Ab initio theory of phase transitions and thermoelasticity of minerals

ARTEM R. OGANOV*, JOHN P. BRODHOLT and G. DAVID PRICE

University College London, Gower Street, London WC1E 6BT, U.K.;
*e-mail: a.oganov@mat.ethz.ch

Introduction

Accurate quantum-mechanical simulations have significantly extended the current picture of the Earth and hold a great promise for the future of the Earth and planetary sciences. Studies of phase transitions, equations of state, elasticity and thermoelastic properties of the Earth-forming minerals are essential to geophysics. This chapter gives a basic background of the physics of the deep Earth and outlines the theory of phase transitions, equations of state, elasticity and thermoelastic properties. A particular emphasis is put on the principles of quantum-mechanical simulations and some recent results relevant to geophysics.

The importance of quantum-mechanical simulations is reflected by the award of the 1998 Nobel Prize in Chemistry to W. Kohn and J. Pople, who were among the pioneers of this field. Areas of application of such simulations are extremely diverse and include studies of the electronic structure, reactivity, catalysis, bulk and surface structure, prediction of materials structures and properties, especially at extreme conditions, calculation of phase diagrams and studies of phase transitions *etc.*

One of the most exciting areas of application of such simulations is the study of the Earth- (and planet-) forming minerals at the extreme conditions of the Earth's interior. One can accurately predict the structures, properties, and behaviour of minerals. This often reveals new aspects of mineral crystal chemistry and allows one to explain geophysical measurements and understand better how the Earth works as a planet. This chapter consists of five major parts – Part I: Brief geophysical introduction, Part II: Thermoelastic properties, Part III: Phase transitions, Part IV: Simulation methods and Part V: Examples from recent studies. Each part is supplied with a number of references for further reading.

Part I: A brief geophysical introduction

This section will be brief by necessity. For further reading we recommend the classical books by Anderson (1989) and Poirier (2000). The reader is also recommended to refer to the book by Verhoogen (1980). A scheme of the Earth's internal structure is shown in Figure 1. Before discussing it in more detail, we make some remarks on the Earth's chemistry.
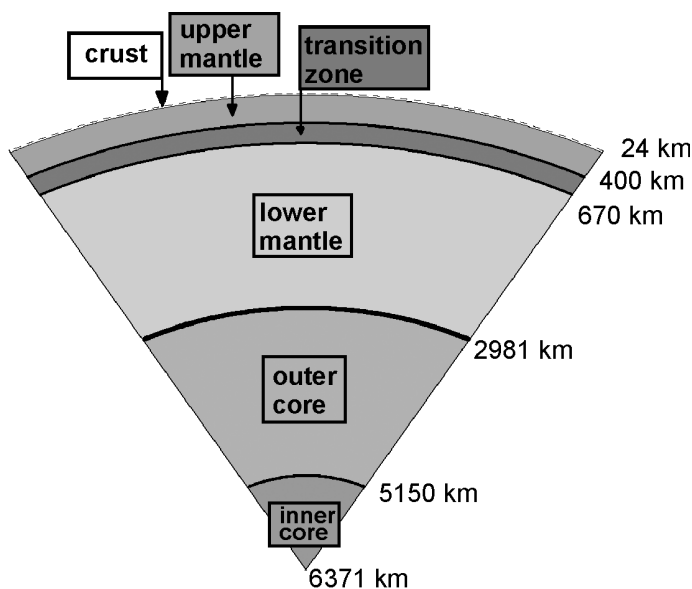
Fig. 1. Scheme of the Earth's internal structure. Mantle convection is responsible for plate tectonics, convection of the outer core – for the generation of the Earth's magnetic field.

The Earth is believed to be in many ways similar to the other terrestrial planets (Mars, Venus, Mercury). These planets are strongly chemically differentiated and have a dense metallic Fe-rich core and silicate mantle. Chondritic models are a good starting point for models of their bulk composition (see Anderson, 1989; Allègre *et al.*, 1995). The relative proportions of refractory elements (*e.g.*, Ca, Al, Sr, Ti, Ba, U, Th, Mg, Si) appear to be very similar in chondritic meteorites, terrestrial planets, Sun and Universe, but terrestrial planets are depleted in volatile elements (*e.g.*, H, noble gases, C, N, O, K, Na, Rb, Cs, S). Relative proportions of major Earth-forming elements in the Universe, Earth and its mantle and crust are given in Table 1.

The presence of two almost immiscible fractions (silicate/oxide crust and mantle and metallic core) results in strong partitioning of elements between them, with siderophiles (*e.g.*, Ni, platinoids, Au, Re) preferentially entering the core, and lithophiles (*e.g.*, Al, Mg, Ca, Na) mainly in the mantle and crust. Chalcophile elements (*e.g.*, Cu, Pb) are distributed between the core and mantle, but prefer the core. Some of the fractionation trends (*e.g.*, mantle and crust depleted in Fe and Ni, but enriched in Ca and Al) can be seen in Table 1.

It is worth noting that at different p/T conditions many elements change their behaviour: *e.g.*, K may become a chalcophile or siderophile element, and Si almost certainly acquires some siderophile properties at very high pressures. This would imply that these elements enter the core; fractionation of K into the core would imply an important source of radiogenic energy (radioactive ^{40}K isotope) within the core.

At high pressures of the lower mantle (> 24 GPa) $MgSiO_3$ with the perovskite structure becomes stable; geochemical models and phase equilibria indicate that this

Table 1. Abundances of some chemical elements (in atoms per 1 atom Si)

Element	The Universe [1]	Whole Earth [2]	Earth's crust [3]	Upper mantle [3]	Lower mantle [3]	Pyrolitic homogeneous mantle [4]
O	20.10	3.73	2.9	3.63	3.63	3.68
Na	0.06	0.06	0.12	0.03	2×10^{-3}	0.02
Mg	1.08	1.06	0.09	0.97	1.09	1.24
Al	0.08	0.09	0.36	0.17	0.06	0.12
Si	1	1	1	1	1	1
P	0.01	–	4×10^{-3}	6×10^{-4}	4×10^{-5}	4×10^{-4}
S	0.52	–	8×10^{-4}	6×10^{-4}	5×10^{-5}	2×10^{-3}
Ca	0.06	0.06	0.14	0.12	0.05	0.09
Cr	0.01	–	1×10^{-4}	5×10^{-3}	0.01	0.01
Fe	0.9	0.9	0.11	0.14	0.14	0.16
Ni	0.05	–	3×10^{-5}	3×10^{-3}	4×10^{-3}	3×10^{-5}

[1] Estimates of Anders & Ebihara (1982); [2] Simple model based on cosmic abundances (Anderson, 1989); [3] Recalculated from data of Anderson (1989); [4] Recalculated from Ringwood (1991).

mineral should be the most abundant mineral in the mantle – in fact, the most abundant mineral in the Earth. We shall consider this mineral in detail later on in this chapter. With the chondritic ratio Mg/Si = 1.06, the lower mantle would contain practically only the perovskite. However, geochemists and geophysicists prefer to describe the composition of the mantle by the pyrolitic model, which has the Mg/Si ratio of 1.24. This results in the mantle enriched in MgO. If the mantle is pyrolitic, the deficit of Si in it may be due to a large Si content in the Earth's core, or due to a non-chondritic bulk Earth composition. Compositions of the Earth's lower mantle and core are still poorly known.

The high density of the core and the presence of the magnetic field suggest that the core is metallic, and the elemental abundances indicate that it must be Fe-rich. It is interesting to mention an early hypothesis, proposed by W.H. Ramsey in 1949, that the Earth's dense core is not chemically different from the mantle and is made of the usual silicates of Mg, Al, Ca, and other elements – the idea being that at very high pressures of the Earth's core (~ 3 Mbar) these silicates will transform into ultradense modifications and (in order to explain the magnetic field) will become metallic. Geochemically this would mean no chemical stratification and invalidity of the chondritic model. Early shock wave experiments of Altshuler's group in Russia (see Zharkov & Kalinin, 1968), subsequently reproduced in other groups and reinforced by theoretical calculations (*e.g.*, Cohen, 1991; Bukowinski, 1994), definitively refuted this hypothesis. Silicate minerals remain insulating and never adopt superdense structures or become metallic at the Earth's core conditions. The core must be Fe-rich.

Detailed information on the density and size of the dense core is obtained from detailed geophysical observations and global seismological models, the most popular of which is PREM (Preliminary Reference Earth Model – Dziewonski & Anderson, 1981). The following regions, separated by major first-order seismic boundaries, are distinguished within the Earth: crust (0–24.4 km, but the depth strongly varies geographically); mantle (24.4–2891 km): upper mantle (24.4–400 km), transition zone (400–670 km), lower mantle (670–2891 km); core (2891–6371 km): outer core

(2891–5150 km) and inner core (5150–6371 km). Figure 1 shows a scheme of the Earth's structure. Seismic discontinuities are due to the compositional changes (24.4 km, 2891 km) and phase transitions (400 km, 670 km, 5150 km). Apart from these major first-order discontinuities, there are weaker discontinuities, whose origin is often poorly understood.

One of the better known minor discontinuities is the 520 km discontinuity (Shearer, 1990) related to the wadsleyite–ringwoodite transition. Although this transition should be present globally, the discontinuity is small and only locally observable (Deuss & Woodhouse, 2001); it is not included in current seismological models. In principle, even global discontinuities may become seismically invisible if they are spread over a large depth interval.

Apart from PREM, there are other global seismological models, *e.g.*, ak135 (Kennett *et al.*, 1995), which generally agree with PREM within 1–2% on seismic velocities (Fig. 2). The parameters given by these models as a function of depth are: pressure (p), compressional and shear seismic velocities (V_p and V_s), density (ρ), acceleration due to gravity (g), adiabatic bulk and shear moduli (K_S and G), seismic parameter ($\Phi = K_S/\rho$), Poisson ratio (ν), attenuation Q-factors for the compressional and shear velocities (Q_p and Q_s), and Bullen parameter (η). Some of these profiles are shown in Figures 3 and 4.

