{\mathcal{H}}_{0L}$ are known (Eqs. (111) and (112)). To make $\mathcal{W}_{kL}$ for higher k it is readily seen that

$$\mathcal{W}_{k+1L}(\mathbf{r}) = -\varepsilon\mathcal{W}_{kL}(\mathbf{r}) - \mathcal{H}_{kL}(\mathbf{r})$$

Using Eq. (D.19) for the quantity in square brackets, Eq. (D.16), the resulting integral is

$$\begin{aligned}&\int \mathcal{H}^*_{k_1L_1}(\varepsilon,\gamma_1;\mathbf{r}-\mathbf{R}_1)\mathcal{H}_{k_2L_2}(\varepsilon,\gamma_2;\mathbf{r}-\mathbf{R}_2)\,d^3r =\\ &\qquad 4\pi(-1)^{\ell_1}\sum_M C_{L_1L_2M}\mathcal{W}_{k'M}(\varepsilon,\gamma;\mathbf{R})\end{aligned} \tag{D.24}$$

where k', γ and $\mathbf{R}$ the same as in Eq. (D.21). Note that Eq. (D.24) encompasses matrix elements between two generalised Gaussians, by virtue of Eq. (133), $H_{k+1L}(\varepsilon=0) = -4\pi G_{kL}$.

D.1. The momentum and position operators acting on a smooth Hankel function

Following the usual rules of Fourier transforms, the gradient operator $\nabla_{\mathbf{r}}$ and position operator $\mathbf{r}$ act in reciprocal space as follows. Use the definition Eqs. (115) and (117) with Eq. (A.11) to obtain

$$\begin{aligned}\nabla_p\,\mathcal{H}_{kL}(\varepsilon,\gamma;\mathbf{r}) &= \frac{1}{(2\pi)^3}\int \widehat{\mathcal{H}}_{kL}(\varepsilon,\gamma;\mathbf{q})\,\nabla e^{i\mathbf{q}\cdot\mathbf{r}}\,d^3q\\ &= -\frac{1}{(2\pi)^3}\int e^{i\mathbf{q}\cdot\mathbf{r}}(-i\mathbf{q})\widehat{h}_0(\varepsilon,\gamma;q)y_L(-i\mathbf{q})\,d^3q\\ &= -\frac{1}{(2\pi)^3}\int d^3q\,e^{i\mathbf{q}\cdot\mathbf{r}}\sum_M\left\{C^{(+)}_{ML;p}\widehat{\mathcal{H}}_{kM}(\mathbf{q}) + C^{(-)}_{ML;p}\mathcal{H}_{k+1M}(\mathbf{q})\right\}\\ &= -\sum_M\left\{C^{(+)}_{ML;p}\mathcal{H}_{kM}(\mathbf{r}) + C^{(-)}_{ML;p}\mathcal{H}_{k+1M}(\mathbf{r})\right\}\end{aligned} \tag{D.25}$$

$C^{(\pm)}_{ML;p}$ has two nonzero elements for fixed L; p. The position operator is more complicated, and we restrict the derivation to $k=0$ case.

$$\begin{aligned}\mathbf{r}\,\mathcal{H}_{0L}(\varepsilon,\gamma;\mathbf{r}) &= \frac{1}{(2\pi)^3}\int \widehat{\mathcal{H}}_{0L}(\varepsilon,\gamma;\mathbf{q})\,\frac{\partial}{i\partial\mathbf{q}}e^{i\mathbf{q}\cdot\mathbf{r}}\,d^3q\\ &= \frac{1}{(2\pi)^3}\int e^{i\mathbf{q}\cdot\mathbf{r}}\,[i\nabla_{\mathbf{q}}\widehat{h}_0(\varepsilon,\gamma;q)]y_L(-i\mathbf{q})\\ &\quad + \widehat{h}_0(\varepsilon,\gamma;q)[i\nabla_{\mathbf{q}}y_L(-i\mathbf{q})]\,d^3q\end{aligned}$$

The first term contains

$$\begin{aligned}[i\nabla_{\mathbf{q}}\widehat{h}_0(\varepsilon,\gamma;q)] &= 2i\mathbf{q}\,(\widehat{w}_0 - \gamma\widehat{h}_0) = 2(-i\mathbf{q})\,\dot{\widehat{h}}_0\\ [i\nabla_{\mathbf{q}}\widehat{h}_0(\varepsilon,\gamma;q)]y_L(-i\mathbf{q}) &= 2\sum_M\left\{C^{(+)}_{ML;p}\dot{\widehat{\mathcal{H}}}_{0M}(\mathbf{q}) + C^{(-)}_{ML;p}\widehat{\mathcal{H}}_{1M}(\mathbf{q})\right\}\end{aligned}$$

using Eq. (A.11). The second term follows directly from Eq. (A.10) to yield

$$\begin{aligned}\mathbf{r}_p\,\mathcal{H}_{0L}(\varepsilon,\gamma;\mathbf{r}) = \sum_M\Big\{\, &C^{(+)}_{ML;p}(2\dot{\mathcal{H}}_{0M}(\mathbf{r})) + C^{(-)}_{ML;p}(2\dot{\mathcal{H}}_{1M}(\mathbf{r})\\ &+(2\ell+1)\mathcal{H}_{0M}(\mathbf{r}))\,\Big\}\end{aligned} \tag{D.26}$$

D.2. Matrix elements of position and gradient operators

Eq. (D.25) shows that $\nabla\mathcal{H}_{kL}(\mathbf{r})$ generates a linear combination of the same family of $\mathcal{H}_{kL}(\mathbf{r})$, evaluated at the vector connecting atom centres. Thus, matrix elements of the gradient operator are linear combinations of the same functions. A similar situation applies to matrix elements of $\mathbf{r}$, though they entail matrix elements of the overlap between $\mathcal{H}_{kL}$ and $\dot{\mathcal{H}}_{kL}$. These integrals can be evaluated in reciprocal space using Parseval's identity, expanding $\widehat{\mathcal{H}}^*_1\dot{\widehat{\mathcal{H}}}_2$ analogously to Eq. (D.16). The radial part of this product can be expanded as:

$$\begin{aligned}\widehat{h}_0(\varepsilon_1,\gamma_1,q)\dot{\widehat{h}}_0(\varepsilon_2,\gamma_2,q) &= \frac{-(4\pi)^2e^{\gamma_1(\varepsilon_1-q^2)}e^{\gamma_2(\varepsilon_2-q^2)}}{(\varepsilon_1-q^2)(\varepsilon_2-q^2)^2}\\ &\quad +\gamma_2\,\widehat{h}_0(\varepsilon_1,\gamma_1,q)\widehat{h}_0(\varepsilon_2,\gamma_2,q)\\ &= \frac{-4\pi}{\varepsilon_1-\varepsilon_2}\Big\{\, e^{(\varepsilon_1-\varepsilon_2)\gamma_1}\widehat{w}_0(\varepsilon_2,\gamma,q) + [\gamma_2+(\varepsilon_1-\varepsilon_2)^{-1}]\\ &\times e^{-\varepsilon_2\gamma_1-\varepsilon_1\gamma_2}\left(e^{\varepsilon_1\gamma}\,\widehat{h}_0(\varepsilon_2,\gamma,q) - e^{\varepsilon_2\gamma}\,\widehat{h}_0(\varepsilon_1,\gamma,q)\right)\Big\}\end{aligned} \tag{D.27}$$

