Chapter 2.1

Density Functional Theory in Mineral Physics

Lars Stixrude

Density functional theory of the electronic structure of condensed matter is reviewed with an emphasis on its application to geophysics. The review is placed within the context of our attempts to understand planetary interiors and the unique features of these regions that lead us to use band-structure theory. The foundations of density functional theory are briefly discussed, as are its scope and limitations. Special attention is paid to commonly used approximations of the theory, including those of the exchange-correlation potential and the structure of the electronic core. Some of the important computational methods are reviewed, including the linearized augmented plane-wave method and the plane-wave pseudopotential method. Examples of applications of density functional theory to the study of the equation of state, crystalline structure, phase stability, and elasticity of earth materials are described. Some critical areas for further development are identified.

2.1.1 Introduction

Planetary interiors represent a unique environment in the universe in which the behavior of condensed matter presents a considerable challenge. The nature and the evolution of planetary interiors, even that of our own Earth, are complex, poorly understood, and difficult to predict with current theoretical understanding. In contrast, we have a much better understanding in many ways of the interiors of distant stars. For example, we are

H. Aoki et al. (eds), *Physics Meets Mineralogy*

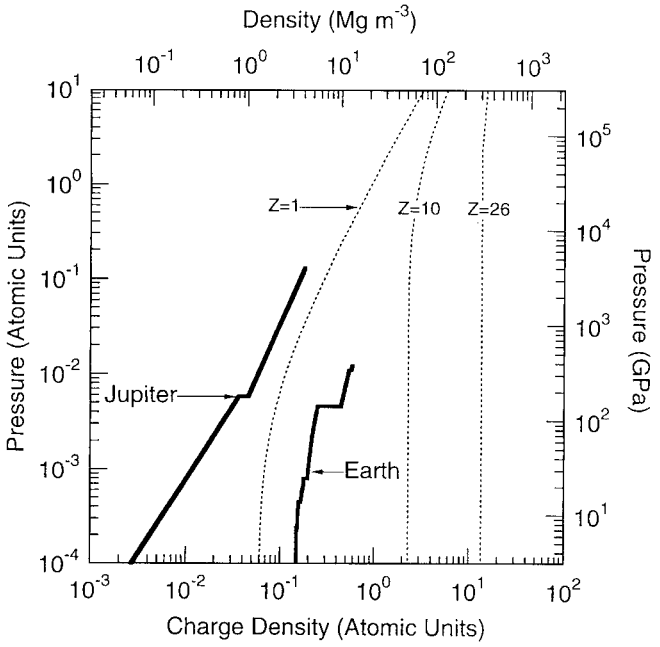

Figure 2.1.1 Pressure in the interior of Jupiter [3] and Earth [4] as a function of mass density (top) and charge density (bottom). The charge density has been calculated from the observed mass density with the assumption that the number of electrons is one-half the number of nucleons. Planetary structures are compared with limiting high-density equation of state (2.1.1) for three values of the atomic number Z.

able to calculate the structures and evolutionary history of stars with some certainty, an exercise that is not yet possible for the Earth.

There are sound physical reasons for this apparent anomaly. The stellar interior is extraordinarily simple from a condensed-matter physicist's point of view. Because the pressure is so high, the electrons obey an almost trivial limiting behavior, the uniform electron gas [1]. The fundamental reason is that the kinetic energy of electrons increases with the charge density ρ as $\rho^{2/3}$, whereas the potential energy binding the electrons to the nuclei increases only as $\rho^{1/3}$; the kinetic energy dominates at high pressure and the electrons become unbound (see Ref. [2] for an extended discussion).

The contrast with planetary interiors can be illustrated when the pressure is expressed in terms of atomic units, 1 atomic unit (29.4 TPa) being comparable with the pressure required for complete ionization and the formation of a degenerate electron gas. The structures of planets are such that pressures are much less than unity (Fig. 2.1.1). The behavior of planetary materials will be far from plasmalike and therefore much more complex.

A more useful pressure scale in the context of planetary interiors is formed from an energy typical of the spacing between electronic bands (1 eV) and a volume typically

occupied by a valence electron in a mineral (25 bohr^3 $\approx$ 4 Å^3). This pressure scale ($\sim$100 GPa) is characteristic of the Earth's interior and the bulk modulus of typical earth-forming constituents. We must then expect to find in planetary interiors not only significant compression and phase transitions, but also electronic transitions (e.g., insulator to metal) and substantial changes in the mechanisms of bonding, all of which complicate our picture of planetary structure and evolution.

From the computational point of view, the recognition that planetary interiors are characterized by complex multiphase behavior places tremendous demands on the required accuracy of theoretical methods. They must be general, applicable to essentially all classes of elements, and must not make any assumptions regarding the nature of bonding. Energies and volumes must be accurate to well within typical heats and volumes of solid–solid transformations. In the case of the Earth, these requirements rule out essentially all weak screening approaches that treat condensed matter in perturbative fashion, beginning with the free-electron gas. Indeed, early calculations for which one such approach was used [5] incorrectly predicted that iron is substantially lighter than the Earth's core [6]. In the case of Jupiter, weak screening approximations have proved fruitful for investigations of planetary structure but are unlikely to successfully capture a priori important details such as the structure of the molecular-to-atomic transition [1, 3].

The following sections review in some detail modern methods of first-principles theory as they have been applied in the geophysics literature. The fundamental approximations on which current implementations of density functional theory are based are discussed. I then discuss applications of the theory to the derivation of observable quantities of geophysical interest. In Section 2.1.3 I review computational methods for solving the equations. Finally, I discuss some important unsolved problems in the behavior of earth materials and possible future directions.

2.1.2 Theory

From the point of view of any first-principles theory, solids are composed of nuclei and electrons; atoms and ions are constructs that play no primary role. This departure from our usual way of thinking about minerals and solids is essential and has the following important consequences. We may expect our theory to be equally applicable to the entire range of conditions encountered in planets (and even stars), the entire range of bonding environments encompassed by this enormous range of pressures and temperatures, and all elements of the periodic table.

To illustrate this way of thinking about solids and to introduce some important concepts, consider first the properties of the simplest system, the uniform electron gas with embedded nuclei. The total energy consists of the kinetic energy of the electrons and three distinct contributions to the potential energy: (1) Coulomb interactions among nuclei and electrons, (2) electron exchange, and (3) electron correlation. The

first contribution is straightforward and involves only sums over point charges and/or integrals over the (uniform) electronic charge density.

Exchange and correlation account for local deviations from uniform charge that arise from the tendency of electrons to avoid each other. Correlation accounts for the mutual Coulomb repulsion, whereas exchange embodies the Pauli exclusion principle and the resulting tendency of electrons of the same spin to avoid each other. The net effect is that each electron can be thought of as digging a hole of reduced charge density about itself. Certain properties of the exchange-correlation hole are well understood; it is known for instance that its integrated charge must exactly balance that of the electron. Exchange and correlation reduce the total energy by reducing the Coulomb repulsion between electrons.