Pressure distribution can be calculated straightforwardly from the density distribution using the following equation:

$$\frac{dp}{dr} = -4\pi G\rho r^{-2}\int_0^r \rho r^2 dr = -g\rho \ , \tag{1}$$

where G, r, and g are the gravitational constant, radius, and acceleration due to gravity.
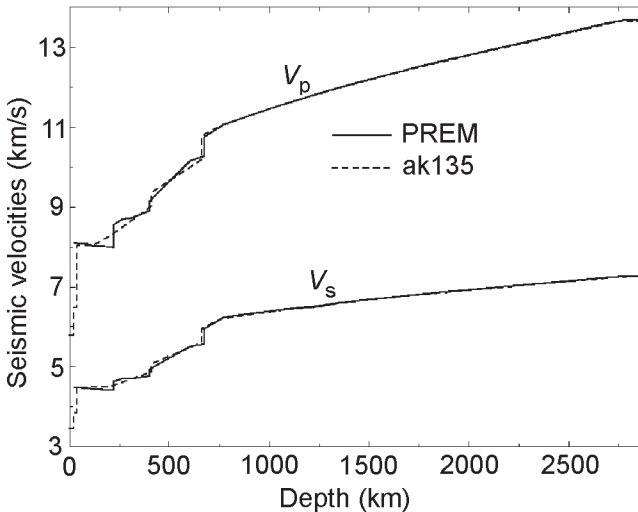

Fig. 2. Comparison of the PREM and ak135 profiles of seismic wave velocities in the Earth's mantle and crust.
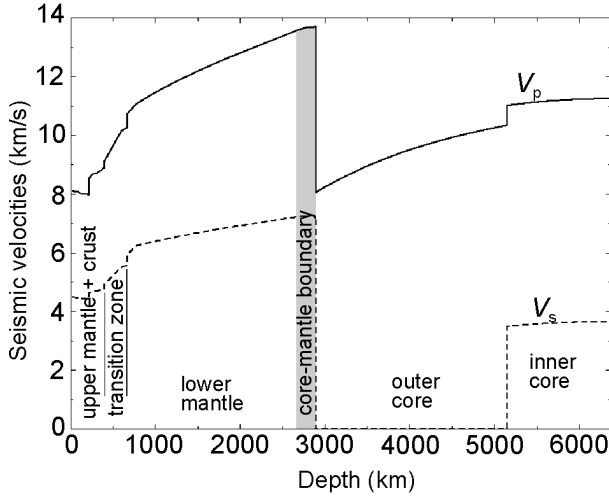
Fig. 3. PREM seismic velocity profiles. The major regions of the Earth are specified. The core–mantle boundary region D″ (grey shading) is the major seismic boundary in the Earth. Shear waves do not propagate in the outer core, which leads to the conclusion that the outer core is molten. The inner core is solid.

Derivation of a temperature profile is much more complicated, and currently there is no commonly accepted thermal model of the Earth. Both temperature and composition can in principle be determined from seismological models given some knowledge of the properties of the Earth's materials. The famous Birch's law (Poirier, 2000) states that compressional seismic velocities depend only on the composition and density of the material:

$$V_p = a(\bar{M}) + b\rho, \tag{2}$$

where $\bar{M}$ is the average atomic mass, a and b constants, ρ the density. This empirical law works surprisingly well. For the outer core it suggests $\bar{M} = 49.3$ (Poirier, 2000), supporting the Fe-rich composition and presence of light elements (the atomic mass of Fe is 55.8).
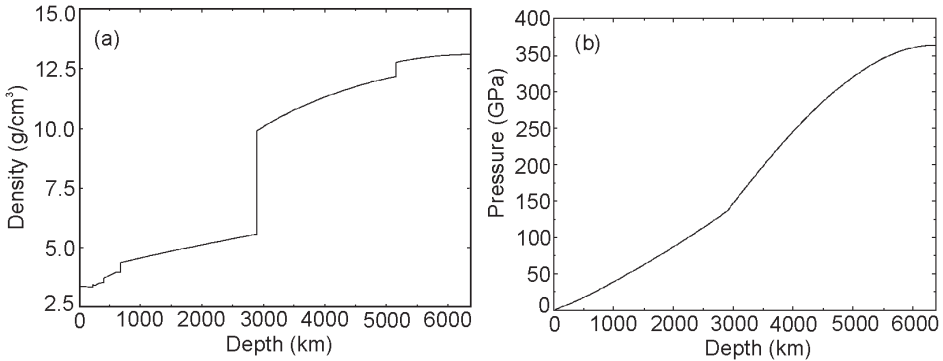

Fig. 4. PREM density profile *(a)* and pressure profile *(b)*. Pressure is continuous, but there is a large kink in its slope at the core–mantle boundary due to the large density jump occurring there.

The latest *ab initio* simulations (Vočadlo *et al.*, 2000) and experiments indicate that Fe is hexagonal close packed in the inner core. The inner core is highly seismically anisotropic, with the fastest direction of seismic waves along the axis of the Earth's rotation. This anisotropy implies a high degree of crystal alignment, whose cause is unknown. Using *ab initio* molecular dynamics simulations, Alfè *et al.* (1999) calculated the melting curve of pure Fe and concluded that for a pure Fe core the temperature at the inner–outer core boundary is 6700 ± 600 K. Using the density jump at this boundary as a constraint, Alfè *et al.* (2002) were able to put forward a compositional model for the Earth's core (inner core: 8.5% Si + S and 0.2% O; outer core: 10% Si + S and 8% O); remarkably, this composition has $\bar{M} = 49.38$. The temperature at which these compositions are at equilibrium is 5600 K, a more rigorous estimate of the temperature at the inner-outer core boundary.

Vigorous convection of the outer core, like a giant dynamo-machine, generates the Earth's magnetic field. By constructing an adiabatic temperature profile for the outer core, Alfè *et al.* (private communication) obtained $T \approx 4200$ K at the core–mantle boundary.

The Earth's mantle consists mainly of Mg-silicates with $\bar{M} = 21.3$ from Birch's law (Poirier, 2000). Compared to $\bar{M} = 20.12$ for $MgSiO_3$, 20.15 for MgO, and 20.13 for Mg_2SiO_4, this implies an $\sim 10\%$ substitution of Mg by Fe. Fe is mainly in the form of Fe^{2+}, which is in a high-spin state at low pressures, but may transform into low-spin non-magnetic Fe^{2+} at high pressures. This 'magnetic collapse' has attracted much attention in both theoretical (Sherman, 1991; Isaak *et al.*, 1993; Cohen *et al.*, 1997; Cohen, 1999; Fang *et al.*, 1999) and experimental (Pasternak *et al.*, 1997; Badro *et al.*, 1999) literature, but it remains highly unclear whether the high-spin-to-low-spin transition in Fe^{2+} ions occurs in the mantle.

Phase transitions of Mg-silicates determine the seismic structure of the mantle (Fig. 5; Helffrich, 2000; Chudinovskikh & Boehler, 2001). Whether there is any compositional stratification is an open question. Pyrolitic lower mantle would consist of $(Mg,Fe)SiO_3$ perovskite (~ 75 vol%), magnesiowüstite $(Mg,Fe)O$ (~ 20 vol%), and $CaSiO_3$ perovskite (~ 5 vol%).

Knowledge of the thermoelastic properties of $(Mg,Fe)SiO_3$ perovskite is crucial for constructing thermal and compositional models of the mantle. Stixrude *et al.* (1992), using early measurements of thermoelastic parameters of $(Mg,Fe)SiO_3$ perovskite (Knittle & Jeanloz, 1986; Mao *et al.*, 1991) with an unusually high thermal expansion coefficient, arrived at the conclusion that the lower mantle must be $\sim 100\%$ $(Mg,Fe)SiO_3$ perovskite. This would imply a compositional difference between the lower mantle and pyrolitic upper mantle, absence of chemical mixing between the upper and lower mantle and layered (*i.e.* upper and lower mantle convecting separately), rather than whole-mantle, convection. The bulk lower mantle composition would then be chondritic (*e.g.*, in Mg/Si ratio) rather than pyrolitic. Later measurements of thermal expansion of this mineral (Wang *et al.*, 1994; Funamori *et al.*, 1996; Fiquet *et al.*, 2000) yielded much lower values. To have a pure perovskite lower mantle with these values, one must have temperatures that are unacceptably high. Whole-mantle or intermediate convection models are currently preferred, being also consistent with seismic tomography images. However, geochemical studies provide an argument in favour of the layered mantle
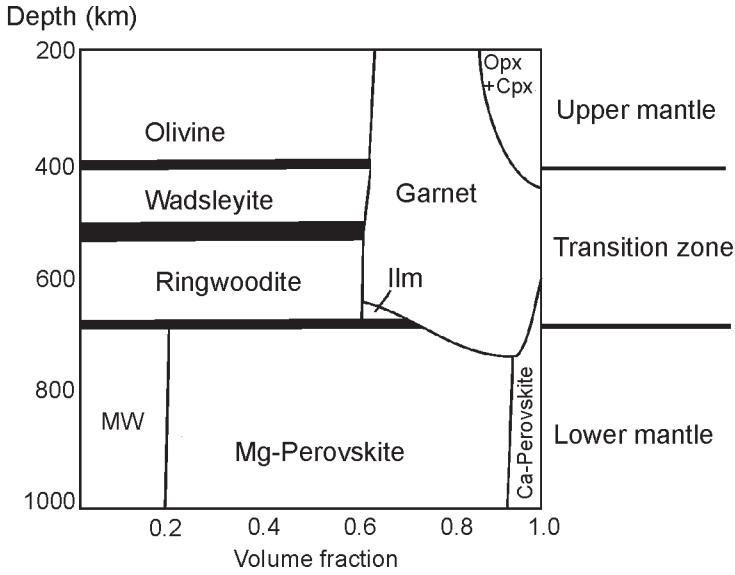
Fig. 5. Schematic phase relations in pyrolite (after Ringwood, 1991). Opx and Cpx are ortho- and clinopyroxene, Ilm is $MgSiO_3$ ilmenite (akimotoite), MW is magnesiowüstite $(Mg,Fe)O$, 'Mg-perovskite' and 'Ca-perovskite' stand for $MgSiO_3$ and $CaSiO_3$ perovskites.

convection, indicating two chemically distinct sources of mantle magmas. The style of mantle convection remains an open question (see Poirier, 2000).

The temperature gradient in a convecting system (*e.g.*, mantle) must be adiabatic or higher. The adiabatic temperature gradient can be calculated from the thermodynamic equality:

$$\left(\frac{\partial T}{\partial p}\right)_S = \frac{\gamma T}{K_S},$$ (3)

where γ and K_S are the Grüneisen parameter and the adiabatic bulk modulus, respectively.

A fundamental geophysical relation (see Jackson, 1998) exists:

$$1 - g^{-1}\frac{d\Phi}{dr} = \left(\frac{\partial K_S}{\partial p}\right)_S + \frac{\tau\alpha\Phi}{g}\left(1 + \frac{\left(\frac{\partial K_S}{\partial T}\right)_p}{\alpha K_S}\right),$$ (4)

which describes self-compression of a chemically homogeneous layer characterised by a superadiabatic temperature gradient τ; α is the thermal expansion coefficient. If the temperature distribution is adiabatic, the following relation must be obeyed:

$$\eta = \left(\frac{\partial K_S}{\partial p}\right)_S + g^{-1}\frac{d\Phi}{dr} = \Phi\frac{d\rho}{dp} = 1,$$ (5)

where η is known as the Bullen parameter. If its values deviate from 1, it indicates that either the temperature gradient is non-adiabatic, or the chemical composition varies with depth.

In a chemically homogeneous system, the superadiabatic gradient τ is related to the deviations of η from unity:

$$\tau = \frac{g}{\alpha}\left(\Phi^{-1} - \frac{d\rho}{dp}\right) = \frac{g}{\alpha\Phi}\left(1 - \eta\right). \tag{6}$$

In PREM, $\eta = 0.99 \pm 0.01$ throughout the lower mantle, supporting the view that the lower mantle is adiabatic and chemically homogeneous. However, in the ak135 model $\eta = 0.94 \pm 0.02$, which implies very large superadiabatic gradients, 0.3–0.7 K/km (see Jackson, 1998).