Substituting this radial function into the square brackets in Eq. (D.19), we obtain

$$\begin{aligned}\left\langle \mathcal{H}_{k_1L_1}\dot{\mathcal{H}}_{k_2L_2}\right\rangle &= \int \mathcal{H}^*_{k_1L_1}(\varepsilon_1,\gamma_1;\mathbf{r}-\mathbf{R}_1)\dot{\mathcal{H}}_{k_2L_2}(\varepsilon_2,\gamma_2;\mathbf{r}-\mathbf{R}_2)\,d^3r\\ &= \frac{4\pi(-1)^{\ell_1}}{\varepsilon_1-\varepsilon_2}\sum_M C_{L_1L_2M}\times\Big\{\, e^{(\varepsilon_1-\varepsilon_2)\gamma_1}\mathcal{W}_{k'M}(\varepsilon_2) +\\ &\quad [\gamma_2+(\varepsilon_1-\varepsilon_2)^{-1}]\,e^{-\varepsilon_2\gamma_1-\varepsilon_1\gamma_2}\\ &\quad\times\left(e^{\varepsilon_1\gamma}\,\mathcal{H}_{k'M}(\varepsilon_2) - e^{\varepsilon_2\gamma}\,\mathcal{H}_{k'M}(\varepsilon_1)\right)\Big\}\end{aligned} \tag{D.28}$$

where $\mathcal{W}_{k'M}$ and $\mathcal{H}_{k'M}$ are evaluated with the same arguments k', γ and $\mathbf{R}$ as in Eq. (D.21). When $\varepsilon_2, \varepsilon_1 \to \varepsilon$, the expression in curly brackets reduces to

$$[\dot{\mathcal{W}}_{k'M}(\varepsilon) + (\gamma_2-\gamma_1)\mathcal{W}_{k'M}(\varepsilon)]/2$$

Combining Eqs. (D.25), (D.26), (D.21) and (D.28)

$$\begin{aligned}&\left\langle \mathcal{H}_{k_1L_1}\nabla_p\mathcal{H}_{k_2L_2}\right\rangle\\ &\quad= \int \mathcal{H}^*_{k_1L_1}(\varepsilon_1,\gamma_1;\mathbf{r}-\mathbf{R}_1)\nabla_p\mathcal{H}_{k_2L_2}(\varepsilon_2,\gamma_2;\mathbf{r}-\mathbf{R}_2)\,d^3r\\ &\quad= -\sum_M\left\{C^{(+)}_{ML_2;p}\left\langle \mathcal{H}_{k_1L_1}\mathcal{H}_{k_2M}\right\rangle + C^{(-)}_{ML_2;p}\left\langle \mathcal{H}_{k_1L_1}\mathcal{H}_{k_2+1M}\right\rangle\right\}\end{aligned} \tag{D.29}$$

$$\begin{aligned}\left\langle \mathcal{H}_{0L_1}(\mathbf{r}-\mathbf{R}_2)_p \mathcal{H}_{0L_2}\right\rangle &= \int \mathcal{H}^*_{0L_1}(\varepsilon_1,\gamma_1;\mathbf{r}-\mathbf{R}_1)(\mathbf{r}-\mathbf{R}_2)_p \\ &\times \mathcal{H}_{0L_2}(\varepsilon_2,\gamma_2;\mathbf{r}-\mathbf{R}_2)d^3r \\ &= \sum_M \Big\{ C^{(+)}_{ML_2;p} 2\left\langle \mathcal{H}_{0L_1}\dot{\mathcal{H}}_{0M}\right\rangle + C^{(-)}_{ML_2;p}\left(2\left\langle \mathcal{H}_{0L_1}\dot{\mathcal{H}}_{1M}\right\rangle \right. \\ &+ (2\ell_2+1)\left\langle \mathcal{H}_{0L_1}\mathcal{H}_{0M}\right\rangle\Big)\Big\} \end{aligned} \tag{D.30}$$