The total energy of our simple system is readily evaluated as a function of charge density; the equation of state then follows from differentiation. With the assumption that the nuclei are in a close-packed arrangement and with only the leading-order high-density contributions to exchange and correlation included, the equation of state is [7]

$$P = 0.176 r_s^{-5}[1 - (0.407Z^{2/3} + 0.207)r_s], \tag{2.1.1}$$

where P is the static (athermal) pressure, Z is the nuclear charge, and the Wigner–Seitz radius,

$$r_s = \left(\frac{3}{4\pi\rho}\right)^{1/3}, \tag{2.1.2}$$

is a measure of the average spacing between electrons. The first term in Eq. (2.1.1) is the kinetic contribution, the second is due to the Coulomb attraction of the nuclei for the electrons and mutual repulsion of the electrons, and the third is due to exchange. Correlation, which is smaller than exchange at high density, has been neglected, as has the mutual Coulomb repulsion of the nuclei.

Comparisons with the structure of planetary interiors reveal some fundamentally important aspects of planetary matter (Fig. 2.1.1). First, the net Coulomb attraction provided by the nuclei plays an essential role at planetary densities – different mean nuclear charges account to first order for the difference in mean charge (and mass) density between Jupiter and Earth. Second, screening also has a first-order effect on the equation of state, accounting for the much lower densities of planets at a given pressure than those predicted by Eq. (2.1.1). In planetary matter, the charge density is substantially enhanced in the vicinity of the nucleus, reducing the ability of the point charges to attract the remaining (valence) electrons. Screening is weaker in the case of Jupiter because it contains dominantly lighter elements and because the pressures are much greater. Nevertheless, for all the planets screening is so strong that it must be accounted for. In the case of the terrestrial planets, the charge density near the nuclei

is so much higher than in the interstitial region that this difference plays a central role in the design of modern computational methods.

2.1.2.1 Density Functional Theory

We turn now from simple to real systems and at the same time from analytically expressible results to necessarily elaborate computations. Although the electronic structure will be nontrivial, we retain the charge density as a central concept. This is appealing because this quantity is experimentally observable; it is precisely what is measured by an x-ray-diffraction experiment.

The general problem we are faced with in a nonuniform, nondegenerate electron system is formidable. Given a periodic potential set by the positions of the nuclei, we must solve the Schrödinger equation for the total wave function $\Psi(r_1, r_2, \ldots, r_N)$ of a system of N interacting electrons, where N is of the order of Avogadro's number. Density functional theory [8, 9] is a powerful and, in principle, exact method of dealing with this problem in a tractable way (see Ref. [10] for reviews).

The essence of this theory is the proof that the ground-state properties of a material are a unique functional of the charge density $\rho(\mathbf{r})$. Among these properties are the ground-state total energy,

$$E = T + U\,[\rho(\mathbf{r})] + E_{\mathrm{xc}}\,[\rho(\mathbf{r})]\,, \tag{2.1.3}$$

and its derivatives (pressure, elastic constants, etc.), where T is the kinetic energy of a system of noninteracting electrons with the same charge density as that of the interacting system, U is the electrostatic (Coulomb) energy, including the electrostatic interaction between the nuclei, and E_{xc} is the exchange-correlation energy. A variational principle leads to a set of single-particle, Schrödinger-like, Kohn–Sham equations,

$$[-\nabla^2 + V_{\mathrm{KS}}]\psi_i = \epsilon_i \psi_i\,, \tag{2.1.4}$$

where ψ_i is now the wave function of a single electron, ϵ_i is the corresponding eigenvalue, and the effective potential is

$$V_{\mathrm{KS}}\,[\rho(\mathbf{r})] = \sum_{i=1}^{N} \frac{2Z_i}{|\mathbf{r} - \mathbf{R}_i|} + \int \frac{2\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}\,\mathrm{d}\mathbf{r}' + V_{\mathrm{xc}}\,[\rho(\mathbf{r})]\,, \tag{2.1.5}$$

where the first two terms are Coulomb potentials that are due to the nuclei and the other electrons, respectively, the last is the exchange-correlation potential, and the units are Rydberg atomic units: $\hbar^2/2m = 1$, $e^2 = 2$, energy in Ry and length is in bohrs.

The power of density functional theory is that it allows us to calculate, in principle, the exact many-body total energy of a system from a set of single-particle equations. The solution to the Kohn–Sham equations is that of the set of coupled generalized

eigenvalue equations:

$$H_{ij}(\mathbf{k})\psi_j(\mathbf{r},\mathbf{k}) = \epsilon_j(\mathbf{k})O_{ij}(\mathbf{k})\psi_j(\mathbf{r},\mathbf{k}), \tag{2.1.6}$$

$$H_{ij}(\mathbf{k}) = \int \psi_i^*(\mathbf{r},\mathbf{k})(-\nabla^2 + V_{\mathrm{KS}})\psi_j(\mathbf{r},\mathbf{k})\,\mathrm{d}\mathbf{r}, \tag{2.1.7}$$

$$O_{ij}(\mathbf{k}) = \int \psi_i^*(\mathbf{r},\mathbf{k})\psi_j(\mathbf{r},\mathbf{k})\,\mathrm{d}\mathbf{r}, \tag{2.1.8}$$

where **H** and **O** are the Hamiltonian and the overlap matrices, respectively, and **k** is a vector in reciprocal space. Because the Kohn–Sham potential is a functional of the charge density, the equations must be solved self-consistently together with the definition of the charge density in terms of the wave functions:

$$\rho(\mathbf{r}) = \int \sum_i n[E_F - \epsilon_i(\mathbf{k})]\psi_i^*(\mathbf{r},\mathbf{k})\psi_i(\mathbf{r},\mathbf{k})\,\mathrm{d}\mathbf{k}, \tag{2.1.9}$$

where n is the occupation number and E_F is the Fermi energy.

2.1.2.2 Approximations

Exchange-Correlation Potential

The Kohn–Sham equations are exact. That only approximate solutions have been possible to date is a limitation imposed only by our current ignorance of the exact exchange-correlation functional. If the exact exchange-correlation functional were known, we would be able to obtain exact solutions. All other terms in the Kohn–Sham equations are straightforward and readily evaluated.

The exchange-correlation functional is known precisely only for simple systems such as the uniform electron gas (Fig. 2.1.2). The exchange portion is known analytically, as are the leading-order contributions to correlation in the limit of high density [12]:

$$\begin{aligned} V_{\mathrm{xc}} &= \frac{\partial}{\partial\rho}(\rho E_{\mathrm{xc}}), \\ E_{\mathrm{xc}} &= -\frac{3}{4\pi}\left(\frac{9\pi}{4}\right)^{1/3} r_s^{-1} + A\ln r_s + B, \end{aligned} \tag{2.1.10}$$

where the first contribution to E_{xc} is exchange and the constants $A = (1 - \ln 2)/\pi$ and $B = -0.046644$ [13]. At other densities, accurate values of the exchange-correlation potential are known from quantum Monte Carlo calculations [14], which have been represented in a parametric form that obeys the high-density limiting behavior [Eqs. (2.1.10)] [15].