According to Verhoogen (1980), the core–mantle boundary layer D″ is essentially a thermal boundary layer with a temperature jump of $\sim$ 1200 K. From seismology, this layer is highly variable in thickness, highly heterogeneous and elastically anisotropic, and has very small or in some places even negative velocity gradients. An intriguing possibility is the partial melting of this region. The solidus of the pyrolite mantle was determined experimentally by Zerr *et al.* (1998); at the core–mantle boundary melting would start at $\sim$ 4300 K, which is similar to the temperatures of the core near the core–mantle boundary. The presence of a melt would imply high electrical conductivity due to ionic diffusivity. The electrical conductivity of the lower mantle is indeed high ($\sim$ 1–10 S/m on average – Xu *et al.*, 2000).

Knittle & Jeanloz (1991) consider the D″ layer as a chemical reaction zone between the core and mantle. They experimentally observed a reaction, which can be schematically written as follows:

$$(Mg,Fe)SiO_3 + Fe = MgSiO_3 + SiO_2 + FeO + FeSi. \tag{7}$$
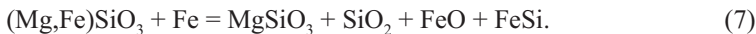
Iron oxide and silicide at the high pressure of the core–mantle boundary are metallic and should be soluble in the core. The reaction of Knittle & Jeanloz (1991) might drive Fe (as well as Si and O) from the mantle into the core. This opens an interesting possibility of the still growing core.

Another interesting question is the nature of minor seismic discontinuities in the lower mantle. *E.g.*, the locally observed 1200 km discontinuity[1] has been attributed to a tetragonal-to-cubic transition in $CaSiO_3$ perovskite (Stixrude *et al.*, 1996; Chizmeshya *et al.*, 1996), found in linear response all-electron LAPW calculations. More approximate pseudopotential calculations of Karki (1997) and Warren *et al.* (1998), however, did not find this transition, and the stable phase of $CaSiO_3$ perovskite in their simulations was always cubic. Improved experimental and theoretical techniques can or soon will be able to resolve such questions.

Our understanding of the mantle mineralogy can be greatly improved by studies of mantle inclusions – *e.g.*, Harte *et al.* (1999) found several lower mantle minerals

---

[1] Vinnik *et al.* (1998); Le Stunff *et al.* (1995) suggested that discontinuities at 785–950 km and 1200 km might be global.

(including $MgSiO_3$ with up to 10% Al_2O_3) in inclusions in diamonds. Most inclusions studied so far have upper mantle or transition zone origin, however.

The transition zone (400–670 km) is quite diverse mineralogically, and might possess exotic properties. It may host large amounts of water: both wadsleyite and ringwoodite can contain up to 2–3 wt% H_2O (see Fiquet, 2001 and references therein). It has also been suggested (Angel *et al.*, 1996 and references therein) that five-coordinate Si, which is unusual for inorganic compounds, can play an important role in the transition zone, determining its transport properties.

The upper mantle consists predominantly of olivine, garnet, and pyroxenes. Its major features include: (i) ultralow velocity zone at variable depths, roughly between 50–100 km and 220 km (Anderson, 1989), (ii) a seismic discontinuity (Lehmann discontinuity) at 220 km depth, (iii) strong elastic anisotropy above 220 km, and (iv) compositional heterogeneity in the upper 150 km (see Ringwood, 1991). The Lehmann discontinuity is, possibly, due to the *Pbca*–*C*2/*c* transition in pyroxenes (see Mendelssohn & Price, 1997). The ultralow velocity zone is interpreted as a region of partial melting and low viscosity (asthenosphere) beneath the rigid lithosphere. Anisotropy in this region is a consequence of preferred orientation of crystals caused by convective flow.

A comprehensive review of mantle mineralogy can be found in Fiquet (2001). Overall, mantle mineralogy is dominated by a handful of mineral species with quite dense structures. Very large (*e.g.*, Na, K, Ca, Sr, Cl, U, Th) and very small (*e.g.*, Li, Be, B) atoms cannot enter these minerals, and concentrate in mantle magmas and fluids, rising to the surface of the Earth and forming its crust. This is why the rich mineral list of the crust is dominated by mantle incompatible elements. The most abundant minerals of the crust are feldspars $(Na,K,Ca)(Al,Si)_4O_8$.

The Earth's mantle is convecting at the velocities of several mm/year. These velocities are highly variable, and may be a few times lower in the lower mantle due to its higher viscosity. Microscopically, this solid state convection can occur by dislocational or diffusional creep.

It is important to understand the 3D thermal structure of the mantle underlying this convection, in particular how large are the thermal anomalies associated with the 'cold slabs' and 'hot spots'. Cold slabs may have a different mineralogy from the rest of the mantle, because of the lower temperatures and different bulk chemical composition. An aluminous phase (possibly, $MgAl_2O_4$ in the $CaTi_2O_4$ or $CaFe_2O_4$ structures – Kesson *et al.*, 1994, 1998) would be present in the slabs. H_2O, liberated from hydrous silicates at high pressures, might form ice VII in sufficiently cold slabs (*e.g.*, at 700 K and 15 GPa – Bina & Navrotsky, 2000).

Studying the topography of major seismic discontinuities, it is possible to get some ideas on lateral temperature variations (Helffrich, 2000), knowing the dp/dT slopes of the corresponding phase transitions. Most slabs turn out to be ~ 400–700 K colder than the average mantle at 660 km; the Tonga slab seems to be ~ 1200 K colder. The temperature anomaly below the Iceland hot spot is +180 K at 660 km depth (Helffrich, 2000). It is possible to extract the same information as well as other important characteristics of the mantle from seismic tomography.

In seismic tomography, one inverts numerous seismological measurements to obtain *three-dimensional* distributions of seismic wave velocities. It is common to represent results in terms of perturbations of seismic velocities relative to the average velocity at each depth: *e.g.*, $\Delta V_S / V_S$. Most recent tomography maps have similar qualitative features (locations of 'cold' and 'hot' regions) and roughly similar magnitudes of the velocity perturbations. Perhaps, the most reliable global seismic tomography maps currently available are those of Masters *et al.* (2000).

In the first approximation, low velocities can be attributed to high temperatures, and high velocities to low temperatures. There is a remarkable correlation between surface tectonics and tomographic images down to the core–mantle boundary. Most cold regions (probably, representing the slabs) penetrate the 670 km boundary and even seem to reach the core–mantle boundary region (Masters *et al.*, 2000), although in some tomographic images (Kennett *et al.*, 1998) most slabs seem to disappear somewhere between 1000 km and 2000 km depths.

From PREM, the relative variation of shear and compressional seismic wave velocities due to pressure alone:

$$R_{\mathrm{T}} = \left(\frac{\partial \ln V_{\mathrm{s}}}{\partial \ln V_{\mathrm{p}}} \right)_{\mathrm{T}} \tag{8}$$

is 0.7 (Anderson, 1989). A similarly defined parameter,

$$R_{\mathrm{p}} = \left(\frac{\partial \ln V_{\mathrm{s}}}{\partial \ln V_{\mathrm{p}}} \right)_{\mathrm{p}}, \tag{9}$$

measuring the same ratio, but due to temperature effects alone, is much larger: it increases from 1.7 to 2.6 between the depths of 1000 km and 2000 km according to seismic tomography (Robertson & Woodhouse, 1996).

The large difference between R_{T} and R_{p} has been a puzzle for geophysicists over the last 15 years. The seismological values of R_{T} are well reproduced in the laboratory (see Anderson, 1989) and in theoretical simulations (Oganov *et al.*, 2001b; Parker & Wall, 1991); it is the large seismologically measured value of R_{p} that needs to be explained for the lower mantle. Three possible explanations exist: (i) intrinsic anharmonicity (Anderson, 1989), (ii) anelastic effects (Karato, 1993) and (iii) large scale compositional heterogeneity (*e.g.*, Masters *et al.*, 2000). The first two effects must be present; the question is how important these effects are.

Chemical heterogeneity is believed to be very important only in the D″ layer, which can be a chemical reaction zone (Knittle & Jeanloz, 1991). In the next section we shall consider in detail elastic properties of crystals, equations of state and thermodynamic properties. This theory plays a central role in the interpretation of seismological data.

Part II: Thermoelastic properties

Thermodynamics

Thermodynamic properties are, perhaps, the most important properties of a crystal – they define its stability field; their derivatives with respect to pressure, temperature, and volume describe the behaviour of the crystal at changing conditions, its equation of state (EOS) and response functions such as the elastic constants and thermal expansion.

In thermodynamic theory of condensed matter, a fundamental role is played by the partition function:

$$Z = \sum_n e^{-E_n/k_B T},\tag{10}$$

where the summation is carried out over all discrete energy levels of the system. Once Z is known, all thermodynamic properties can be obtained straightforwardly, *e.g.* the Helmholtz free energy:

$$F = E_0 - k_B T \ln Z,\tag{11}$$

where E_0 is the ground state energy (at 0 K) including the energy of zero-point motion.

However, it is extremely difficult to obtain all the energy levels experimentally or theoretically, and their number is overwhelmingly large for solids. The harmonic approximation gives a first approximation to the distribution of the energy levels E_n in solids; it is usually accurate for the most populated lowest excited vibrational levels. The key concept here is that of a non-interacting (ideal) gas of quasiparticles called phonons, and the key quantity is the phonon density of states, $g(\omega)$. Anharmonic effects can often be included as corrections to the harmonic results.

Harmonic approximation

In this approximation, it is assumed that all interatomic interactions are analogous to perfectly elastic springs. According to quantum mechanics, the energy levels of a harmonic oscillator are discrete; the solution of the Schrödinger equation yields:

$$E_n = \hbar(½ + n)\omega,\tag{12}$$

where n is the quantum number and ω the frequency of the vibration. For $n = 2, 3, \ldots$ we have doubly, triply, … excited vibrations – such vibrations are called overtones, and due to high energies such excited states have low populations p_n increasing with temperature according to the Boltzmann statistics: $p_n \sim e^{-E_n/k_B T}$. Energy levels in real crystals are well described by Equation 12 only for the lowest quantum numbers n, but these represent the most populated, and thus the most important vibrational excitations.

In the harmonic model all atoms vibrate around their equilibrium positions. Diffusion becomes impossible because of the infinite energy required for removing an atom or for an atom to become unbound to its original position. Such high-temperature phenomena as melting and diffusion in real materials are essentially anharmonic and cannot be accounted for by the harmonic approximation. The harmonic potential

becomes inaccurate at large displacements and higher-order overtones are poorly described by the harmonic approximation. These large displacements become accessible only at high temperatures, and this is where the harmonic approximation breaks down.