References

[1] C. Herring, Phys. Rev. 57 (1940) 1169–1177, http://dx.doi.org/10.1103/PhysRev.57.1169.
[2] J.C. Slater, Phys. Rev. 51 (1937) 846–851, http://dx.doi.org/10.1103/PhysRev.51.846.
[3] J. Korringa, Phys. Rep. 238 (1994) 341–360, http://dx.doi.org/10.1016/0370-1573(94)90122-8.
[4] P.E. Blöchl, Phys. Rev. B 50 (1994) 17953–17979, http://dx.doi.org/10.1103/PhysRevB.50.17953.
[5] O.K. Andersen, Phys. Rev. B 12 (1975) 3060–3083, http://dx.doi.org/10.1103/PhysRevB.12.3060.
[6] K. Lejaeghere, G. Bihlmayer, T. Bjorkman, P. Blaha, S. Blugel, V. Blum, D. Caliste, I.E. Castelli, S.J. Clark, A. Dal Corso, S. de Gironcoli, T. Deutsch, J.K. Dewhurst, I. Di Marco, C. Draxl, M. Duak, O. Eriksson, J.A. Flores-Livas, K.F. Garrity, L. Genovese, P. Giannozzi, M. Giantomassi, S. Goedecker, X. Gonze, O. Granas, E.K.U. Gross, A. Gulans, F. Gygi, D.R. Hamann, P.J. Hasnip, N.A.W. Holzwarth, D. Iuan, D.B. Jochym, F. Jollet, D. Jones, G. Kresse, K. Koepernik, E. Kucukbenli, Y.O. Kvashnin, I.L.M. Locht, S. Lubeck, M. Marsman, N. Marzari, U. Nitzsche, L. Nordstrom, T. Ozaki, L. Paulatto, C.J. Pickard, W. Poelmans, M.I.J. Probert, K. Refson, M. Richter, G.-M. Rignanese, S. Saha, M. Scheffler, M. Schlipf, K. Schwarz, S. Sharma, F. Tavazza, P. Thunstrom, A. Tkatchenko, M. Torrent, D. Vanderbilt, M.J. van Setten, V. Van Speybroeck, J.M. Wills, J.R. Yates, G.-X. Zhang, S. Cottenier, Science 351 (2016) aad3000, http://dx.doi.org/10.1126/science.aad3000.
[7] M. Methfessel, D. Hennig, M. Scheffler, Phys. Rev. B 46 (1992) 4816–4829, http://dx.doi.org/10.1103/PhysRevB.46.4816.
[8] A.I. Liechtenstein, I.I. Mazin, C.O. Rodriguez, O. Jepsen, O.K. Andersen, M. Methfessel, Phys. Rev. B 44 (1991) 5388–5391, http://dx.doi.org/10.1103/PhysRevB.44.5388.
[9] M. Methfessel, M. van Schilfgaarde, M. Scheffler, Phys. Rev. Lett. 70 (1993) 29–32, http://dx.doi.org/10.1103/PhysRevLett.70.29.
[10] M.I. Katsnelson, V.P. Antropov, M. van Schilfgaarde, B.N. Harmon, JETP Lett. 62 (1995) 439–443.
[11] V.P. Antropov, M.I. Katsnelson, M. van Schilfgaarde, B.N. Harmon, Phys. Rev. Lett. 75 (1995) 729–732, http://dx.doi.org/10.1103/PhysRevLett.75.729.
[12] T. Kotani, H. Akai, Phys. Rev. B 54 (1996) 16502–16514, http://dx.doi.org/10.1103/PhysRevB.54.16502.
[13] F. Aryasetiawan, O. Gunnarsson, Phys. Rev. Lett. 74 (1995) 3221–3224, http://dx.doi.org/10.1103/PhysRevLett.74.3221.
[14] S.V. Faleev, F. Léonard, D.A. Stewart, M. van Schilfgaarde, Phys. Rev. B 71 (2005) 195422, http://dx.doi.org/10.1103/PhysRevB.71.195422.
[15] T. Kotani, M. van Schilfgaarde, Solid State Commun. 121 (2002) 461–465, http://dx.doi.org/10.1016/S0038-1098(02)00028-5.
[16] S.V. Faleev, M. van Schilfgaarde, T. Kotani, Phys. Rev. Lett. 93 (2004) 126406, http://dx.doi.org/10.1103/PhysRevLett.93.126406.
[17] S. Choi, A. Kutepov, K. Haule, M. van Schilfgaarde, G. Kotliar, npj Quantum Mater. 1 (2016) 16001, http://dx.doi.org/10.1038/npjquantmats.2016.1.
[18] O.K. Andersen, O. Jepsen, Phys. Rev. Lett. 53 (1984) 2571–2574, http://dx.doi.org/10.1103/PhysRevLett.53.2571.
[19] F. Caruso, P. Rinke, X. Ren, A. Rubio, M. Scheffler, Phys. Rev. B 88 (2013) 075105, http://dx.doi.org/10.1103/PhysRevB.88.075105.
[20] M.J. van Setten, F. Weigend, F. Evers, J. Chem. Theory Comput. 9 (2013) 232–246, http://dx.doi.org/10.1021/ct300648t.
[21] M.A. Marques, M.J. Oliveira, T. Burnus, Comput. Phys. Comm. 183 (2012) 2272–2281, http://dx.doi.org/10.1016/j.cpc.2012.05.007.
[22] A.N. Chantis, K.D. Belashchenko, E.Y. Tsymbal, M. van Schilfgaarde, Phys. Rev. Lett. 98 (2007) 046601, http://dx.doi.org/10.1103/PhysRevLett.98.046601.
[23] A.N. Chantis, K.D. Belashchenko, D.L. Smith, E.Y. Tsymbal, M. van Schilfgaarde, R.C. Albers, Phys. Rev. Lett. 99 (2007) 196603, http://dx.doi.org/10.1103/PhysRevLett.99.196603.
[24] K.D. Belashchenko, A.A. Kovalev, M. van Schilfgaarde, Phys. Rev. Lett. 117 (2016) 207204, http://dx.doi.org/10.1103/PhysRevLett.117.207204.
[25] K.D. Belashchenko, A.A. Kovalev, M. van Schilfgaarde, Phys. Rev. Mater. 3 (2019) 011401, http://dx.doi.org/10.1103/PhysRevMaterials.3.011401.
[26] F. Aryasetiawan, Phys. Rev. B 46 (1992) 13051–13064, http://dx.doi.org/10.1103/PhysRevB.46.13051.
[27] K. Haule, Phys. Rev. B 75 (2007) 155113, http://dx.doi.org/10.1103/PhysRevB.75.155113.
[28] O. Parcollet, M. Ferrero, T. Ayral, H. Hafermann, I. Krivenko, L. Messio, P. Seth, Comput. Phys. Comm. 196 (2015) 398–415, http://dx.doi.org/10.1016/j.cpc.2015.04.023.