The precision of modern condensed-matter computations has made the accurate representation of the exchange-correlation potential of the uniform electron gas an important issue. In this context, it is important to be aware that approximate representations of V_{xc} have appeared frequently in the geophysical literature and are still

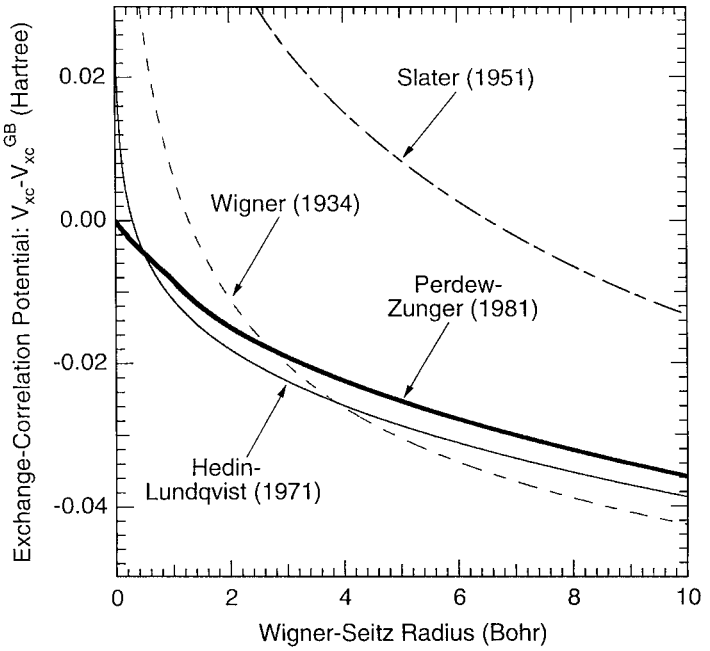

Figure 2.1.2 Difference between the exchange-correlation potential and its high-density limit [Eqs. (2.1.10)] in (bold curve) the local-density approximation and (other curves) three commonly used approximations to the local-density approximation. For the Slater [11] result, I have used $\alpha = 2/3$, which yields the pure exchange potential.

in use. Of these, the Hedin–Lundqvist [16] expression is most similar to the accurate Perdew–Zunger parameterization, that of Wigner [17] the least. The Wigner approximation shows a much stronger dependence on density than the accurate potential and leads to significant errors in density functional computations for solids. None of the commonly used approximate expressions satisfy the correct high-density limiting behavior [Eqs. (2.1.10)].

The charge density in real materials is highly nonuniform, and the exchange-correlation potential cannot be evaluated. Fortunately, simple approximations of the exchange-correlation potential have been very successful. The local-density approximation (LDA) is based on the uniform electron gas, taking into account nonuniformity to lowest order by setting V_{xc} at every point in the crystal to that of the uniform electron gas with a density equal to the local charge density [10].

The success of the LDA can be understood at a fundamental level in terms of the satisfaction of exact sum rules for the exchange-correlation hole [18]. For example, the LDA correctly predicts an exchange-correlation hole of unit charge. Ultimately, the appropriateness of the LDA can be judged only by comparison of its predictions with observation. Here the LDA has been remarkably successful. The LDA has been shown to yield excellent agreement with experiment for a wide variety of insulators, metals, and semiconductors and for bulk, surface, and defect properties. The LDA

shows some general failures, such as a tendency to underpredict bandgaps, and, from a geophysical point of view, more serious failures for certain materials. For example, LDA fails to predict the correct ground state of iron.

These failures have prompted the development of new exchange-correlation functionals. One shortcoming of the LDA may be its local character, that is, its inability to distinguish among electrons of different angular momenta or energy. Generalized gradient approximations (GGAs) partially remedy this by including a dependence on local charge-density gradients in addition to the density itself. Some care must be taken in constructing gradient approximations; a straightforward Taylor series expansion in the charge-density gradient about the LDA result fails completely because it violates the sum rule for the exchange-correlation hole. The most widely used GGA satisfies the sum rules exactly [13]. This approximation and its forerunners have been shown to yield agreement with experimental data that is usually as good as the LDA and often substantially better [19]. For example, GGAs correctly predict the bcc phase as the ground state of iron [20]. The relationship of the GGA to the LDA can be expressed in terms of the enhancement factor,

$$F(r_s, s) = \frac{V_{xc}^{GGA}(r_s, s)}{V_x(r_s)}, \tag{2.1.11}$$

where V_x is the exchange potential and F is a function of the charge density and the nondimensional charge-density gradient, $s = (24\pi^2)^{-1/3}|\nabla\rho/\rho^{4/3}|$ (Fig. 2.1.3).

Frozen-Core Approximation

The physical motivation for this approximation is the observation that only the valence electrons participate in bonding and in the response of the crystal to most perturbations of interest. Unless the perturbation is of very high energy (comparable with the binding energy of the core states), the tightly bound core states remain essentially unchanged. The frozen-core approximation is satisfied to a high degree of accuracy for many applications, for example in the case of finite strains of magnitudes typically encountered in the Earth's interior.

Within this approximation, the charge density of the core electrons is just that of the free atom, which can be found readily. We then need solve for only the valence electrons in Eq. (2.1.4), often a considerable computational advantage. An important technical point is that, although in many cases the choice is obvious, there is no fundamentally sound way to decide a priori which electrons are core and which are valence. Some care is required; for example, the $3p$ electrons in iron must be treated as valence electrons as they are found to deform substantially at pressures comparable with those in the Earth's core [21].

Pseudopotential Approximation

This approximation goes one step beyond the frozen-core approximation. It replaces the nucleus and the core electrons with a simpler object, the pseudopotential, which

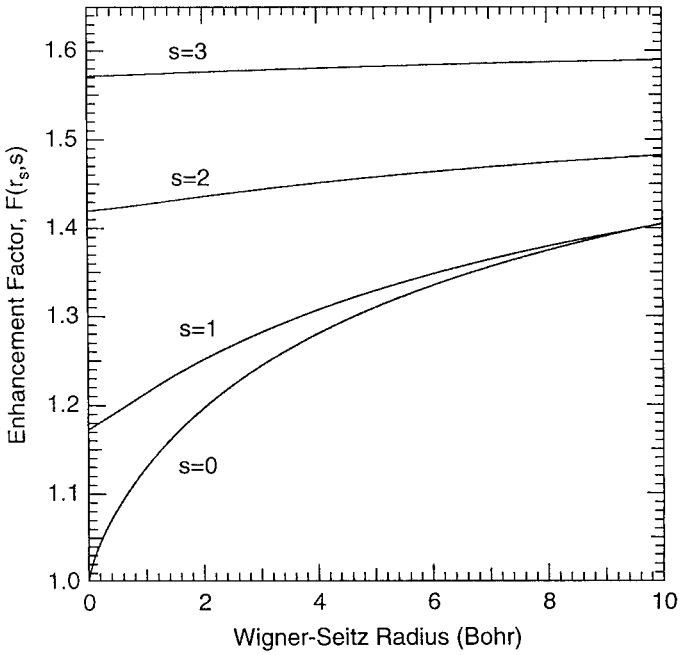

Figure 2.1.3 Effect of charge-density gradients on the exchange-correlation potential according to the GGA. The enhancement factor for zero gradient ($s = 0$) reflects the contribution of correlation to V_{xc}.

has the same scattering properties [22]. The pseudopotential is chosen such that the valence wave function in the free atom is the same as the all-electron solution beyond some cutoff radius, but nodeless within this radius. The advantages of the pseudopotential method are that (1) spatial variations in the pseudopotential are much less rapid than the bare Coulomb potential of the nucleus and (2) we need solve for only the (pseudo-) wave functions of the valence electrons, which show much less rapid spatial variation than the core electrons or the valence electrons in the core region. This means that in the solution of the Kohn–Sham equations, potential and charge density can be represented by a particularly simple, complete, and orthogonal set of basis functions (plane waves) of manageable size; with this basis set, evaluation of total energies, stresses, and forces acting on the atoms is particularly efficient.