The harmonic vibrational energy of a crystal can be determined via the phonon frequencies ω and temperature T. Each phonon contributes to the vibrational energy according to the following formula:

$$E_{\text{vib}}(\omega,T) = \frac{1}{2}\hbar\omega + \frac{\hbar\omega}{\exp\left(\dfrac{\hbar\omega}{k_{\text{B}}T}\right)-1}, \tag{13}$$

where $\hbar$ is the Planck constant and k_{B} the Boltzmann constant. The first, temperature independent, term in Equation 13 is the zero-point energy originating from quantum motion of atoms; this contribution dominates E_{vib} at low temperatures. The second, temperature dependent term (thermal energy, or heat content) represents the energy change upon heating from 0 K to the temperature T. Differentiation of Equation 13 with respect to T produces the heat capacity of a harmonic oscillator:

$$C_{\text{V,vib}}(\omega,T) = k_{\text{B}}\left(\frac{\hbar\omega}{k_{\text{B}}T}\right)^2 \frac{\exp\left(\dfrac{\hbar\omega}{k_{\text{B}}T}\right)}{\left(\exp\left(\dfrac{\hbar\omega}{k_{\text{B}}T}\right)-1\right)^2}, \tag{14}$$

and it is easy to obtain its vibrational entropy:

$$S_{\text{vib}}(\omega,T) = -k_{\text{B}}\ln\left[1-\exp\left(-\frac{\hbar\omega}{k_{\text{B}}T}\right)\right] + \frac{1}{T}\frac{\hbar\omega}{\exp\left(\dfrac{\hbar\omega}{k_{\text{B}}T}\right)-1} \tag{15}$$

and the Helmholtz free energy:

$$F_{\text{vib}}(\omega,T) = \frac{1}{2}\hbar\omega + k_{\text{B}}T\ln\left[1-\exp\left(-\frac{\hbar\omega}{k_{\text{B}}T}\right)\right]. \tag{16}$$

In a crystal with n atoms in the unit cell, there are $3n$ different normal modes, whose frequencies form a spectrum called the phonon spectrum, or phonon density of states $g(\omega)$. As in the harmonic approximation the normal modes are independent and non-interacting, all thermodynamic properties can be obtained as a natural generalisation of Equations 13–16 by integrating over frequencies, *e.g.*:

$$C_{\text{V}}(T) = \int_0^{\omega_{\text{max}}} C_{\text{V}}(\omega,T)g(\omega)d\omega = k_{\text{B}}\int_0^{\omega_{\text{max}}} g(\omega)\left(\frac{\hbar\omega}{k_{\text{B}}T}\right)^2 \frac{\exp\left(\dfrac{\hbar\omega}{k_{\text{B}}T}\right)}{\left(\exp\left(\dfrac{\hbar\omega}{k_{\text{B}}T}\right)-1\right)^2} d\omega. \tag{17}$$

Once the density of states $g(\omega)$ is known, all thermodynamic functions can be easily computed. IR or Raman spectra probe only some of the vibrations with wavevectors close to zero (*e.g.*, having the wavelengths of the scale of the whole crystal; however, most vibrations have shorter wavelengths), and therefore do not give the full information on the phonon spectrum. The $g(\omega)$ can be determined experimentally by inelastic neutron scattering (*e.g.*, Dove, 1993; Choudhury *et al.,* 2002 [this volume]), or theoretically using lattice dynamics or molecular dynamics simulations.

Often reasonable approximations to $g(\omega)$ can be obtained using the IR and/or Raman spectra and/or the elastic properties. In the Debye model, the phonon spectrum is described by a parabolic function truncated at some maximum frequency $\omega_D = \dfrac{k\theta_D}{\hbar}$:

$$g(\omega) = 9n\left(\frac{\hbar}{k_B\theta_D}\right)^2 \omega^2, \tag{18}$$

where n is the number of atoms in the unit cell of volume V_0, and θ_D is the characteristic Debye temperature, the only parameter in this model. This density of states is based on the assumption that all modes obey the same (linear) dispersion law as the long wavelength acoustic modes, related to the elastic constants of the solid.

With Equation 18 for $g(\omega)$, the thermodynamic functions take the following forms:

$$E_{vib} = \frac{9}{8}k_B n\theta_D + 3k_B bTD\left(\frac{\theta_D}{T}\right), \tag{19}$$

$$C_V(T) = \left(\frac{dE_{vib}}{dT}\right)_V = 3k_B n\left[4D\frac{\theta_D}{T} - \frac{3\theta_D/T}{e^{\theta_D/T}-1}\right], \tag{20}$$

$$S(T) = \int_0^T \frac{C_p}{T}dT = k_B n\left[4D\left(\frac{\theta_D}{T}\right) - 3\ln\left(1 - e^{\theta_D/T}\right)\right], \tag{21}$$

where $\quad D(x) = \dfrac{3}{x^3}\int_0^x \dfrac{x^3 dx}{e^x - 1}, \quad x = \dfrac{\theta_D}{T}.$

The Debye temperature is determined by the elastic properties or, more precisely, the average sound velocity $\langle v \rangle$ in a crystal:

$$\theta_D = \frac{\hbar}{k_B}\left(\frac{6\pi^2 n}{V_0}\right)^{\frac{1}{3}}\langle v \rangle. \tag{22}$$

The mean sound velocity can be accurately calculated from the elastic constant tensor (Robie & Edwards, 1966) or using an approximate formula:

$$\langle v \rangle = \left(\frac{1}{V_p^3} + \frac{2}{V_s^3}\right)^{-\frac{1}{3}}, \tag{23}$$

where V_p and V_s are the longitudinal and transverse sound velocities, respectively.

The Debye model is very effective in application to chemically and structurally simple crystals, but fails in more complicated cases. Kieffer (1979a) gave a comprehensive critical review of the Debye model and in the same seminal cycle of papers proposed a new model (Kieffer, 1979b, 1979c, 1980, 1982). This model is based on the following main points: (i) the acoustic modes have a sinusoidal dispersion law $\omega(\mathbf{k})$ (not linear, as assumed in the Debye model) and are described by three different densities of states, (ii) optic modes are described by separate densities of states: monochromatic ('Einstein') peaks for high-frequency optic modes and by a continuum distribution for the low-frequency optic modes. In comparison with the Debye model this model has a much greater accuracy, but at the cost of a number of additional parameters.

In the Agoshkov model (Agoshkov, 1985; Agoshkov *et al.*, 1994), all the acoustic and low-frequency optic modes are described by a single Debye-like parabolic density of states $g(\omega)$, while the high-frequency modes, like in the Kieffer model, are described as Einstein oscillators. Komada & Westrum (1997) proposed another model, which, like the Debye model, has only one parameter.

Shortcomings of the harmonic approximation

The harmonic approximation greatly simplifies calculations, but has fundamental flaws. First, within this approximation, all vibrational modes are independent – there is no exchange of energy and momentum between them. Non-interacting phonons can freely travel within the crystal, leading to the infinite mean free path of the phonons and infinite thermal conductivity of the harmonic crystal. In real crystals, thermal conductivity is finite due to phonon–phonon collisions, scattering on defects, and finite size of samples.

Diffusion and melting cannot be explained within the harmonic approximation. The same is true for displacive phase transitions – although the harmonic approximation can indicate such a transition by showing a soft mode, no properties can be calculated for the dynamically disordered phase.

In the harmonic approximation there is no thermal expansion. Related to this is the equality $C_V = C_p$, whereas experiment indicates $C_V < C_p$.

From general thermodynamics,

$$C_p = C_V(1+\alpha^2 K_T V/C_V), \tag{24}$$

where α is the thermal expansion coefficient, K_T isothermal bulk modulus, and V molar volume. When $\alpha = 0$, $C_V = C_p$.

Also, experiment indicates (Gillet *et al.*, 1999) that at high temperatures the isochoric heat capacity can differ from $3nR$, while in the harmonic approximation $3nR$ is the high-temperature limit (the Dulong–Petit limit). Similarly, the vibrational energy can exceed the harmonic quota of $3k_BT$ per atom (or k_BT per mode). It can be shown (Oganov, 2002) that for the classical anharmonic oscillator the total vibrational energy is:

$$E_{vib} = k_BT - \frac{1}{6}\frac{\partial^3 E_{pot}}{\partial x^3}\langle x^3\rangle - \frac{1}{4}\frac{\partial^4 E_{pot}}{\partial x^4}\langle x^4\rangle - \ldots \quad, \tag{25}$$

where $E_{pot}(x)$ is the potential energy as a function of displacement x. Equation 25 differs

from the high-temperature harmonic expression $E_{vib} = k_BT$. Obviously, the second and third terms in Equation 25 would be absent in the harmonic approximation or in the case of small displacements, and we would recover the Dulong–Petit value of k_BT for the energy per mode at high temperatures.

Perhaps, the most spectacular failure of the harmonic approximation is its wrong conclusion that vibrational frequencies do not depend on temperature and pressure. To account for such experimentally observed strong dependences, one has to use more realistic approximations. The first approximation, combining the simplicity of the harmonic approximation while correcting its most serious errors is the quasiharmonic approximation.

Quasiharmonic approximation (QHA)

In this approximation, it is assumed that the solid behaves like a harmonic solid at each volume, but the phonon frequencies depend on volume. It is assumed that they depend only on volume – *i.e.*, heating at constant volume does not change the frequencies.

In the QHA phonon modes are still independent and non-interacting. Thermodynamic functions at constant volume, as before, are given by Equation 17 for the heat capacity C_V and analogous relations for other properties. C_V still tends to $3nR$ at high temperatures. Melting, diffusion, and dynamically disordered phases are beyond the scope of this approximation, which breaks down at high temperatures. Thermal conductivity is still infinite.

However crude, this approximation corrects the worst errors of the harmonic approximation. Introducing a volume dependence of the frequencies is enough to create non-zero thermal expansion and account for $C_V < C_p$ (*cf.* Eqn. 24). This is the first approximation for the thermal EOS of solids, which can be effectively used in conjunction with realistic interatomic potentials (Parker & Price, 1989; Kantorovich, 1995; Gale, 1998) or quantum-mechanical approaches (Baroni *et al.*, 1987, 2001; Karki *et al.*, 1999, 2000a, 2000b).

Beyond the QHA: Intrinsic anharmonicity

At temperatures roughly below $\frac{1}{2} - \frac{2}{3}$ of the melting temperature, the QHA is quite accurate. Only at higher temperatures do its errors become large. All the effects beyond the QHA are termed *intrinsic anharmonic effects*. To the lowest order, the intrinsic anharmonic contribution to the free energy F_a is proportional to T^2 (Zharkov & Kalinin, 1968; Gillet *et al.*, 1999) at high temperatures:

$$F(V,T) = F_{qha}(V,T) + 3nk_B[\tfrac{1}{2}a(V)T^2 + \ldots], \tag{26}$$

from which

$$
\begin{aligned}
&E(V,T) = E_{qha}(V,T) - 3nk_B[\tfrac{1}{2}a(V)T^2 + \ldots], \\
&S(V,T) = S_{qha}(V,T) - 3nk_B[a(V)T + \ldots], \\
&C_V(T) = C_{V,qha}(T) - 3nk_B[a(V)T + \ldots].
\end{aligned}
\tag{27}
$$

One important consequence of intrinsic anharmonicity is that vibrational frequencies depend not only on volume; identical volume changes, produced by changing pressure and temperature, result in different changes in the frequencies. This is clearly demonstrated by precise experiments (see, *e.g.*, Gillet *et al.*, 2000, 1999) and computer simulations (Winkler & Dove, 1992). The intrinsic anharmonicity parameter of an *i*-th phonon mode at a wavevector **k** is given by:

$$a_{i\mathbf{k}} = \left(\frac{\partial \ln \omega_{i\mathbf{k}}}{\partial T}\right)_V = \alpha(\gamma_{i\mathbf{k}}^{\mathrm{T}} - \gamma_{i\mathbf{k}}^{\mathrm{p}}), \qquad (28)$$

where α is the thermal expansion coefficient, and $\gamma_{i\mathbf{k}}^{\mathrm{T}}$ and $\gamma_{i\mathbf{k}}^{\mathrm{p}}$ are constant-temperature and constant-pressure mode Grüneisen parameters (see below), respectively. Intrinsic anharmonic parameters $a_{i\mathbf{k}}$ (Eqn. 28) are equal to the coefficients a previously used in Equations 26 and 27 for anharmonic thermodynamic properties (a expresses the total anharmonicity, while $a_{i\mathbf{k}}$ are the mode anharmonic parameters). Since these parameters are measurable, the anharmonic corrections can be directly evaluated from experiment.