[29] M.W. Finnis, A.T. Paxton, M. Methfessel, M. van Schilfgaarde, Phys. Rev. Lett. 81 (1998) 5149–5152, http://dx.doi.org/10.1103/PhysRevLett.81.5149.
[30] M.W. Finnis, Interatomic Forces in Condensed Matter, Oxford University Press, Oxford, 2003.
[31] M. Methfessel, M. van Schilfgaarde, R.A. Casali, in: H. Dreyssé (Ed.), Electronic Structure and Physical Properties of Solids: The Uses of the LMTO Method, in: Lecture Notes in Physics, vol. 535, Springer, Berlin, 2000, p. 114.
[32] O. Gunnarsson, O. Jepsen, O.K. Andersen, Phys. Rev. B 27 (1983) 7144–7168, http://dx.doi.org/10.1103/PhysRevB.27.7144.
[33] O.K. Andersen, T. Saha-Dasgupta, R.W. Tank, G.K.C. Arcangeli, O. Jepsen, G. Krier, in: H. Dreyssé (Ed.), Electronic Structure and Physical Properties of Solids: The Uses of the LMTO Method, in: Lecture Notes in Physics, vol. 535, Springer, Berlin, 2000, p. 3.
[34] O.K. Andersen, A.V. Postnikov, S.Y. Savrasov, in: W.H. Butler, P.H. Dederichs, A. Gonis, R.L. Weaver (Eds.), Applications of Multiple Scattering Theory to Materials Science, Vol. 253, Materials Research Society, Pittsburgh, 1992, pp. 37–70.
[35] O.K. Andersen, Z. Pawlowska, O. Jepsen, Phys. Rev. B 34 (1986) 5253–5269, http://dx.doi.org/10.1103/PhysRevB.34.5253.
[36] M. van Schilfgaarde, W.R.L. Lambrecht, in: L. Colombo, A. Gonis, P. Turchi (Eds.), Tight-Binding Approach to Computational Materials Science, in: Lecture Notes in Physics, vol. 491, Springer, Berlin, 1998, p. 137.
[37] M.P.L. Sancho, J.M.L. Sancho, J. Rubio, J. Phys. F 15 (1985) 851–858, http://dx.doi.org/10.1088/0305-4608/15/4/009.
[38] A.-B. Chen, Y.-M. Lai-Hsu, W. Chen, Phys. Rev. B 39 (1989) 923–929, http://dx.doi.org/10.1103/PhysRevB.39.923.
[39] A.K. Rajagopal, J. Callaway, Phys. Rev. B 7 (1973) 1912–1919, http://dx.doi.org/10.1103/PhysRevB.7.1912.
[40] G. Vignale, M. Rasolt, Phys. Rev. Lett. 59 (1987) 2360–2363, http://dx.doi.org/10.1103/PhysRevLett.59.2360.
[41] G. Vignale, M. Rasolt, Phys. Rev. B 37 (1988) 10685–10696, http://dx.doi.org/10.1103/PhysRevB.37.10685.
[42] H. Ebert, J. Phys. Condens. Matter 1 (1989) 9111–9116, http://dx.doi.org/10.1088/0953-8984/1/46/005.
[43] I. Turek, V. Drchal, J. Kudrnovský, M. Šob, P. Weinberger, Electronic Structure of Disordered Alloys, Surfaces and Interfaces, Springer, Boston, 1997.
[44] R. Feder, F. Rosicky, B. Ackermann, Z. Phys. B 52 (1983) 31–36, http://dx.doi.org/10.1007/BF01305895.
[45] J. Kudrnovský, V. Drchal, Phys. Rev. B 41 (1990) 7515–7528, http://dx.doi.org/10.1103/PhysRevB.41.7515.
[46] A. Devonport, A. Vishina, R. Singh, M. Edwards, K. Zheng, J. Domenico, N. Rizzo, C. Kopas, M. van Schilfgaarde, N. Newman, J. Magn. Magn. Mater. 460 (2018) 193–202, http://dx.doi.org/10.1016/j.jmmm.2018.03.054.
[47] V. Alena, (Ph.D. thesis), Kings College London, 2019.
[48] M. Abd El Qader, R.K. Singh, S.N. Galvin, L. Yu, J.M. Rowell, N. Newman, Appl. Phys. Lett. 104 (2014) 022602, http://dx.doi.org/10.1063/1.4862195.
[49] V.P. Antropov, M.I. Katsnelson, B.N. Harmon, M. van Schilfgaarde, D. Kusnezov, Phys. Rev. B 54 (1996) 1019–1035, http://dx.doi.org/10.1103/PhysRevB.54.1019.
[50] M. van Schilfgaarde, I.A. Abrikosov, B. Johansson, Nature 400 (1999) 46–49, http://dx.doi.org/10.1038/21848.
[51] A. Bulgac, D. Kusnezov, Phys. Rev. A 42 (1990) 5045–5048, http://dx.doi.org/10.1103/PhysRevA.42.5045.
[52] T. Kotani, H. Kino, H. Akai, J. Phys. Soc. Japan 84 (2015) 034702, http://dx.doi.org/10.7566/JPSJ.84.034702.
[53] E. Bott, M. Methfessel, W. Krabs, P.C. Schmidt, J. Math. Phys. 39 (1998) 3393–3425, http://dx.doi.org/10.1063/1.532437.
[54] T. Kotani, M. van Schilfgaarde, Phys. Rev. B 81 (2010) 125117, http://dx.doi.org/10.1103/PhysRevB.81.125117.
[55] J.M. Soler, A.R. Williams, Phys. Rev. B 40 (1989) 1560–1564, http://dx.doi.org/10.1103/PhysRevB.40.1560.
[56] M. van Schilfgaarde, T. Kotani, S.V. Faleev, Phys. Rev. B 74 (2006) 245125, http://dx.doi.org/10.1103/PhysRevB.74.245125.
[57] D.J. Singh, L. Nordström, Planewaves, Pseudopotentials and the LAPW Method, second ed., Springer, Berlin, 2006.
[58] J. Kuneš, P. Novák, R. Schmid, P. Blaha, K. Schwarz, Phys. Rev. B 64 (2001) 153102, http://dx.doi.org/10.1103/PhysRevB.64.153102.
[59] C. Friedrich, M.C. Müller, S. Blügel, Phys. Rev. B 83 (2011) 081101, http://dx.doi.org/10.1103/PhysRevB.83.081101.
[60] H. Jiang, P. Blaha, Phys. Rev. B 93 (2016) 115203, http://dx.doi.org/10.1103/PhysRevB.93.115203.
[61] J. Harris, Phys. Rev. B 31 (1985) 1770–1779, http://dx.doi.org/10.1103/PhysRevB.31.1770.