The pseudopotential is an approximation to the potential that the valence electrons "see" and its construction is nonunique; different pseudopotentials may yield significantly different predictions of bulk properties. Several different methods for constructing pseudopotentials have been developed [23–25]. Care must be taken to demonstrate the transferability of the pseudopotentials generated by a particular method and to compare with all-electron calculations where these are available. When these conditions

are met, the error that is due to the pseudopotential is generally small (a few percent in volume for earth materials).

2.1.2.3 Derivation of Observables

Total Energy and Band Structure

For a given arrangement of nuclei (crystal structure) we may solve the equations of density functional theory under one or more of the above approximations to determine the total energy, charge density, and the quasiparticle eigenvalue spectrum (electronic band structure). By examining the dependence of the total energy on perturbations to the volume V or shape of the crystal (described by the deviatoric strain tensor ϵ'_{ij}) or to the positions of the atoms, we can, in principle, deduce the Helmholtz free energy F as a function of V, ϵ'_{ij}, and T. For example, the pressure and the equation of state are simply given by the variation of the total energy with volume.

We may determine the elastic constants from total-energy calculations. For small deviatoric strains under hydrostatic stress [26],

$$F(V, \epsilon'_{ij}, T) = F_0(V) + F_{\mathrm{TH}}(V, T) + \frac{1}{2} c_{ijkl}(V, T)\epsilon'_{ij}\epsilon'_{kl}, \tag{2.1.12}$$

where F_0 is the static (zero-temperature) contribution, F_{TH} is due to the thermal excitation of electrons and phonons, and c_{ijkl} is the elastic-constant tensor. This equation shows that combinations of elastic constants are related to the difference in total energy between a strained and an unstrained lattice.

It is possible in principle to calculate thermal contributions to the thermodynamic and thermoelastic properties of crystals. Calculating thermal properties is much more difficult than calculating static properties, however. The reason is simple: The atomic vibrations induced by finite temperature break the symmetry of the crystal so that it is now periodic in only a time-averaged sense. In the context of total-energy calculations, our task is then to evaluate the partition function, an integral over all atomic configurations realized by a crystal at high temperature. This is essentially impossible. More efficient ways of evaluating thermal free energies from first principles are required. Some future directions are indicated in Section 2.1.5.

Forces, Stresses, and Structures

The Hellman–Feynman theorem allows us to calculate first derivatives of the total energy directly in terms of the ground-state wave functions. The application of this theorem allows us to determine the forces acting on every atom and the stresses acting on the lattice.

This is important for two related reasons. First, it allows us to determine ground-state crystal structures very effectively. This has become possible only recently for relatively complex structures such as $MgSiO_3$ perovskite [27]. The key innovation

has been the development of a structural optimization strategy based on a pseudo-Lagrangian that treats the components of the strain tensor and the atomic positions as dynamical variables [28]. The optimization is performed at constant pressure. At each step of the dynamical trajectory, the Hellman–Feynman forces and stresses [29] acting on the nuclei and the lattice parameters, respectively, are evaluated and used to generate the next configuration. The optimization is complete when the forces on the nuclei vanish and the stress is hydrostatic and balances the applied pressure.

Second, once the equilibrium structure at a given pressure is determined, we can calculate the elastic constants. We do this in a straightforward way by applying a deviatoric strain to the lattice and calculating the resulting stress tensor. The elastic constant c_{ijkl} is then given by the ratio of stress σ_{ij} to strain ϵ_{ij}:

$$\sigma_{ij} = c_{ijkl}\epsilon_{kl}. \tag{2.1.13}$$

Care must be taken to reoptimize the positions of the atoms in each strained configuration, as vibrational modes typically couple with lattice strains in silicate structures.

2.1.3 Computation

2.1.3.1 Methods

First-principles methods solve Eqs. (2.1.6)–(2.1.8) by expanding the wave functions in a basis

$$\psi_i(\mathbf{r}, \mathbf{k}) = \sum_{j=1}^{N} c_{ij}(\mathbf{k})\phi_j(\mathbf{r}, \mathbf{k}), \tag{2.1.14}$$

where N is the number of basis functions ϕ_j and c_{ij} are the coefficients to be determined by solution of the Kohn–Sham equations.

The linearized augmented plane-wave (LAPW) method is the current state of the art in density functional theory computations. It makes no essential approximations beyond that to the exchange-correlation functional, allowing us to solve routinely for all electrons, both core and valence. For example, it makes no approximations of the shape of the charge density or potential. The accurate representation of the potential and the core states means that the LAPW method is equally applicable to all elements of the periodic table and over the entire range of densities of interest in planetary or astrophysical studies.

The LAPW method differs from its forerunner, the APW method, in that the APW method assumes a spherically symmetric potential near the nuclei [30, 31]. Because of its precise representation of the potential, the LAPW method is sometimes referred to as the full-potential LAPW (FLAPW). LAPW shares the ability to represent precisely the full potential and the charge density with the full-potential linear muffin-tin orbital

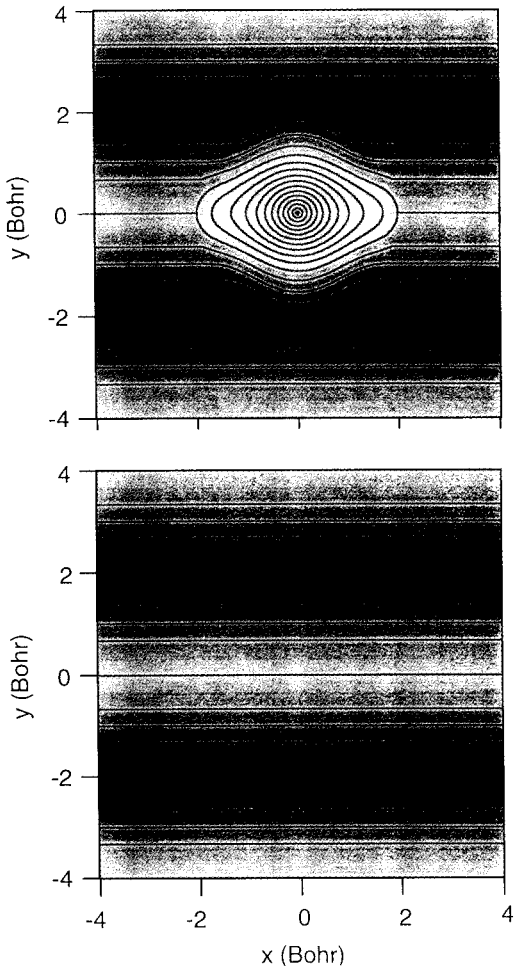

Figure 2.1.4 (Top) One LAPW basis function in the vicinity of a hydrogen nucleus located at the origin: $G = (0, \pi/2, 0)$, $l_{max} = 6$, $E_l(\text{Ry}) = -(l+1)^{-2}$; (bottom) the plane wave $G = (0, \pi/2, 0)$.

(FP-LMTO) method [32]. The FP-LMTO method is very similar to the LAPW method in its capabilities and level of accuracy, differing primarily in the details of the basis functions.