The full anharmonic theory of thermodynamic properties of crystals is very complicated (see Wallace, 1998). Computationally, all intrinsic anharmonic effects can be fully accounted for by the use of Monte Carlo or molecular dynamics simulations (Allen & Tildesley, 1987): these methods involve a full sampling of the energy hypersurface without any assumptions regarding its shape or smallness of atomic displacements; these methods are equally applicable to liquids and gases.

Equation of state (EOS)

General thermodynamic formulation

An equation of state (*i.e.*, the *p–V–T* relationship) is one of the most important characteristics of a material. Indeed, accurate EOSs of minerals are necessary for the interpretation of seismological observations.

Generally, from thermodynamics:

$$p = -\left(\frac{\partial F}{\partial V}\right)_T \qquad\qquad \text{isothermal EOS,}$$

$$p = -\left(\frac{\partial E}{\partial V}\right)_S \qquad\qquad \text{adiabatic EOS.}$$

Explicit analytical EOS can only be written for an ideal gas (where interatomic interactions are absent and there are no problems in calculating exactly the partition function and thermodynamic properties). For solids and liquids interatomic interactions are essential, and all analytical EOSs are by necessity approximate. Even worse, interactions between atoms make phase transitions possible, and EOS becomes discontinuous (*i.e.*, non-analytical) at phase transitions. All the approximate EOS formulations are valid only for one phase, and generally the accuracy of the EOS is best at conditions far from phase transitions.

Consider the isothermal EOS: $p = -(\partial F/\partial V)_T$, taking the QHA as the starting point. Using indices *i* and **k** to denote the number of the phonon branch and the wavevector **k**, we have:

$$F(T) = E_0 \frac{1}{2}\sum_{i,\mathbf{k}} \hbar\omega_{i\mathbf{k}} + k_{\mathrm{B}}T\sum_{i,\mathbf{k}}\left[1 - \exp\left(-\frac{\hbar\omega_{i\mathbf{k}}}{k_{\mathrm{B}}T}\right)\right],$$

differentiating which one arrives at

$$p(V,T) = p_{\mathrm{st}}(V) + \frac{1}{2}\sum_{i,\mathbf{k}}\hbar\frac{\gamma_{i\mathbf{k}}\omega_{i\mathbf{k}}}{V} + \sum_{i,\mathbf{k}}\frac{\gamma_{i\mathbf{k}}E_{\mathrm{th},i\mathbf{k}}}{V}, \tag{29}$$

where $\gamma_{i\mathbf{k}} = -\left(\dfrac{\partial\ln\omega}{\partial\ln V}\right)_T$ is the mode Grüneisen parameter.

Defining average Grüneisen parameters, re-write Equation 29:

$$p(V,T) = p_{\mathrm{st}}(V) + \langle\gamma_{i\mathbf{k}}(V)\rangle\frac{E_{\mathrm{zp}}(V)}{V} + \Gamma(V,T)\frac{E_{\mathrm{th}}(V,T)}{V}, \tag{30}$$

where Γ is the so called thermodynamic Grüneisen parameter defined as a weighted average of the mode Grüneisen parameters:

$$\Gamma = \frac{\displaystyle\sum_{i,\mathbf{k}}\gamma_{i\mathbf{k}}E_{\mathrm{th},i\mathbf{k}}}{E_{\mathrm{th}}} = \frac{\displaystyle\sum_{i,\mathbf{k}}\gamma_{i\mathbf{k}}C_{\mathrm{V},i\mathbf{k}}}{C_{\mathrm{V}}}, \tag{31}$$

where $C_{\mathrm{V},i\mathbf{k}}$ is the contribution of the *i*-th mode (at the **k** wavevector) to the total isochoric heat capacity (Eqn. 17). The term $\langle\gamma_{i\mathbf{k}}(V)\rangle\dfrac{E_{\mathrm{zp}}(V)}{V}$ in Equation 30 describes pressure due to zero-point motion.

For the thermal expansion one has:

$$\alpha = \Gamma\frac{C_{\mathrm{V}}}{K_{\mathrm{T}}V}. \tag{32}$$

At high temperatures, where all phonons are fully excited, Equation 31 reduces to a constant high-temperature limit, which is just the arithmetic mean over all mode Grüneisen parameters: $\Gamma \to \langle\gamma_{i\mathbf{k}}\rangle$. In the classical approximation, which is the basis of standard molecular dynamics and Monte Carlo simulations, thermodynamic Grüneisen parameter is close to $\langle\gamma_{i\mathbf{k}}\rangle$ (Welch *et al.*, 1978), but also includes a temperature-dependent correction due to intrinsic anharmonic effects.

Generally, Γ is a function of temperature as well as volume (*cf.* Eqn. 31), often with a non-monotonic temperature dependence (Oganov *et al.*, 2000; Price *et al.*, 1987). This dependence is usually small and neglected in the Mie–Grüneisen approximation. The best value for the Mie–Grüneisen approximation is $\langle\gamma_{i\mathbf{k}}\rangle$ (denoted below as γ), which is correct in the high-temperature limit and also appears in the expression for the pressure of zero-point vibrations.

The famous Mie–Grüneisen thermal EOS is then written as:

$$p(V,T) = p_{0\mathrm{K}}(V) + p_{\mathrm{th}}(V,T) = p_{0\mathrm{K}}(V) + \gamma(V)\frac{E_{\mathrm{th}}(V,T)}{V}. \tag{33}$$

The volume dependence of γ is described by a power law:

$$\gamma(V) = \gamma_0 \left(\frac{V}{V_0} \right)^q,$$

(34)

with q usually assumed to be constant. In many studies (*e.g.*, Oganov *et al.*, 2000), however, it was found that $q = \left(\dfrac{\partial \ln \gamma}{\partial \ln V} \right)_T$ varies very strongly with volume.

Although derived within the QHA, the Mie–Grüneisen EOS can be formulated at any level of theory. Generally speaking, Equation 31 will become invalid due to a correction for intrinsic anharmonicity, but in such cases effective Grüneisen parameter can be defined, implicitly subsuming the anharmonic effects.

It is possible to separate explicitly the intrinsic anharmonic effects (*e.g.*, Zharkov & Kalinin, 1968; Gillet *et al.*, 1999; Dorogokupets, 2000). Making use of the high-temperature result for each mode $F_a(V, T) = \dfrac{1}{2} k_B a_i(V) T^2$ (Eqn. 26), the anharmonic thermal pressure per mode is:

$$p_a = -\frac{1}{2} k_B a_i(V) \left(\frac{\partial \ln a_i}{\partial \ln V} \right)_T \frac{T^2}{V}.$$

(35)

Analytical representations of the equation of state
Assuming that K varies linearly with pressure and denoting $K_0' = (\partial K/\partial p)_{T, p=0}$, we obtain the Murnaghan EOS:

$$p = \frac{K_0}{K_0'} \left[\left(\frac{V}{V_0} \right)^{-K_0'} - 1 \right].$$

(36)

This simple EOS works well in the compression range $|(V - V_0)/V_0| < 0.1$–0.2. At larger compressions higher-order derivatives of K become important. Instead of including these non-linearities of successively higher orders, it is better to construct an approximate model for the energy as a function of $x = V_0/V$ or some other measure of strain. Using this method, some of the most popular EOSs are obtained – families of Birch–Murnaghan and logarithmic EOSs of different orders. Vinet EOS follows from a general scaled interatomic potential.

Birch–Murnaghan EOSs. These EOSs are based on the polynomial expansion of the energy:

$$E = E_0 + af^2 + bf^3 + cf^4 + \ldots$$

(37)

in powers of the Eulerian strain f_E:

$$f_E = \frac{1}{2} \left[1 - \left(\frac{V}{V_0} \right)^{\frac{2}{3}} \right].$$

(38)

It is advantageous to use the Eulerian finite strain rather than the Lagrangian strain:

$$f_L = \frac{1}{2}\left[\left(\frac{V}{V_0}\right)^{\frac{2}{3}} - 1\right],$$ (39)

because Eulerian strain leads to a better description of the correct $E(V)$ dependences with fewer terms in the expansion (37). However, for infinitesimal strains both definitions become equivalent, and

$$f_E = f_L = \frac{da}{a_0} = -\frac{1}{3}\frac{dV}{V_0}.$$ (40)

The second-order Birch–Murnaghan EOS (BM2) is:

$$p = \frac{3}{2}K_0\left[\left(\frac{V_0}{V}\right)^{\frac{7}{3}} - \left(\frac{V_0}{V}\right)^{\frac{5}{3}}\right],$$ (41)

which corresponds to the energy:

$$E = E_0 + \frac{9}{8}K_0V_0\left[\left(\frac{V_0}{V}\right)^{\frac{2}{3}} - 1\right]^2.$$ (42)

The third-order BM EOS is

$$p(V) = \frac{3}{2}K_0\left[\left(\frac{V_0}{V}\right)^{\frac{7}{3}} - \left(\frac{V_0}{V}\right)^{\frac{5}{3}}\right]\left\{1 + \xi\left[\left(\frac{V_0}{V}\right)^{\frac{2}{3}} - 1\right]\right\},$$ (43)

where $\xi = \frac{3}{4}(K_0' - 4)$.

When $K_0' = 4$, the third-order Birch–Murnaghan EOS (BM3) reduces to the second-order one (BM2). In other words, BM2 EOS implicitly assumes $K_0' = 4$. It is remarkable that this estimate of the lowest order of finite strain theory is often very close to experimental values of K_0' of many crystals (especially with close-packed structures). This is why BM EOSs work well in many cases. Their main problem is rather poor convergence with respect to the number of terms beyond the second order for large strains: it turns out that the fourth-order energy term in Equation 37 is often larger than the third-order one. The third-order BM EOS is very accurate when K_0' is close to 4; when it is significantly different, this EOS gives poor results at high compressions. For $K_0' < 4$ at large compressions (*e.g.*, for $V/V_0 \approx 0.5$ and $K_0' = 2$) BM3 can give even negative pressures (Poirier & Tarantola, 1998)! Being based on a polynomial expansion of the interatomic potential (of the type $\sum_{n=1,2,3,...} a_n R^{-2n}$), this EOS may become unreliable for extrapolations (Cohen *et al.*, 2000).

It is possible to derive systematically higher-order BM EOS, but this appears to be of little use since the number of parameters involved will be too large; only the fourth-order BM EOS:

$$p = 3K_0 f_E (1 + 2f_E)^{\frac{5}{2}} \left\{ 1 + \frac{3}{2}(K_0' - 4)f_E + \frac{3}{2}\left[K_0 K_0'' + (K_0' - 4)(K_0' - 3) + \frac{35}{9} \right] f_E^2 \right\} \quad (44)$$

is sometimes used. BM4 EOS seems to be very accurate in a wide range of compressions.