[62] W.M.C. Foulkes, R. Haydock, Phys. Rev. B 39 (1989) 12520–12536, http://dx.doi.org/10.1103/PhysRevB.39.12520.
[63] V.I. Anisimov, J. Zaanen, O.K. Andersen, Phys. Rev. B 44 (1991) 943–954, http://dx.doi.org/10.1103/PhysRevB.44.943.
[64] V.I. Anisimov, I.V. Solovyev, M.A. Korotin, M.T. Czyzyk, G.A. Sawatzky, Phys. Rev. B 48 (1993) 16929–16934, http://dx.doi.org/10.1103/PhysRevB.48.16929.
[65] V.I. Anisimov, F. Aryasetiawan, A.I. Lichtenstein, J. Phys. Condens. Matter 9 (1997) 767–808, http://dx.doi.org/10.1088/0953-8984/9/4/002.
[66] A.I. Liechtenstein, V.I. Anisimov, J. Zaanen, Phys. Rev. B 52 (1995) R5467–R5470, http://dx.doi.org/10.1103/PhysRevB.52.R5467.
[67] M.T. Czyzyk, G.A. Sawatzky, Phys. Rev. B 49 (1994) 14211–14228, http://dx.doi.org/10.1103/PhysRevB.49.14211.
[68] A.G. Petukhov, I.I. Mazin, L. Chioncel, A.I. Lichtenstein, Phys. Rev. B 67 (2003) 153106, http://dx.doi.org/10.1103/PhysRevB.67.153106.
[69] V.I. Anisimov, O. Gunnarsson, Phys. Rev. B 43 (1991) 7570–7574, http://dx.doi.org/10.1103/PhysRevB.43.7570.
[70] M. Cococcioni, S. de Gironcoli, Phys. Rev. B 71 (2005) 035105, http://dx.doi.org/10.1103/PhysRevB.71.035105.
[71] E.U. Condon, G.H. Shortley, The Theory of Atomic Spectra, Cambridge University Press, Cambridge, 1953.
[72] Note that Eq.(2) in Ref [73] incorrectly uses 50/429 (taken from Ref. [65]) instead of 1/858 as the contribution of $F^6(ff)$ to the average J, thus the actual J used were about 25% smaller than the ones listed. However, since this did not correspond to the correct average J, it makes an error in the double counting term which shifts the f-bands somewhat relative to the other bands. Also the calculation used the Hartree–Fock unscreened atomic values of the Slater integrals F^k for $k \geq 2$..
[73] P. Larson, W.R.L. Lambrecht, A. Chantis, M. van Schilfgaarde, Phys. Rev. B 75 (2007) 045114, http://dx.doi.org/10.1103/PhysRevB.75.045114.
[74] S. Keshavarz, J. Schött, A.J. Millis, Y.O. Kvashnin, Phys. Rev. B 97 (2018) 184404, http://dx.doi.org/10.1103/PhysRevB.97.184404.
[75] S.L. Dudarev, G.A. Botton, S.Y. Savrasov, C.J. Humphreys, A.P. Sutton, Phys. Rev. B 57 (1998) 1505–1509, http://dx.doi.org/10.1103/PhysRevB.57.1505.
[76] A. Boonchun, W.R.L. Lambrecht, Phys. Status Solidi b 248 (2011) 1043–1051, http://dx.doi.org/10.1002/pssb.201046328.
[77] F. Brivio, K.T. Butler, A. Walsh, M. van Schilfgaarde, Phys. Rev. B 89 (2014) 155204, http://dx.doi.org/10.1103/PhysRevB.89.155204.
[78] A. Svane, N.E. Christensen, M. Cardona, A.N. Chantis, M. van Schilfgaarde, T. Kotani, Phys. Rev. B 81 (2010) 245120, http://dx.doi.org/10.1103/PhysRevB.81.245120.
[79] T. Kotani, J. Phys. Soc. Japan 83 (2014) 094711, http://dx.doi.org/10.7566/JPSJ.83.094711.
[80] N.E. Zein, S.Y. Savrasov, G. Kotliar, Phys. Rev. Lett. 96 (2006) 226403, http://dx.doi.org/10.1103/PhysRevLett.96.226403.
[81] K. Schwarz, P. Blaha, G. Madsen, Comput. Phys. Comm. 147 (2002) 71–76, http://dx.doi.org/10.1016/S0010-4655(02)00206-0.
[82] L. Hedin, Phys. Rev. 139 (1965) A796–A823, http://dx.doi.org/10.1103/PhysRev.139.A796.
[83] M. van Schilfgaarde, T. Kotani, S. Faleev, Phys. Rev. Lett. 96 (2006) 226402, http://dx.doi.org/10.1103/PhysRevLett.96.226402.
[84] T. Kotani, M. van Schilfgaarde, S.V. Faleev, Phys. Rev. B 76 (2007) 165106, http://dx.doi.org/10.1103/PhysRevB.76.165106.
[85] F. Aryasetiawan, O. Gunnarsson, Phys. Rev. B 49 (1994) 16214–16222, http://dx.doi.org/10.1103/PhysRevB.49.16214.
[86] S. Acharya, C. Weber, E. Plekhanov, D. Pashov, A. Taraphder, M. Van Schilfgaarde, Phys. Rev. X 8 (2018) 021038, http://dx.doi.org/10.1103/PhysRevX.8.021038.
[87] E.L. Shirley, Phys. Rev. B 54 (1996) 7758–7764, http://dx.doi.org/10.1103/PhysRevB.54.7758.
[88] B. Holm, U. von Barth, Phys. Rev. B 57 (1998) 2108–2117, http://dx.doi.org/10.1103/PhysRevB.57.2108.
[89] M. Grumet, P. Liu, M. Kaltak, J. Klimeš, G. Kresse, Phys. Rev. B 98 (2018) 155143, http://dx.doi.org/10.1103/PhysRevB.98.155143.
[90] K.D. Belashchenko, V.P. Antropov, N.E. Zein, Phys. Rev. B 73 (2006) 073105, http://dx.doi.org/10.1103/PhysRevB.73.073105.
[91] D. Tamme, R. Schepe, K. Henneberger, Phys. Rev. Lett. 83 (1999) 241, http://dx.doi.org/10.1103/PhysRevLett.83.241.
[92] F. Aryasetiawan, O. Gunnarsson, Phys. Rev. Lett. 74 (1995) 3221–3224, http://dx.doi.org/10.1103/PhysRevLett.74.3221.
[93] M. Cazzaniga, H. Cercellier, M. Holzmann, C. Monney, P. Aebi, G. Onida, V. Olevano, Phys. Rev. B 85 (2012) 195111, http://dx.doi.org/10.1103/PhysRevB.85.195111.
[94] R. Sakuma, P. Werner, F. Aryasetiawan, Phys. Rev. B 88 (2013) 235110, http://dx.doi.org/10.1103/PhysRevB.88.235110.
[95] P. Liu, B. Kim, X.-Q. Chen, D.D. Sarma, G. Kresse, C. Franchini, Phys. Rev. Mater 2 (2018) 075003, http://dx.doi.org/10.1103/PhysRevMaterials.2.075003.
[96] F. Bruneval, T. Rangel, S.M. Hamed, M. Shao, C. Yang, J.B. Neaton, Comput. Phys. Comm. 208 (2016) 149–161, http://dx.doi.org/10.1016/j.cpc.2016.06.019.