The accuracy and the flexibility of the LAPW method are derived from its basis, which explicitly treats the first-order partitioning of space into near-nucleus regions, where the charge density and its spatial variability are large, and interstitial regions, where the charge density varies more slowly (Fig. 2.1.4) [33–35]. These two regions are delimited by the construction of so-called muffin-tin spheres of radius R^{α}_{MT} centered on each nucleus α. A dual-basis set is constructed, consisting of plane waves in the interstitial regions that are matched continuously to more rapidly varying functions inside the spheres. Within the muffin-tin spheres ($r' < R^{\alpha}_{\text{MT}}$),

$$\phi^{\mathbf{k}+\mathbf{G}}(\mathbf{r}) = \left[a^{\alpha}_{lm} u_l\left(E^{\alpha}_l, r'\right) + b^{\alpha}_{lm} \dot{u}_l\left(E^{\alpha}_l, r'\right)\right] Y_{lm}(\mathbf{r}'/r'), \tag{2.1.15}$$

and for $r' > R_{\mathrm{MT}}^{\alpha}$,

$$\phi^{\mathbf{k}+\mathbf{G}}(\mathbf{r}) = \exp[i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}], \tag{2.1.16}$$

where $\mathbf{r}' = \mathbf{r} - \mathbf{R}_{\alpha}$, $\mathbf{R}_{\alpha}$ are the positions of the nuclei, $\mathbf{G}$ is a wave vector, u_l and $\dot{u}_l$ are the solution to the radial part of the Schrödinger equation and its energy derivative, respectively, for the spherically symmetric portion of the potential inside the muffin-tin sphere at energy E_l, and the coefficients a and b are determined by requiring continuity of the basis function and its first radial derivative on the muffin-tin sphere.

With this basis set, all-electron calculations for silicates or transition metals typically require of the order of 100 basis functions per atom. The primary disadvantage of the LAPW method is that the complexity of the basis functions makes it relatively intensive computationally. In practice, this limits the size of the system that can be studied. Even so, LAPW computations for structures as complex as that of $MgSiO_3$ perovskite (20 atoms in the unit cell) have been performed [36].

Basis sets consisting solely of plane waves, because of their analytical simplicity, have some advantages over the LAPW basis. However, all-electron calculations are virtually impossible with a plane-wave basis set; the number of basis functions needed to represent the rapid spatial oscillations of the core region is much too large to be practical. For this reason, the plane-wave basis is generally linked in practice to the pseudopotential approximation, in which the Fourier content of charge density and potential are limited by design.

2.1.3.2 Convergence

There are two primary convergence issues: the size of the basis and the integrations over reciprocal space [e.g., Eq. (2.1.9)]. Both basis sets have the property of smooth convergence; this means that convergence of the computations is readily assessed; quantities of interest vary smoothly as the basis-set size is increased. In the LAPW method, the size of the basis set is described by the dimensionless quantity $R_{\mathrm{MT}}K_{\max}$, where $K_{\max}$ is the maximum wave number of the plane waves included in the basis set. In the pseudopotential method, the size of the basis is set by the maximum kinetic energy of the plane waves $E_{\mathrm{cut}} = K_{\max}^2$ in atomic units. Typical values for computations of silicates are $R_{\mathrm{MT}}K_{\max} = 7$ and $E_{\mathrm{cut}} = 40-80$ Ry, depending on the pseudopotential that is used.

Sampling of the Brillouin zone is treated with the special-points method, which has been shown to yield rapid convergence [37]. For insulators, only a few points (1–10) are typically needed to achieve fully converged total energies; metals require denser sampling because of the often complex structure of the Fermi surface. The special-points method constructs a uniform grid of k points of specified resolution in the first Brillouin zone. The resulting set of k points is divided into subgroups (stars) of symmetrically equivalent points. The Kohn–Sham equations are solved for only one

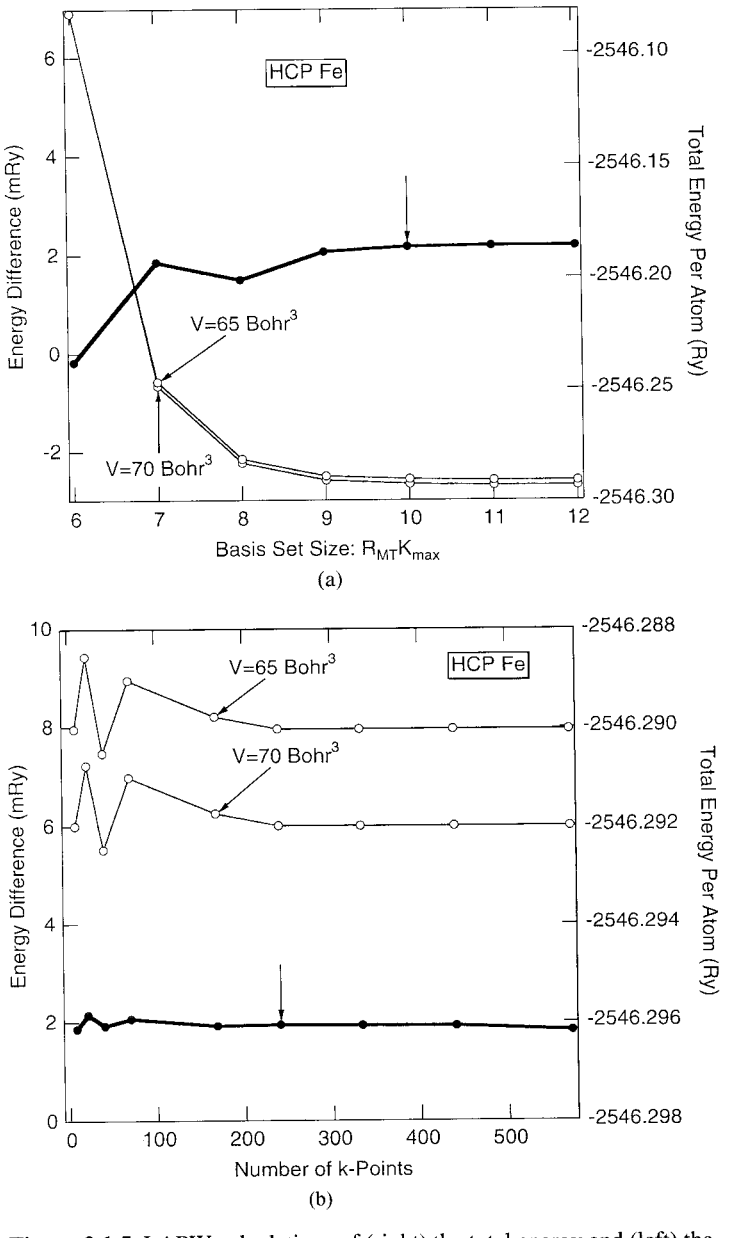

Figure 2.1.5 LAPW calculations of (right) the total energy and (left) the difference in total energy of hcp iron at two different volumes as a function of (a) the size of the basis and (b) the number of k points in reciprocal space.

member of each star, and the wave functions at other points in the star are reconstructed with the appropriate symmetry operations, weighting the contribution of each star by its degeneracy.