Logarithmic EOSs. Drawbacks of the Birch–Murnaghan EOSs have led Poirier & Tarantola (1998) to derive a new EOS, based on the polynomial expansion (37) in powers of the natural (also called logarithmic, or Hencky) strain:

$$f_H = \ln\left(\frac{l}{l_0}\right) = \frac{1}{3}\ln\left(\frac{V}{V_0}\right), \quad (45)$$

where l is the length of the sample (l_0 in the unstrained sample). The third-order logarithmic EOS is:

$$p = K_0 \frac{V_0}{V}\left[\ln\left(\frac{V_0}{V}\right) + \frac{1}{2}(K_0' - 2)\ln^2\left(\frac{V_0}{V}\right) \right]. \quad (46)$$

In the fourth order, the logarithmic EOS is:

$$p = K_0 \frac{V_0}{V}\ln\left(\frac{V_0}{V}\right)\left\{ 1 + \frac{1}{2}(K_0' - 2)\ln\left(\frac{V_0}{V}\right) + \frac{1}{6}\left[1 + K_0 K_0'' + (K_0' - 2) + (K_0' - 2)^2 \right]\ln^2\left(\frac{V_0}{V}\right) \right\}. \quad (47)$$

In the case of MgO examined in Table 2, the third-order logarithmic EOS shows a very poor performance compared to BM3 or Vinet EOS. It can be expected that this EOS can work well in those rare cases where $K_0' \approx 2$.

Table 2. Analytical EOSs fitted to theoretical p–V data on periclase (MgO)

p_{max} (N)	MUR		BM2 ($K_0' = 4$)	BM3		Logarithmic third-order		Vinet	
	K_0	K_0'	K_0	K_0	K_0'	K_0	K_0'	K_0	K_0'
165 GPa (12)	159.18	3.588	158.53	151.66	4.213	138.42	5.594	146.61	4.653
150 GPa (11)	158.66	3.605	158.13	151.79	4.208	140.02	5.477	147.25	4.624
105 GPa (8)	156.55	3.688	156.86	151.74	4.210	144.16	5.157	148.75	4.545
60 GPa (5)	154.70	3.788	155.29	152.12	4.186	148.43	4.781	150.67	4.416
30 GPa (3)	154.03	3.846	154.13	152.96	4.110	151.32	4.448	152.41	4.246
0 GPa	$K_0 = 153.23$; $K_0' = 4.2 \pm 0.1$ (from elastic constants)								

MUR – Murnaghan EOS. Data were obtained by pseudopotential plane wave calculations within the generalised gradient approximation (GGA). This table shows EOS parameters obtained for a pressure range 0–p_{max} with N data points. Zero-pressure volume V_0 was fixed at the theoretical value 77.64 Å^3. The last row represents independent calculations of the elastic constants.

Logarithmic EOS corresponds to an unphysical interactomic potential of the type

$$\sum_{n=1,2,3,\dots} a_n (\ln R)^n$$

(Cohen *et al.*, 2000). Even though it can be expanded as $-a/R + bR^{-(3K_0' - 7)}$ (Poirier & Tarantola, 1998), one can note that the repulsion exponent is too small (5 for $K_0' = 4$) compared to the usual values $\sim$ 9–12. This is why, although based on a potentially powerful idea of the Hencky finite strain, the logarithmic EOS is not generally successful. Vinet EOS is the best choice of an EOS for solids – it combines simplicity, high accuracy, and universal applicability to solids with any character of chemical bonding, and up to extremely high compressions (at least 2–3-fold compressions).

Vinet EOS (Vinet *et al.*, 1986, 1989)*.* This EOS, also known as the Universal EOS, is sometimes considered as one of the major recent achievements of solid-state physics (Sutton, 1993). Like BM and logarithmic EOSs, Vinet EOS is a family of EOSs of different orders. The most remarkable feature is very fast convergence with respect to the order of EOS – one seldom needs to go beyond the third-order Vinet EOS.

The third-order Vinet EOS is:

$$p = 3K_0 \frac{1 - (V/V_0)^{1/3}}{(V/V_0)^{2/3}} \exp\left\{ \frac{3}{2}(K_0' - 1)\left[1 - \left(\frac{V}{V_0}\right)^{1/3} \right] \right\}. \tag{48}$$

This EOS is based on a universal scaled binding curve:

$$E = E_0(1 + a)\exp(-a), \tag{49}$$

where E_0 is the bond energy at equilibrium, $a = (R - R_0)/l$, $l = \sqrt{E_0 \Big/ \left(\dfrac{\partial^2 E}{\partial R^2}\right)}$ being a

scaling length roughly measuring the width of the potential well, and R the Wigner–Seitz radius (the average radius of a sphere in the solid containing 1 atom). The potential (49) was first used in 1930s by Rydberg for fitting potential curves of molecules and obtaining their anharmonic coefficients; it turned out (see Sutton, 1993 and references therein) that it describes very accurately systems with different types of chemical bonding in solids, molecules, adsorbates *etc.* The Vinet EOS proved to be very accurate for fitting the EOS of solid hydrogen (Loubeyre *et al.*, 1996; Cohen *et al.*, 2000) in the pressure range 0–120 GPa, roughly to the 8-fold compression. The universal binding curve (Eqn. 49), however, does not accurately describe long-range forces in ionic and van der Waals crystals, which means that for such materials Vinet EOS will not give accurate description of large expansion ($V_0/V << 1$). For compression, however, it should be sufficiently accurate, since compressional behaviour is dominated by interatomic repulsion, well accounted for by Equation 49.

In very rare cases a higher-order Vinet EOS may be needed; such higher-order versions of the Vinet EOS already exist (Vinet *et al.*, 1989); fourth-order Vinet EOS has been applied to solid H_2 at extreme compressions (Cohen *et al.*, 2000) and has lead to significant

improvements of the description of experimental *p–V* data. At extreme compressions ($V/V_0 < 0.1$), where solids approach the free-electron regime, other EOSs are required. Such EOSs, manifesting the Thomas–Fermi limiting behaviour at extreme compressions, are developed and discussed in detail by Holzapfel (1996) and Hama & Suito (1996).

Table 2 gives a comparison of several EOSs fitted to theoretical data on MgO. Fits were performed in different pressure ranges. Ideally, EOS parameters (V_0, K_0, and K_0') should be independent of the pressure range of fitting and consistent with zero-pressure values of these parameters. K_0 and K_0' were also determined from the independently calculated elastic constants at several pressures. The best results are obtained with the BM3 and Vinet EOS. For MgO BM3 slightly outperforms Vinet EOS, however, there are more examples of the opposite. Table 2 also shows that fitted values of the EOS parameters (especially K_0') strongly depend on the EOS used. As the pressure range of fitting tends to zero, all EOSs become equivalent.

Since both K' and γ come from anharmonicity, an intriguing possibility arises of a general relation between them. This possibility has been widely discussed since 1939, when Slater suggested the first solution of the problem:

$$\gamma_s = \frac{1}{2} K' - \frac{1}{6}. \tag{50}$$

Later approaches resulted in similar equations, the difference being in the value of the constant subtracted from K': 1/2, 5/6 or 0.95. If any of the relations of the type (50) were accurate, it would greatly simplify the construction of thermal EOS. Although some correlation between γ and K' does exist, the correlation is too poor to be useful (Wallace, 1998).

EOS, internal strain, and phase transitions
All the EOSs discussed above were based on the assumption that crystal structures compress uniformly, *i.e.* there is no relaxation of the unit cell shape or atomic positions. For some crystals (*e.g.*, MgO) this is true: there are no internal degrees of freedom, and the unit cell shape is fixed by symmetry. For most crystals and all glasses, however, this is not the case. As emphasised by Angel & Ross (1996), all the classical EOSs are less successful for crystals with internal degrees of freedom. Effects of atomic relaxation (often called the effects of the internal strain) lead to renormalised values of the EOS parameters. These EOSs perform particularly poorly in the vicinity of phase transitions.

Let us develop a simple qualitative model of the internal strain effects. Homogeneous compression leads to a configuration with atoms generally off their equilibrium positions. We denote such displacements as Q. Denoting the energy of the crystal without atomic relaxation as E_0, in the lowest order (*i.e.*, in the harmonic approximation) the energy E with positional relaxation will be:

$$E = E_0 - \frac{1}{2} kQ^2,$$

where k is the force constant. For the static pressure $p = -\left(\dfrac{\partial E}{\partial V}\right)$ we find

$$p(V) = -\left(\frac{\partial E_0}{\partial V}\right) + \frac{1}{2}\frac{\partial(kQ^2)}{\partial V} = p_0(V) + kQ\left(\frac{\partial Q}{\partial V}\right) + \frac{1}{2}Q^2\left(\frac{\partial k}{\partial V}\right), \qquad (51)$$

where p_0 is the pressure corresponding to the homogeneous compression.

Let us represent Q in the lowest order as a linear function of volume and take the reference $Q_i = 0$ at $V = V_0$. Then,

$$Q_i = -a(V - V_0).$$

Using a strictly harmonic approximation and neglecting the volume dependence of the force constant k (for a more general treatment see Oganov, 2002) we obtain:

$$p(V) = p_0(V) + a^2 k_0(V - V_0). \qquad (52)$$

The term $p_0(V)$ is well described by conventional EOSs (*e.g.*, Vinet EOS), whereas this may not be the case for the total EOS. The bulk modulus $K = -V\left(\frac{\partial p}{\partial V}\right)_T$ is always lowered by the relaxational effects, in the simplest approximation (Eqn. 52):

$$K(V) = K_0(V) - a^2 k_0 V, \qquad (53)$$

whereas K' is predicted to be increased.

Complex structures with many degrees of freedom are relatively 'soft' and often have large K'_0 significantly exceeding 'normal' $K'_0 = 4$. It is well known that internal strains always soften elastic constants (*e.g.*, Catti, 1989) and, in particular, the bulk modulus. *E.g.*, quartz SiO_2, despite consisting of extremely rigid SiO_2 tetrahedra, has a very low bulk modulus $K_0 = 37.12$ GPa and high $K'_0 = 5.99$ (Angel *et al.*, 1997). Perhaps, the highest known $K'_0 = 13$ was found in amphibole grunerite (Zhang *et al.*, 1992), whose structure has very many degrees of freedom.