[97] T. Olsen, K.S. Thygesen, Phys. Rev. B 86 (2012) 081103, http://dx.doi.org/10.1103/PhysRevB.86.081103.
[98] F. Bruneval, J. Chem. Phys. 136 (2012) 194107, http://dx.doi.org/10.1063/1.4718428.
[99] P. Koval, D. Foerster, D. Sánchez-Portal, Phys. Rev. B 89 (2014) 155417, http://dx.doi.org/10.1103/PhysRevB.89.155417.
[100] N. Salas-Illanes, (Ph.D. thesis), Humboldt-Universität zu Berlin, 2019.
[101] S. Ismail-Beigi, J. Phys. Condens. Matter 29 (2017) 385501, http://dx.doi.org/10.1088/1361-648X/aa7803.
[102] A.L. Kutepov, Phys. Rev. B 95 (2017) 195120, http://dx.doi.org/10.1103/PhysRevB.95.195120.
[103] I. Aguilera, C. Friedrich, S. Blügel, Phys. Rev. B 91 (2015) 125129, http://dx.doi.org/10.1103/PhysRevB.91.125129.
[104] R. Shaltaf, G.-M. Rignanese, X. Gonze, F. Giustino, A. Pasquarello, Phys. Rev. Lett. 100 (2008) 186401, http://dx.doi.org/10.1103/PhysRevLett.100.186401.
[105] M. Shishkin, M. Marsman, G. Kresse, Phys. Rev. Lett. 99 (2007) 246403, http://dx.doi.org/10.1103/PhysRevLett.99.246403.
[106] M. Betzinger, C. Friedrich, A. Görling, S. Blügel, Phys. Rev. B 92 (2015) 245101, http://dx.doi.org/10.1103/PhysRevB.92.245101.
[107] A.N. Chantis, M. van Schilfgaarde, T. Kotani, Phys. Rev. Lett. 96 (2006) 086405, http://dx.doi.org/10.1103/PhysRevLett.96.086405.
[108] A.N. Chantis, M. Cardona, N.E. Christensen, D.L. Smith, M. van Schilfgaarde, T. Kotani, A. Svane, R.C. Albers, Phys. Rev. B 78 (2008) 075208, http://dx.doi.org/10.1103/PhysRevB.78.075208.
[109] J.-W. Luo, A.N. Chantis, M. van Schilfgaarde, G. Bester, A. Zunger, Phys. Rev. Lett. 104 (2010) 066405, http://dx.doi.org/10.1103/PhysRevLett.104.066405.
[110] T. Kotani, M. van Schilfgaarde, J. Phys. Condens. Matter. 20 (2008) 295214, http://dx.doi.org/10.1088/0953-8984/20/29/295214.
[111] L. Ke, M. van Schilfgaarde, J. Pulikkotil, T. Kotani, V. Antropov, Phys. Rev. B 83 (2011) 060404, http://dx.doi.org/10.1103/PhysRevB.83.060404.
[112] S.V. Faleev, O.N. Mryasov, M. van Schilfgaarde, Phys. Rev. B 85 (2012) 174433, http://dx.doi.org/10.1103/PhysRevB.85.174433.
[113] T. Kotani, M. van Schilfgaarde, Phys. Rev. B 81 (2010) 125201, http://dx.doi.org/10.1103/PhysRevB.81.125201.
[114] N.E. Christensen, A. Svane, R. Laskowski, B. Palanivel, P. Modak, A.N. Chantis, M. van Schilfgaarde, T. Kotani, Phys. Rev. B 81 (2010) 045203, http://dx.doi.org/10.1103/PhysRevB.81.045203.
[115] L. Sponza, P. Pisanti, A. Vishina, D. Pashov, C. Weber, M. van Schilfgaarde, S. Acharya, J. Vidal, G. Kotliar, Phys. Rev. B 95 (2017) 041112, http://dx.doi.org/10.1103/PhysRevB.95.041112.
[116] B. Cunningham, M. Grüning, P. Azarhoosh, D. Pashov, M. van Schilfgaarde, Phys. Rev. Mater. 2 (2018) 034603, http://dx.doi.org/10.1103/PhysRevMaterials.2.034603.
[117] S. Chambers, T. Droubay, T. Kaspar, M. Gutowski, M. van Schilfgaarde, Surf. Sci. 554 (2004) 81–89, http://dx.doi.org/10.1016/j.susc.2004.02.021.
[118] M. Gatti, F. Bruneval, V. Olevano, L. Reining, Phys. Rev. Lett. 99 (2007) 266402, http://dx.doi.org/10.1103/PhysRevLett.99.266402.
[119] C. Weber, S. Acharya, B. Cunningham, M. Grüning, L. Zhang, H. Zhao, Y. Tan, Y. Zhang, C. Zhang, K. Liu, M.V. Schilfgaarde, M. Shalaby, arXiv:1901.08139 [link]. URL https://arxiv.org/abs/1901.08139v1.
[120] Y. Tan, H. Zhao, L.Z. Yan, L. Zhang, Y. Zhang, C. Weber, S. Acharya, B. Cunningham, M. Gruning, K. Liu, M.V. Schilfgaarde, M. Shalaby, 2019 44th International Conference on Infrared, Millimeter, and Terahertz Waves (IRMMW-THz), Vol. 1, 2019, http://dx.doi.org/10.1109/IRMMW-THz.2019.8874246.
[121] M. van Schilfgaarde, M.I. Katsnelson, Phys. Rev. B 83 (2011) 081409, http://dx.doi.org/10.1103/PhysRevB.83.081409.
[122] W. Jauch, M. Reehuis, Phys. Rev. B 65 (2002) 125111, http://dx.doi.org/10.1103/PhysRevB.65.125111.
[123] W. Jauch, M. Reehuis, Phys. Rev. B 70 (2004) 195121, http://dx.doi.org/10.1103/PhysRevB.70.195121.
[124] M. Heiblum, M.V. Fischetti, W.P. Dumke, D.J. Frank, I.M. Anderson, C.M. Knoedler, L. Osterling, Phys. Rev. Lett. 58 (1987) 816–819, http://dx.doi.org/10.1103/PhysRevLett.58.816.
[125] T. Ruf, M. Cardona, Phys. Rev. B 41 (1990) 10747–10753, http://dx.doi.org/10.1103/PhysRevB.41.10747.
[126] M. Cardona, C.M. Penchina, N. Shevchik, J. Tejeda, Solid State Commun. 11 (1972) 1655–1658, http://dx.doi.org/10.1016/0038-1098(72)90764-8.
[127] S. Albrecht, L. Reining, R. Del Sole, G. Onida, Phys. Rev. Lett. 80 (1998) 4510–4513, http://dx.doi.org/10.1103/PhysRevLett.80.4510.
[128] M. Rohlfing, S.G. Louie, Phys. Rev. Lett. 81 (1998) 2312–2315, http://dx.doi.org/10.1103/PhysRevLett.81.2312.
[129] Cunningham, B. and Grüning, M. and Pashov, D. and van Schilfgaarde, M..
[130] H. Fröhlich, Adv. Phys. 3 (1954) 325–361, http://dx.doi.org/10.1080/00018735400101213.
[131] W.R.L. Lambrecht, C. Bhandari, M. van Schilfgaarde, Phys. Rev. Mater. 1 (2017) 043802, http://dx.doi.org/10.1103/PhysRevMaterials.1.043802.
[132] M. Cardona, M.L.W. Thewalt, Rev. Modern Phys. 77 (2005) 1173–1224, http://dx.doi.org/10.1103/RevModPhys.77.1173.