The results of convergence tests typical of a transition metal are shown in Fig. 2.1.5. Convergence of the total energy to within a few tenths of a millirydberg ($\sim$50 parts per 10^9 in the case of iron) are routinely achieved. This level of convergence is essential for making accurate predictions – typical solid–solid heats of transformation are of the order of a few millirydbergs. A general feature of convergence in either LAPW or pseudopotential methods is that energy differences – between two volumes of one structure or between two different structures – converge much faster than total energies.

2.1.4 Some Applications

2.1.4.1 Equation of State

The error that is due to the LDA can be evaluated by a comparison of the results of LAPW calculations, which make no further essential approximations beyond the LDA, with experiment (Fig. 2.1.6). In investigations of silicates and oxides of geophysical interest, it has been found that errors in volumes are typically 1%–4%, with theoretical volumes being uniformly smaller than experimental [36, 39–41]. Part of this small difference is due to the higher temperatures of experiments (300 K) compared with the athermal calculations. This is a highly satisfactory level of agreement for a theory that is parameter free and independent of experiment. All-electron LDA computations of transition metals show errors of similar magnitude in the zero-pressure volume; for the $3d$ and the $4d$ metals, the calculations uniformly underestimate the experimental volumes, whereas for the $5d$ metals, the situation is more complex [42]. For the heaviest materials, additional effects such as spin-orbit coupling, often neglected in computations, may become important and contribute to the discrepancy between theory and experiment.

The GGA improves the agreement between theory and experimental equations of state for most materials, including the $3d$ transition metals. In the case of iron, LAPW and FP-LMTO calculations differ from the experimentally measured room-temperature equation of state by 3% at zero pressure and by less than 1% at core pressures [21, 32, 43]; agreement with high-temperature Hugoniot data is equally good [44].

Pseudopotential calculations make additional approximations that lead to additional errors (Fig. 2.1.6). These are small in magnitude and comparable in size with the LDA error. At this level of detail, different pseudopotentials yield results that differ from each other and from the all-electron LDA result from LAPW. Because the pseudopotential method is nearly as accurate as the much more elaborate LAPW method, it is often

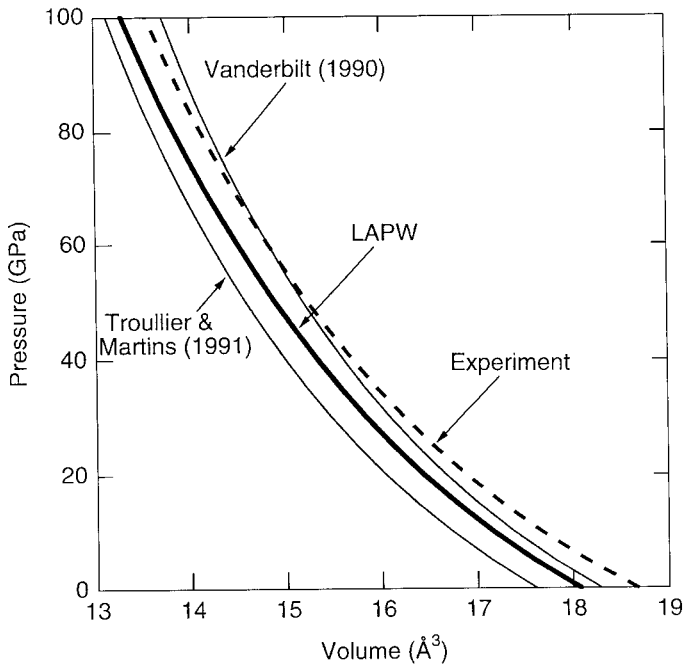

Figure 2.1.6 Equation of state of MgO from (dashed curve) experiment [38], (bold curve) all-electron LDA (LAPW) calculations, (thin curves) pseudopotential calculations based on the indicated potentials.

preferred for many applications, as its computational advantages allow much larger and more complex systems to be studied.

2.1.4.2 Structural Optimization

The structure and compression mechanisms of a number of complex silicates have been studied with density functional theory at high pressure, including $MgSiO_3$ enstatite and perovskite, Mg_2SiO_4 forsterite, and SiO_2 in the quartz, stishovite, $CaCl_2$, and columbite structures [27, 40, 41, 45–50]. These investigations (1) provide an important test of the approximations on which first-principles methods are based (2) illustrate in detail often not obtainable by experiment the nature of compression mechanisms, and (3) provide a sensitive test of the hypothesis that some minerals undergo high-order symmetry-invariant phase transformations.

When the method of Wentzcovitch [28] is used, the optimization of complex crystal structures such as forsterite is an efficient procedure. Typically, of the order of 10–20 iterations are required for full structural convergence in this mineral with 7 internal degrees of freedom and three lattice parameters [48]. The results of first-principles calculations show that volume compression is primarily accommodated by nearly isotropic compression of the MgO_6 octahedra, which are much softer than the SiO_4

tetrahedra. Changes in bond lengths and angles are smooth and monotonic, unlike the results of x-ray diffraction experiments. In particular, density functional theory does not support the proposal that compression mechanisms change suddenly at pressures near 9 GPa. Brodholt et al. [51], by using different pseudopotentials, also found no evidence for sudden changes in compression mechanisms.

2.1.4.3 Phase Stability

In many ways, phase stability provides the most stringent test of first-principles methods. The reason is that we are comparing total energies computed for two different structures with different basis sets and Brillouin zones at the level of heats of transformations, generally a miniscule fraction of the total energy (less than 1 part per 10^6).

First-principles LDA results for transformations in oxides and silicates have shown excellent agreement with experiment [39–41, 50, 52–54]. In the case of transition metals, the form of the exchange-correlation potential is critical. LDA fails to predict the correct ground state of iron, finding incorrectly that the hcp phase has a lower total energy than the bcc. The GGA correctly recovers the bcc ground state (Fig. 2.1.7). Moreover, it accurately predicts the pressure of the phase transition from bcc to hcp

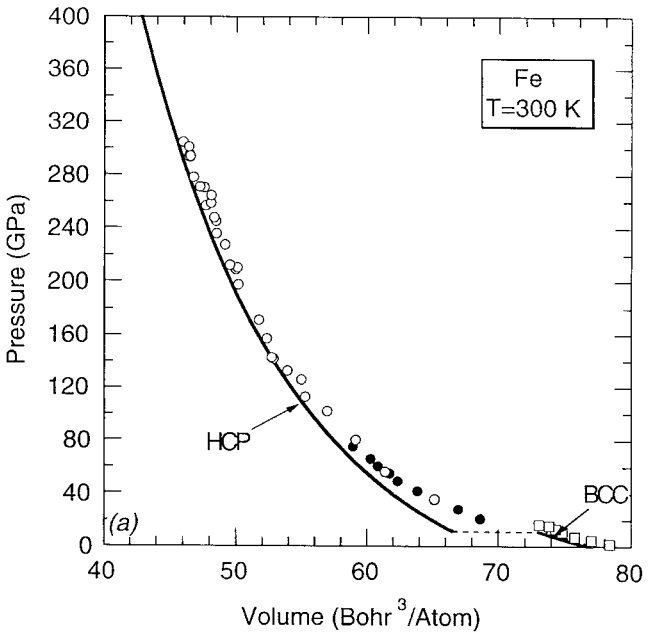

Figure 2.1.7 GGA equations of state of ferromagnetic bcc and nonmagnetic hcp (solid curve) phases of iron compared with the experimental data of Jephcoat et al. [55] (bcc, open squares; hcp, filled circles) and Mao et al. [56] (hcp, open circles). The dashed line indicates the predicted phase-transition pressure of 11 GPa. (From Ref. [44].)

near 11 GPa. This is an important result because the energetics are particularly subtle in the case of this transition as it involves a ferromagnetic and a nonmagnetic phase.