Results of *ab initio* calculations of sillimanite (Al_2SiO_5), presented in Figure 6, illustrate these points. Without allowing the atomic coordinates and the shape of the unit cell to relax, the structure is much stiffer. BM3 EOS fitted to the $p(V)$ data has $K_0 = 199.7$ GPa and 160 GPa, $K'_0 = 4.0$ and 2.7 for the unrelaxed and relaxed EOSs, respectively. In agreement with the prediction in Equation 53, the bulk modulus is lowered by relaxation. $K'_0 = 4.0$ of the unrelaxed EOS is typical of crystals without internal degrees of freedom (*e.g.*, close-packed oxides and metals). For the relaxed EOS $K'_0 = 2.7$, but values close to 4 would result from fitting in narrower pressure ranges. Low values of K' can be related to phase transitions, particularly if the transition (like in sillimanite) is accompanied by softening and collapse of the structure (in sillimanite, a soft longitudinal acoustic mode develops, leading to an isosymmetric phase transition at 38 GPa. In the vicinity of this transition K' is negative). To study EOS near phase transitions, one needs to go beyond the harmonic approximation built in the model (53). This can be done using the Landau expansion of the internal energy in powers of Q including the full elastic constants tensor and allowed couplings of the order parameter and lattice strains.
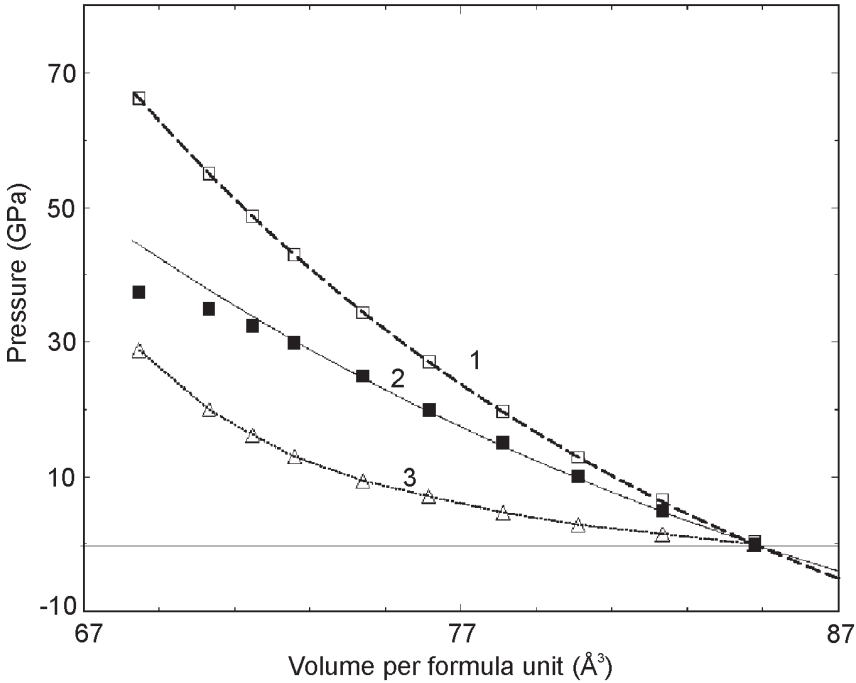
Fig. 6. Effects of internal strains on EOS. (1) unrelaxed EOS (symbols: calculated points, line: BM3 fit), (2) full EOS including relaxation (symbols: calculated points, line: BM3 fit), (3) the difference caused by relaxation. Note that in the pre-transition region the full EOS is poorly fit, while the unrelaxed EOS is well represented by BM3 EOS everywhere. The relaxational correction is very large.

Elastic constants

The importance of elastic constants for the Earth sciences springs from the fact that the main information about the deep Earth is obtained seismologically, and seismic wave velocities are related to the elastic properties of the Earth-forming rocks and minerals. A number of excellent books and reviews consider the elastic properties; we mention Nye (1998), Sirotin & Shaskolskaya (1975), Wallace (1998), Alexandrov & Prodaivoda (1993), Born & Huang (1954), Belikov *et al.* (1970), Fedorov (1968).

Elastic constants characterise the ability of a material to deform under small stresses. They are be described by a fourth rank tensor C_{ijkl}, relating the second rank stress tensor σ_{ij} to the (also second rank) strain tensor e_{kl} via the generalised Hooke's law:

$$\sigma_{ij} = C_{ijkl}e_{kl}, \tag{54}$$

where multiplication follows the rules of tensor multiplication (see Nye, 1998). Equation 54 can be simplified using the Voigt notation (Nye, 1998), in which pairs of indices '11', '22', '33', '12', '13', '23' are represented by only one symbol: 1, 2, 3, 6, 5, and 4, respectively. Instead of Equation 54 we write:

$$\sigma_i = C_{ij}e_j. \tag{55}$$

Here we use infinitesimal strains; under an infinitesimal strain, the lattice vectors matrix a'_{ij} of the strained crystal is obtained from the old lattice matrix a^0_{ij} and the strain tensor e_{ij} using the relation:

$$a'_{ij} = (\delta_{ij} + e_{ij})a^0_{ij}. \tag{56}$$

In the original tensor notation and in the Voigt notation (Nye, 1998), the $(\delta_{ij} + e_{ij})$ matrix is represented as follows:

$$\begin{pmatrix} 1+e_{11} & e_{12} & e_{13} \\ e_{12} & 1+e_{22} & e_{23} \\ e_{13} & e_{23} & 1+e_{33} \end{pmatrix} = \begin{pmatrix} 1+e_1 & +e_6/2 & e_5/2 \\ e_6/2 & 1+e_2 & e_4/2 \\ e_5/2 & +e_4/2 & 1+e_3 \end{pmatrix}. \tag{57}$$

The Voigt notation allows one to write the elastic constants conveniently as a symmetric 6×6 matrix; this is adequate for most situations (except, *e.g.*, for a general transformation of the coordinate system, where only the original fourth-rank tensor representation can be used to derive the transformed elastic constants). The number of components of a fourth rank tensor is 81; the Voigt notation reduces this to 36. The thermodynamic equality $C_{ij} = C_{ji}$ makes the 6×6 matrix of the elastic constants symmetric relative to its main diagonal, reducing the number of independent constants to 21. Symmetry further reduces this number to 13 for monoclinic, 9 for orthorhombic, 6 or 7 (depending on the point group symmetry) for trigonal and tetragonal, 5 for hexagonal and 3 for cubic crystals; for isotropic (amorphous) solids there are only 2 independent elastic constants. The elastic constants and their anisotropy are closely related to crystal structures. *E.g.*, for andalusite and sillimanite (Al_2SiO_5) this relation was clearly demonstrated by Vaughan & Weidner (1978).

One can define the inverse tensor S_{ijkl} (or, in Voigt notations, S_{ij}), often called the elastic compliance tensor:

$$\{S_{ijkl}\} = \{C_{ijkl}\}^{-1} \text{ or } \{S_{ij}\} = \{C_{ij}\}^{-1}. \tag{58}$$

For an orthorhombic crystal, this matrix inversion leads to the following explicit relations between C_{ij} and S_{ij} (Belikov *et al.*, 1970):

$$\begin{aligned} S_{11} &= (C_{22}C_{33} - C^2_{23})/D, \\ S_{22} &= (C_{11}C_{33} - C^2_{13})/D, \\ S_{33} &= (C_{11}C_{22} - C^2_{12})/D, \\ S_{44} &= 1/C_{44}, \\ S_{55} &= 1/C_{55}, \\ S_{66} &= 1/C_{66}, \\ S_{12} &= -(C_{12}C_{33} - C_{13}C_{23})/D, \\ S_{13} &= (C_{12}C_{23} - C_{13}C_{22})/D, \\ S_{23} &= -(C_{11}C_{23} - C_{13}C_{12})/D, \end{aligned} \tag{59}$$

where $D = C_{11}C_{22}C_{33} + 2C_{12}C_{13}C_{23} - C^2_{23}C_{11} - C^2_{12}C_{33} - C^2_{13}C_{22}$.

The S_{ij} tensor can be defined via the generalised Hooke's law in its equivalent formulation:

$$e_i = S_{ij}\sigma_j. \tag{60}$$

Linear compressibilities can be easily derived from the S_{ij} tensor. Full expressions for an arbitrary direction in a triclinic crystal can be found in Nye (1998); along the coordinate axes they are:

$$\beta_x = -\frac{1}{a}\left(\frac{\partial a}{\partial p}\right)_T = \sum_{j=1}^{3} S_{1j} = S_{11} + S_{12} + S_{13}$$

$$\beta_y = -\frac{1}{b}\left(\frac{\partial b}{\partial p}\right)_T = \sum_{j=1}^{3} S_{2j} = S_{22} + S_{12} + S_{23} \qquad (61)$$

$$\beta_z = -\frac{1}{c}\left(\frac{\partial c}{\partial p}\right)_T = \sum_{j=1}^{3} S_{3j} = S_{33} + S_{13} + S_{23},$$

where a, b, c are linear dimensions along the axes of the coordinate system[2]. The bulk compressibility is:

$$\beta = -\frac{1}{V}\left(\frac{\partial V}{\partial p}\right)_T = \beta_x + \beta_y + \beta_z =$$

$$= \sum_{i=1}^{3}\sum_{j=1}^{3} S_{ij} = S_{11} + S_{22} + S_{33} + 2(S_{12} + S_{13} + S_{23}), \qquad (62)$$

where V is the volume (*e.g.*, unit cell volume).

There are two particularly important quantities – bulk modulus K and shear modulus G. These are obtained by special averaging of the individual elastic constants. There are several different schemes of such averaging. The assumption of a homogeneous stress throughout the crystal leads to the Reuss bulk and shear moduli:

$$K_R = [S_{11} + S_{22} + S_{33} + 2(S_{12} + S_{13} + S_{23})]^{-1} = \frac{1}{\beta}, \qquad (63)$$

$$G_R = 15[4(S_{11} + S_{22} + S_{33}) - 4(S_{12} + S_{13} + S_{23}) + 3(S_{44} + S_{55} + S_{66})]^{-1}. \qquad (64)$$

It is the Reuss bulk modulus, explicitly related to compressibility, which is used in constructing EOSs. This is also the modulus appearing in all thermodynamic equations involving the bulk modulus.

Another scheme of averaging, due to Voigt, is based on the assumption of a spatially homogeneous strain and leads to the following expressions:

$$K_V = \frac{1}{9} [C_{11} + C_{22} + C_{33} + 2(C_{12} + C_{13} + C_{23})], \qquad (65)$$

$$G_V = \frac{1}{15} [C_{11} + C_{22} + C_{33} - (C_{12} + C_{13} + C_{23}) + 3(C_{44} + C_{55} + C_{66})]. \qquad (66)$$

For an isotropic polycrystalline aggregate the Voigt and the Reuss schemes give

---

[2] These axes may not coincide with the lattice vectors for non-orthogonal crystal systems. Coordinate systems used in crystal physics are always orthogonal.

upper and lower bounds, respectively, of the corresponding moduli. More accurate estimates are obtained from Voigt–Reuss–Hill averages:

$$K_{\text{VRH}} = \frac{1}{2} \, (K_{\text{V}} + K_{\text{R}}); \ \ G_{\text{VRH}} = \frac{1}{2} \, (G_{\text{V}} + G_{\text{R}}). \tag{67}$$

Slightly more accurate results and tighter bounds are given by the more complicated Hashin–Shtrikman variational averaging scheme.