[133] C. Bhandari, M. van Schilgaarde, T. Kotani, W.R.L. Lambrecht, Phys. Rev. Mater. 2 (2018) 013807, http://dx.doi.org/10.1103/PhysRevMaterials.2.013807.

[134] A. Grüneis, G. Kresse, Y. Hinuma, F. Oba, Phys. Rev. Lett. 112 (2014) 096401, http://dx.doi.org/10.1103/PhysRevLett.112.096401.

[135] A.N. Chantis, M. van Schilfgaarde, T. Kotani, Phys. Rev. B 76 (2007) 165126, http://dx.doi.org/10.1103/PhysRevB.76.165126.

[136] T. Moriya, Spin Fluctuations in Itinerant Electron Magnetism, Springer, Berlin, 1985.

[137] A. Aguayo, I. Mazin, D. Singh, Phys. Rev. Lett. 92 (2004) 147201, http://dx.doi.org/10.1103/PhysRevLett.92.147201.

[138] Y. Nishihara, M. Tokumoto, K. Murata, H. Unoki, Japan. J. Appl. Phys. 26 (1987) L1416–L1418, http://dx.doi.org/10.1143/JJAP.26.L1416.

[139] S. Acharya, D. Pashov, F. Jamet, M. van Schilfgaarde, arXiv:1908.08136v1. URL https://arxiv.org/abs/1908.08136v1.

[140] A. Zunger, S.-H. Wei, L.G. Ferreira, J.E. Bernard, Phys. Rev. Lett. 65 (1990) 353–356, http://dx.doi.org/10.1103/PhysRevLett.65.353.

[141] F. Aryasetiawan, K. Karlsson, Phys. Rev. B 60 (1999) 7419–7428, http://dx.doi.org/10.1103/PhysRevB.60.7419.

[142] E. Şaşioğlu, A. Schindlmayr, C. Friedrich, F. Freimuth, S. Blügel, Phys. Rev. B 81 (2010) 054434, http://dx.doi.org/10.1103/PhysRevB.81.054434.

[143] W.F. Brinkman, T.M. Rice, Phys. Rev. B 2 (1970) 4302–4304, http://dx.doi.org/10.1103/PhysRevB.2.4302.

[144] M.C. Gutzwiller, Phys. Rev. 137 (1965) A1726–A1735, http://dx.doi.org/10.1103/PhysRev.137.A1726.

[145] S.E. Barnes, J. Phys. F 7 (1977) 2637–2647, http://dx.doi.org/10.1088/0305-4608/7/12/022.

[146] P. Coleman, Phys. Rev. B 29 (1984) 3035–3044, http://dx.doi.org/10.1103/PhysRevB.29.3035.

[147] G. Kotliar, A.E. Ruckenstein, Phys. Rev. Lett. 57 (1986) 1362–1365, http://dx.doi.org/10.1103/PhysRevLett.57.1362.

[148] T. Li, P. Wölfle, P.J. Hirschfeld, Phys. Rev. B 40 (1989) 6817–6821, http://dx.doi.org/10.1103/PhysRevB.40.6817.

[149] F. Lechermann, A. Georges, G. Kotliar, O. Parcollet, Phys. Rev. B 76 (2007) 155102, http://dx.doi.org/10.1103/PhysRevB.76.155102.

[150] H. Hasegawa, J. Phys. Soc. Japan 66 (1997) 1391–1397, http://dx.doi.org/10.1143/JPSJ.66.1391.

[151] W. Metzner, D. Vollhardt, Phys. Rev. Lett. 62 (1989) 324–327, http://dx.doi.org/10.1103/PhysRevLett.62.324.

[152] B. Menge, E. Muller-Hartmann, Z. Phys. B 82 (1991) 237–242, http://dx.doi.org/10.1007/BF01324333.

[153] U. Brandt, C. Mielsch, Z. Phys. B 75 (1989) 365–370, http://dx.doi.org/10.1007/BF01321824.

[154] M. Jarrell, Phys. Rev. Lett. 69 (1992) 168–171, http://dx.doi.org/10.1103/PhysRevLett.69.168.

[155] A. Georges, G. Kotliar, Phys. Rev. B 45 (1992) 6479–6483, http://dx.doi.org/10.1103/PhysRevB.45.6479.

[156] A. Georges, G. Kotliar, W. Krauth, M.J. Rozenberg, Rev. Modern Phys. 68 (1996) 13–125, http://dx.doi.org/10.1103/RevModPhys.68.13.

[157] V.I. Anisimov, A.I. Poteryaev, M.A. Korotin, A.O. Anokhin, G. Kotliar, J. Phys. Condens. Matter 9 (1997) 7359–7367, http://dx.doi.org/10.1088/0953-8984/9/35/010.

[158] N. Marzari, D. Vanderbilt, Phys. Rev. B 56 (1997) 12847–12865, http://dx.doi.org/10.1103/PhysRevB.56.12847.

[159] F. Lechermann, A. Georges, A. Poteryaev, S. Biermann, M. Posternak, A. Yamasaki, O.K. Andersen, Phys. Rev. B 74 (2006) 125120, http://dx.doi.org/10.1103/PhysRevB.74.125120.

[160] K. Haule, C.-H. Yee, K. Kim, Phys. Rev. B 81 (2010) 195107, http://dx.doi.org/10.1103/PhysRevB.81.195107.

[161] K. Haule, Private communication.

[162] F. Aryasetiawan, M. Imada, A. Georges, G. Kotliar, S. Biermann, A.I. Lichtenstein, Phys. Rev. B 70 (2004) 195104, http://dx.doi.org/10.1103/PhysRevB.70.195104.

[163] E. Şaşioğlu, C. Friedrich, S. Blügel, Phys. Rev. B 83 (2011) 121101, http://dx.doi.org/10.1103/PhysRevB.83.121101.

[164] A. Georges, in: E. Pavarini, E. Koch, D. Vollhardt, A. Lichtenstein (Eds.), DMFT At 25: Infinite Dimensions: Lecture Notes of the Autumn School on Correlated Electrons 2014, Forschungszentrum Jülich, Jülich, 2014.

[165] T. Maier, M. Jarrell, T. Pruschke, M.H. Hettler, Rev. Modern Phys. 77 (2005) 1027–1080, http://dx.doi.org/10.1103/RevModPhys.77.1027.

[166] E. Gull, A.J. Millis, A.I. Lichtenstein, A.N. Rubtsov, M. Troyer, P. Werner, Rev. Modern Phys. 83 (2011) 349–404, http://dx.doi.org/10.1103/RevModPhys.83.349.

[167] H. Okumura, K. Sato, T. Kotani, Phys. Rev. B 100 (2019) 054419, http://dx.doi.org/10.1103/PhysRevB.100.054419.

[168] V. Antropov, J. Magn. Magn. Mater. 262 (2003) L192–L197, http://dx.doi.org/10.1016/S0304-8853(03)00206-3.

[169] A. Liechtenstein, M. Katsnelson, V. Antropov, V. Gubanov, J. Magn. Magn. Mater. 67 (1987) 65–74, http://dx.doi.org/10.1016/0304-8853(87)90721-9.