2.1.4.4 Elastic Constants

The elastic constants are of central importance geophysically because they govern the passage of seismic waves, our primary source of information on the structure of the Earth's interior. Despite their importance, density functional calculations of the elastic constants of earth materials have appeared only recently. The key development has been that of an efficient structural optimization scheme [27] and calculation of stresses from the Hellman–Feynman theorem. Elastic constants are determined by calculation of the stress generated by deviatoric strains applied to the equilibrium structure. It is straightforward to demonstrate that we are within the linear regime by performing the calculation at a variety of values of the strain magnitude and extrapolating to the limit of zero strain [54]. These calculations show that strains of the order of 1% are appropriate for silicates and oxides. By carefully choosing the symmetry of the applied strain, we can calculate all elements of the elastic-constant tensor with a small number of different strains. For example, the three elastic constants of a cubic mineral can be determined from a single strain; four different strains have been used for orthorhombic minerals (nine independent elastic constants) [57].

The full elastic-constant tensors of a number of silicates and oxides have been determined with the plane-wave pseudopotential method, including that of MgO periclase, $MgSiO_3$ perovskite, Mg_2SiO_4 forsterite and ringwoodite, and SiO_2 in the stishovite, $CaCl_2$, and columbite structures [27, 47, 54, 57–59]. Once the elastic-constant tensor is determined, it is straightforward to calculate the elastic-wave (seismic) velocities in any direction, the elastic anisotropy, and the seismic-wave velocities of isotropic aggregates. Results for $MgSiO_3$ perovskite [47] (Fig. 2.1.8) show several geophysically important features: (1) athermal longitudinal and shear-wave velocities of isotropic aggregates of this mineral nearly parallel those of the lower mantle and are uniformly larger, (2) perovskite remains highly anisotropic throughout the pressure regime of the lower mantle, and (3) its anisotropy changes qualitatively with increasing pressure. The theoretical results are consistent with the hypothesis that perovskite is the most abundant mineral in the lower mantle [66]; they further require that it exist in approximately randomly oriented aggregates, except possibly in the D'' region, where it may partially account for the seismic anisotropy observed there.

2.1.5 Future Directions

This review has only given an indication of the realm of geophysical application of density functional theory. I have reviewed only a subset of the important calculations that have been performed and have not touched on possible future directions.

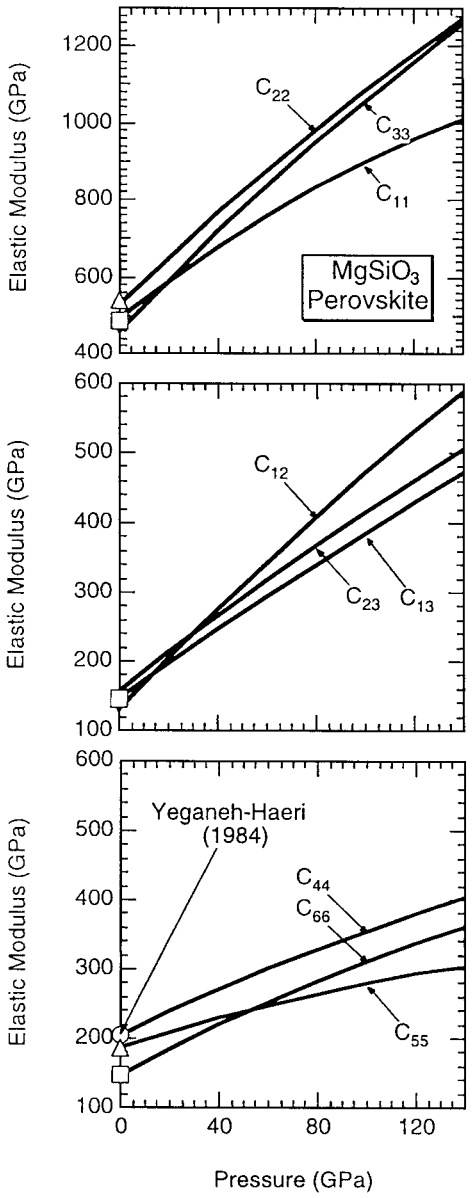

Figure 2.1.8 Elastic constants of $MgSiO_3$ perovskite according to theory [47] (lines) and experiment [60] symbols: c_{11}, c_{12}, c_{44} (○); c_{22}, c_{13}, c_{55} (△); c_{33}, c_{23}, c_{66} (□).

Other applications of density functional theory include the following:

1. The investigation of magnetism in metal alloys, silicates, and oxides. Studies of transition-metal compounds have shown that earth materials display remarkably rich magnetic behavior that has important implications for our understanding of bonding, equations of state, elasticity, and phase stability [61]. This is a challenging area that requires new theoretical developments because transition-metal oxides such as hematite and magnetite are Mott insulators that owe their electronic properties to strong localization of the d electrons. Mott-insulating behavior is fundamentally beyond the scope of band-structure theory as outlined here.
2. First-principles computation of phonon spectra [62, 63]. In addition to making contact with experimental observation of zone-center vibrational frequencies, these predictions allow us to investigate phase stability and, to the extent that thermal properties are quasi harmonic, high-temperature properties [67]. The computation in polar substances is subtle and necessarily involves not only the calculation of force constants, but also that of the dielectric constant and Born effective charge tensors so that coupling to the macroscopic field at zone center is properly accounted for. Efficient calculation involves a technique known as linear response, in which computation of the linear response of the charge density to a perturbation allows second derivatives of the total energy to be computed directly [64].
3. First-principles molecular dynamics simulations at high temperatures. These have now been performed for the first time in a study of liquid iron at core conditions [65]. The method is general and relies on only the principles of the plane-wave pseudopotential method and the computation of stresses and forces outlined here. The first-principles investigation of other solid and fluid earth materials by this technique represents an exciting future direction.

2.1.6 Conclusions

Modern first-principles methods are now capable of realistic predictions of many experimentally observable and geophysically important properties such as the equation of state, phase stability, crystal structure, and elasticity. Parameter free and completely independent of experiment, these methods have been shown to reproduce observations of even subtle features such as phase transitions and the elastic anisotropy with good accuracy. Density functional theory represents the ideal complement to the experimental approach toward studying the behavior of earth materials under extreme conditions. Accelerated progress is predicted on a number of fronts in this challenging field, resulting from the continued interplay of theory and experiment.

Acknowledgment

This work was supported by the U.S. National Science Foundation under grant EAR-9628199.

References

[1] S. Ichimaru, Rev. Mod. Phys. **54**, 1017 (1982).

[2] M. S. T. Bukowinski, Annu. Rev. Earth Planet. Sci., **22**, 167 (1994).

[3] G. Chabrier, Astrophys. J. **391**, 817 (1992).

[4] A. M. Dziewonski and D. L. Anderson, Phys. Earth Planet. Inter. **25**, 297 (1981).

[5] H. Jensen, Z. Phys. **111**, 373 (1938).

[6] F. Birch, J. Geophys. Res. **57**, 227 (1952).

[7] W. B. Hubbard, *Planetary Interiors* (Van Nostrand Reinhold, New York, 1984).

[8] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).

[9] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).

[10] S. Lundqvist and N. H. March, *Theory of the Inhomogeneous Electron Gas* (Plenum, London, 1987).

[11] J. C. Slater, Phys. Rev., **81**, 385 (1951).

[12] M. Gell-Mann and K. A. Brueckner, Phys. Rev. **106**, 364 (1957).

[13] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).

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

[15] J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).

[16] L. Hedin, B. I. Lundqvist, 1971. J. of Phys., **4**, 2064 (1971).

[17] E. Wigner, Phys. Rev., **46**, 1002 (1934).

[18] O. Gunnarsson and B. I. Lundqvist, Phys. Rev. B **13**, 4274 (1976).

[19] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais, Phys. Rev. B **46**, 6671 (1992).

[20] P. Bagno, O. Jepsen, and O. Gunnarson, Phys. Rev. B **40**, 1997 (1989).

[21] L. Stixrude, R. E. Cohen, and D. J. Singh, Phys. Rev. B **50**, 6442 (1994).

[22] W. E. Pickett, Comput. Phys. Rep. **9**, 114 (1989).

[23] D. Vanderbilt, Phys. Rev. B **41**, 7892 (1990).

[24] N. Troullier and J. L. Martins, Phys. Rev. B **43**, 1993 (1991).

[25] J. S. Lin, Q. Qteish, M. C. Payne, and V. Heine, Phys. Rev. B **47**, 4174 (1993).

[26] D. C. Wallace, *Thermodynamics of Crystals*, 1st ed. (Wiley, New York, 1972).

[27] R. M. Wentzcovitch, J. L. Martins, and G. D. Price, Phys. Rev. Lett. **70**, 3947 (1993).

[28] R. M. Wentzcovitch, Phys. Rev. B **44**, 2358 (1991).

[29] O. H. Nielsen and R. Martin, Phys. Rev. B **32**, 3780 (1985).

[30] M. S. T. Bukowinski, Phys. Earth Planet. Inter. **14**, 333 (1977).

[31] M. S. T. Bukowinski, Geophys. Res. Lett. **12**, 536 (1985).

[32] P. Söderlind, J. A. Moriarty, and J. M. Willis, Phys. Rev. B **53**, 14063 (1996).

[33] O. K. Anderson, Phys. Rev. B **12**, 3060 (1975).

[34] S.-H. Wei and H. Krakauer, Phys. Rev. Lett. **55**, 1200 (1985).

[35] D. J. Singh, "*Planewaves, Pseudopotentials, and the LAPW Method*, 1st ed. (Kluwer Academic, Norwall, MA, 1994).

[36] L. Stixrude and R. E. Cohen, Nature (London) **364**, 613 (1993).

[37] H. J. Monkhurst and J. D. Pack, Phys. Rev. B **13**, 5188 (1976).

[38] T. S. Duffy, R. J. Hemley, and H. K. Mao, Phys. Rev. Lett. **74**, 1371 (1995).

[39] M. J. Mehl, R. E. Cohen, and H. Krakauer, J. Geophys. Res. **93**, 8009 (1988).

[40] R. E. Cohen, Am. Mineral. **76**, 733 (1991).

[41] R. E. Cohen, in *High-Pressure Research: Applications to Earth and Planetary Sciences*, Y. Syono and M. H. Manghnani, eds. (TERRAPUB, Tokyo, 1992), pp. 425–431.

[42] M. Sigalas, D. A. Papaconstantopoulos, and N. C. Bacalis, Phys. Rev. B **45**, 5777 (1992).

[43] D. M. Sherman, Earth Planet. Sci. Lett. **153**, 149 (1997).

[44] L. Stixrude, E. Wasserman, and R. E. Cohen, J. Geophys. Res. **102**, 24729 (1997).

[45] R. M. Wentzcovitch, D. A. Hugh-Jones, R. J. Angel, and G. D. Price, Phys. Chem. Minerals **22**, 453 (1995).

[46] R. M. Wentzcovitch, N. L. Ross, and G. D. Price, Phys. Earth Planet. Inter. **90**, 101 (1995).

[47] B. B. Karki, L. Stixrude, S. J. Clark, M. C. Warren, G. J. Ackland, and J. Crain, Am. Mineral. **82**, 635 (1997).

[48] R. E. Wentzcovitch and L. Stixrude, Am. Mineral. **82**, 663 (1997).

[49] R. M. Wentzcovitch, C. da Silva, J. R. Chelikowsky, and N. Binggeli, Phys. Rev. Lett. **80**, 2149 (1998).

[50] B. B. Karki, M. C. Warren, L. Stixrude, G. J. Ackland, and J. Crain, Phys. Rev. B **55**, 3465 (1997).

[51] J. Brodholt, A. Patel, and K. Retson, Amer. Mineral., **81**, 257 (1996).

[52] D. G. Isaak, Phys. Rev. B **47**, 7720 (1993).

[53] K. Kingma, H. K. Mao, and R. J. Hemley, High Pressure Res. **14**, 363 (1996).

[54] B. B. Karki, L. Stixrude, S. J. Clark, M. C. Warren, G. J. Ackland, and J. Crain, Am. Mineral. **82**, 51 (1997).

[55] A. Jephcoat, H. K. Mao, and P. M. Bell, J. Geophys. Res., **91**, 4677 (1986).

[56] H. K. Mao, Y. Wu, L. C. Chen, J. F. Shu, A. P. Jephcoat, J. Geophys. Res., **95**, 21737 (1990).

[57] C. da Silva, L. Stixrude, and R. M. Wentzcovitch, Geophys. Res. Lett. **24**, 1963 (1997).

[58] B. Kiefer, L. Stixrude, and R. M. Wentzcovitch, Geophys. Res. Lett. **24**, 2841 (1997).

[59] B. B. Karki, L. Stixrude, M. C. Warren, G. J. Ackland, and J. Crain, Geophys. Res. Lett. **24**, 3269 (1997).

[60] A. Yeganeh-Haeri, Phys. Earth Planet. Inter. **87**, 111 (1994).

[61] R. E. Cohen, I. I. Mazin, and D. G. Isaak, Science **275**, 654 (1997).

[62] C. Lee and X. Gonze, Phys. Rev. B **51**, 8610 (1995).

[63] L. Stixrude,
H. Krakauer
(1996).

[64] X. Gonze ar
10355 (199

[65] G. A. de Wi
D. Pobson,

[63] L. Stixrude, R. E. Cohen, R. C. Yu, and H. Krakauer, Am. Mineral. **81**, 1293 (1996).

[64] X. Gonze and C. Lee, Phys. Rev. B **55**, 10355 (1997).

[65] G. A. de Wijs, G. Kresse, L. Vocadlo, D. Pobson, D. Aifé, M. J. G. Slan and G. D. Price, Nature **392**, 805 (1998).

[66] B. B. Karki, and L. Stixrude, J. Geophys. Res. **104**, 13025 (1999).

[67] B. B. Karki, R. M. Wentzcovitch, S. de Gironcoli, S. Baroni, Science **286**, 1705 (1999).