[170] V. Antropov, B. Harmon, A. Smirnov, J. Magn. Magn. Mater. 200 (1999) 148–166, http://dx.doi.org/10.1016/S0304-8853(99)00425-4.

[171] M. van Schilfgaarde, V.P. Antropov, J. Appl. Phys. 85 (1999) 4827–4829, http://dx.doi.org/10.1063/1.370495.

[172] A. Franceschetti, S.V. Dudiy, S.V. Barabash, A. Zunger, J. Xu, M. van Schilfgaarde, Phys. Rev. Lett. 97 (2006) 047202, http://dx.doi.org/10.1103/PhysRevLett.97.047202.

[173] J. Okabayashi, M. Mizuguchi, K. Ono, M. Oshima, A. Fujimori, H. Kuramochi, H. Akinaga, Phys. Rev. B 70 (2004) 233305, http://dx.doi.org/10.1103/PhysRevB.70.233305.

[174] M.E. Rose, Elementary Theory of Angular Momentum, Wiley, New York, 1957.

[175] Z.H. Levine, D.C. Allan, Phys. Rev. Lett. 63 (1989) 1719–1722, http://dx.doi.org/10.1103/PhysRevLett.63.1719.

[176] R. Del Sole, R. Girlanda, Phys. Rev. B 48 (1993) 11789–11795, http://dx.doi.org/10.1103/PhysRevB.48.11789.

[177] A.R.H. Preston, A. DeMasi, L.F.J. Piper, K.E. Smith, W.R.L. Lambrecht, A. Boonchun, T. Cheiwchanchamnangij, J. Arnemann, M. van Schilfgaarde, B.J. Ruck, Phys. Rev. B 83 (2011) 205106, http://dx.doi.org/10.1103/PhysRevB.83.205106.

[178] F. Bruneval, N. Vast, L. Reining, M. Izquierdo, F. Sirotti, N. Barrett, Phys. Rev. Lett. 97 (2006) 267601, http://dx.doi.org/10.1103/PhysRevLett.97.267601.

[179] C. Aversa, J.E. Sipe, Phys. Rev. B 52 (1995) 14636–14645, http://dx.doi.org/10.1103/PhysRevB.52.14636.

[180] J.E. Sipe, E. Ghahramani, Phys. Rev. B 48 (1993) 11705–11722, http://dx.doi.org/10.1103/PhysRevB.48.11705.

[181] J.E. Sipe, A.I. Shkrebtii, Phys. Rev. B 61 (2000) 5337–5352, http://dx.doi.org/10.1103/PhysRevB.61.5337.

[182] S.N. Rashkeev, W.R.L. Lambrecht, B. Segall, Phys. Rev. B 57 (1998) 3905–3919, http://dx.doi.org/10.1103/PhysRevB.57.3905.

[183] H. Park, (Ph.D. thesis), Rutgers University-Graduate School-New Brunswick, 2011.

[184] S. Acharya, D. Pashov, C. Weber, H. Park, L. Sponza, M. van Schilfgaarde, Commun. Phys. (2019) http://dx.doi.org/10.1038/s42005-019-0254-1, arXiv:1811.05143 (in press).

[185] T. Suzuki, T. Goto, K. Chiba, T. Shinoda, T. Fukase, H. Kimura, K. Yamada, M. Ohashi, Y. Yamaguchi, Phys. Rev. B 57 (1998) R3229–R3232, http://dx.doi.org/10.1103/PhysRevB.57.R3229.

[186] M. Casula, A. Rubtsov, S. Biermann, Phys. Rev. B 85 (2012) 035115, http://dx.doi.org/10.1103/PhysRevB.85.035115.

[187] S. Laricchia, N. Bonini, M. van Schilfgaarde, Bulletin of the American Physical Society, American Physical Society, 2019, P2200011.

[188] N.V. Prokof'ev, B.V. Svistunov, Phys. Rev. Lett. 81 (1998) 2514–2517, http://dx.doi.org/10.1103/PhysRevLett.81.2514.

[189] D. Pashov, HECToR dCSE Report.

[190] M. van Schilfgaarde, N. Newman, Phys. Rev. Lett. 65 (1990) 2728–2731, http://dx.doi.org/10.1103/PhysRevLett.65.2728.

[191] M. Asta, D. de Fontaine, M. van Schilfgaarde, M. Sluiter, M. Methfessel, Phys. Rev. B 46 (1992) 5055–5072, http://dx.doi.org/10.1103/PhysRevB.46.5055.

[192] M. van Schilfgaarde, I.A. Abrikosov, B. Johansson, Nature 400 (1999) 46–49, http://dx.doi.org/10.1038/21848.

[193] A.T. Paxton, M. v. Schilfgaarde, M. MacKenzie, A.J. Craven, J. Phys. Condens. Matter. 12 (2000) 729–750, http://dx.doi.org/10.1088/0953-8984/12/5/319.

[194] T. Kotani, M. van Schilfgaarde, Solid State Commun. 121 (2002) 461–465, http://dx.doi.org/10.1016/S0038-1098(02)00028-5.

[195] T. Miyake, F. Aryasetiawan, T. Kotani, M. van Schilfgaarde, M. Usuda, K. Terakura, Phys. Rev. B 66 (2002) 245103, http://dx.doi.org/10.1103/PhysRevB.66.245103.

[196] S. Krishnamurthy, M. van Schilfgaarde, N. Newman, Appl. Phys. Lett. 83 (2003) 1761–1763, http://dx.doi.org/10.1063/1.1606873.

[197] J.P. Velev, K.D. Belashchenko, D.A. Stewart, M. van Schilfgaarde, S.S. Jaswal, E.Y. Tsymbal, Phys. Rev. Lett. 95 (2005) 216601, http://dx.doi.org/10.1103/PhysRevLett.95.216601.

[198] A.L. Wysocki, R.F. Sabirianov, M. van Schilfgaarde, K.D. Belashchenko, Phys. Rev. B 80 (2009) 224423, http://dx.doi.org/10.1103/PhysRevB.80.224423.

[199] J.J. Krich, B.I. Halperin, Phys. Rev. Lett. 98 (2007) 226802, http://dx.doi.org/10.1103/PhysRevLett.98.226802.

[200] A.M.A. Leguy, P. Azarhoosh, M.I. Alonso, M. Campoy-Quiles, O.J. Weber, J. Yao, D. Bryant, M.T. Weller, J. Nelson, A. Walsh, M. van Schilfgaarde, P.R.F. Barnes, Nanoscale 8 (2016) 6317–6327, http://dx.doi.org/10.1039/C5NR05435D.

[201] J.D. Jackson, Classical Electrodynamics, Wiley, New York, 1999.