There are two groups of experimental methods of measurement of the elastic constants – (i) static methods (based on determination of stress–strain relations for static stresses) and (ii) high-frequency, or dynamic, methods (*e.g.*, ultrasonic methods, Brillouin spectroscopy *etc.*). Static methods are not used anymore, because of their poor accuracy. Static measurements yield isothermal elastic constants (timescale of the experiment allows thermal equilibrium to be attained within the sample), high-frequency measurements give adiabatic constants (Belikov *et al.*, 1970). The difference is entirely due to the anharmonic effects (see below) and vanishes at 0 K. Adiabatic C_{ij} are always larger (Wallace, 1998):

$$C_{ijkl}^{\text{S}} = C_{ijkl}^{\text{T}} + \frac{TV}{C_{\text{V}}} \, b_{ij} b_{kl} \, , \tag{68}$$

where $b_{ij} = (\partial \sigma_{ij} / \partial T)_{\text{V}}$ are related to the thermal expansion tensor. For the bulk modulus we have:

$$K_{\text{S}} = K_{\text{T}}(1 + \alpha \gamma T) = K_{\text{T}} \left(1 + \frac{\alpha^2 K_{\text{T}} V}{C_{\text{V}}} \right) \, , \tag{69}$$

where α and γ are the thermal expansion and Grüneisen parameter, respectively. Adiabatic and isothermal shear moduli are strictly equal for cubic crystals and usually practically indistinguishable for crystals of other symmetries. Acoustic wave velocities measured in seismological experiments are related to the adiabatic elastic constants.

The elastic constants tensor allows one to calculate the velocities of acoustic waves with an arbitrary propagation direction (see Sirotin & Shaskolskaya, 1975). There are 3 acoustic modes – 1 longitudinal and 2 shear modes. Their average velocities are calculated as follows:

$$V_{\text{p}} = \sqrt{\frac{3K + 4G}{3\rho}} \tag{70}$$

and

$$V_{\text{s}} = \sqrt{\frac{G}{\rho}} \, , \tag{71}$$

where the adiabatic Voigt–Reuss–Hill (or Hashin–Shtrikman) values are used for the bulk and shear moduli.

The energy change (per unit volume) due to a deformation can be described by the adiabatic elastic constants:

$$\frac{\Delta E}{V} = \frac{1}{2} C_{ijkl}^{S} e_{ij} e_{kl} . \tag{72}$$

An analogous relation exists between the Helmholtz free energy and isothermal constants:

$$\frac{\Delta F}{V} = \frac{1}{2} C_{ijkl}^{T} e_{ij} e_{kl} . \tag{73}$$

These relations enable the calculation of elastic constants at zero pressure.

It is well known (Barron & Klein, 1965; Wallace, 1998) that at non-zero pressures there can be several different definitions of elastic constants. Consider a solid at a non-zero hydrostatic pressure p. For the stress tensor we have:

$$\sigma_{ij} = -p\delta_{ij} + C_{ijkl}e_{kl}. \tag{74}$$

Instead of the zero-stress equation (Eqn. 72), the energy density becomes:

$$\frac{\Delta E}{V} = -pe_{ii} + \frac{1}{2}\left[C_{ijkl}^{S} - \frac{1}{2} p(2\delta_{ij}\delta_{kl} - \delta_{il}\delta_{jk} - \delta_{jl}\delta_{ik}) \right] e_{ij} e_{kl} - \frac{1}{2} p\omega_{ik}\omega_{ik}, \tag{75}$$

where ω_{ik} corresponds to a pure rotation of the crystal (rotation tensor), while e_{ij} corresponds to a pure strain. At hydrostatic pressure this gives (Barron & Klein, 1965):

$$C_{ijkl}^{S} = \frac{1}{V}\left(\frac{\partial^2 E}{\partial e_{ij} \partial e_{kl}} \right)_{V} + \frac{1}{2} p(2\delta_{ij}\delta_{kl} - \delta_{il}\delta_{jk} - \delta_{jl}\delta_{ik}). \tag{76}$$

Calculating the second derivatives with respect to the finite Lagrangian strains η_{ij}, different equations are obtained (Wallace, 1998):

$$C_{ijkl}^{S} = \frac{1}{V}\left(\frac{\partial^2 E}{\partial \eta_{ij} \partial \eta_{kl}} \right)_{V} + p(\delta_{ij}\delta_{kl} - \delta_{il}\delta_{jk} - \delta_{jl}\delta_{ik}). \tag{77}$$

Equations 76 and 77 are particularly important, because many simulation packages define the elastic constants as:

$$c_{ijkl}^{S} = \frac{1}{V}\left(\frac{\partial^2 E}{\partial \eta_{ij} \partial \eta_{kl}} \right)_{V}. \tag{78}$$

This definition is equivalent to the definition via stress-strain relations only at zero stress. The definition based on stress–strain relations (Eqns. 76–77) is the most useful one: it is directly related to the mechanical stability criteria and acoustic wave velocities. Cauchy relations, originally derived with the definition via the energy density, can be elegantly formulated in this definition as well (see below). It is essential to distinguish between different definitions of elastic constants and keep in mind Equations 76 and 77.

Cauchy relations

In the absence of internal strains (*i.e.* for crystals with no internal degrees of freedom), in a static lattice where all atoms occupy centrosymmetric positions and all interatomic interactions are central and pairwise (*i.e.*, depend only on distances between atoms, and not on angles), Cauchy relations (Born & Huang, 1954; but take into account Equations 76–77) hold:

$$C_{23} - C_{44} = 2p; \quad C_{31} - C_{55} = 2p; \quad C_{12} - C_{66} = 2p;$$
$$C_{14} - C_{56} = 0; \quad C_{25} - C_{64} = 0; \quad C_{36} - C_{45} = 0. \tag{79}$$

These relations would reduce the maximum number of independent elastic constants to 15; however, they never hold exactly because there always are non-central and many-body contributions to crystal energy. Violations of the Cauchy relations can serve as a useful indicator of the importance of such interactions. While for many alkali halides Cauchy relations hold reasonably well, for alkali earth oxides, also considered as classical ionic solids, Cauchy relations are grossly violated. At high pressures, solid Ar has large Cauchy violations, indicating that many-body interactions are important (Iitaka & Ebisuzaki, 2001).

Mechanical stability

One of the most common types of instabilities in crystals is the so-called mechanical instability, when one or more elastic constants (or their special combinations) tend to zero. The condition of mechanical stability is the positive definiteness of the elastic constants matrix:

$$
\begin{array}{|cccccc|}
\hline
C_{11} & C_{12} & C_{13} & C_{14} & C_{15} & C_{16} \\
C_{21} & C_{22} & C_{23} & C_{24} & C_{25} & C_{26} \\
C_{31} & C_{32} & C_{33} & C_{34} & C_{35} & C_{36} \\
C_{41} & C_{42} & C_{43} & C_{44} & C_{45} & C_{46} \\
C_{51} & C_{52} & C_{53} & C_{54} & C_{55} & C_{56} \\
C_{61} & C_{62} & C_{63} & C_{64} & C_{65} & C_{66} \\
\hline
\end{array}
$$

This is equivalent to the positiveness of all the principal minors of this matrix (principal minors are square sub-matrices symmetrical with respect to the main diagonal – they are indicated by dashed lines in the scheme above). Diagonal elastic constants C_{ii} are positive for all stable crystals. Dynamical stability criteria were first suggested by Born (Born & Huang, 1954) and are often called the Born conditions. In a general form they were discussed in Sirotin & Shaskolskaya (1975) and Fedorov (1968). For crystals of different symmetries these conditions were analysed by Cowley (1976) and Terhune *et al.* (1985). Mechanical stability criteria for crystals under stress must employ the C_{ij} derived from the stress-strain relations (Wang *et al.*, 1993, 1995; Karki, 1997). Violation of any of the mechanical stability conditions leads to softening of an acoustic mode in the vicinity of the Γ point, inducing a ferroelastic phase transition.

Temperature effects and elastic constants

In the harmonic approximation, the elastic constants are temperature-independent. The quasiharmonic and intrinsically anharmonic corrections can actually be large. These contributions to the bulk modulus can be represented (see Dorogokupets, 2000) as additive corrections to the zero-temperature result:

$$K^{T}(V, T) = K_{0\,K}(V) + \Delta K_{qha}^{T}(V, T) + \Delta K_{a}^{T}(V, T), \qquad (80)$$

$$\Delta K_{qha}^{T}(V,T) = p_{th,qha}[1 + \gamma - q] - \gamma^2 T C_{V}/V, \qquad (81)$$

$$\Delta K_{a}^{T}(V,T) = p_{a}\left[1 + m - \left(\frac{\partial \ln m}{\partial \ln V}\right)_{T}\right], \qquad (82)$$

where $q = \left(\dfrac{\partial \ln \gamma}{\partial \ln V}\right)_{T}$, $m = \left(\dfrac{\partial \ln a}{\partial \ln V}\right)_{T}$ (*cf.* Eqn. 35), and formalism of Equations 26, 27, and 35 is used for the anharmonic contributions.

Part III: Phase transitions

The study of phase transitions is of a central importance to modern crystallography, condensed matter physics and chemistry. Phase transitions, being a major factor determining the seismic structure of the Earth, play a special role in geophysics. Here we give a brief review of the theory and classification of phase transitions.

Classifications of phase transitions

Perhaps the first sound classification of phase transitions was proposed by Ehrenfest in 1933 (for a detailed historical and scientific discussion see Jaeger, 1998). *First*-order transitions are those for which the *first* derivatives of the free energy with respect to p and T (*i.e.*, V and S) are discontinuous at the transition point; for *second*-order transitions the *second* derivatives (K and α) are discontinuous, and so forth. Third and higher-order transitions, although theoretically possible, so far have never been reported. Modern variants of the Ehrenfest classification distinguish only between 'first-order' and 'continuous' (*i.e.*, all higher-order) transitions. In a number of cases, the order of the same transition is different at different p–T conditions. Some transitions change from first to second order; the crossover point is called the tricritical point. Among the examples of systems with tricritical crossover are NH_4Cl (Garland & Weiner, 1971), zone-centre cubic–tetragonal transition in $BaTiO_3$ perovskite, possibly the transition from calcite to metastable calcite (II) in $CaCO_3$ (see Hatch & Merrill, 1981), and, possibly, the α–β transition in quartz (SiO_2). Therefore, the order of the transition is not something fundamental. The Ehrenfest classification attempts to describe only thermodynamics of phase transitions without linking it to structural mechanisms or symmetry aspects of transitions. But it is the structure and symmetry relations between the phases that are the most fundamental characteristics of any phase transition.

The first structural classification was due to Buerger (1961), who distinguished between two main types of phase transitions – those with and without changes of the first coordination number. Each of these types was further classified into reconstructive (*i.e.*, requiring formation/breaking of bonds), displacive, order–disorder, electronic *etc.* transitions. *Displacive* phase transitions are accompanied by small atomic displacements, which usually destroy some elements of symmetry. The displacement patterns are determined by the eigenvectors of the soft modes. *Electronic* transitions are accompanied by a change of the electronic or magnetic structure and/or valence state of atoms in the structure; the atomic structure type may or may not change in this process. *Reconstructive* transitions are always first order, involve a latent heat, and require an activation energy for bond breaking.

The Buerger classification, although based on purely structural considerations, naturally gives some insight into thermodynamics and kinetics of phase transitions. However, there is no general and rigorous way of determining coordination numbers of atoms in crystals. In addition to this, Buerger's classification does not discuss symmetry aspects of transitions.

It was first realised by Landau in 1937 (see Landau & Lifshitz, 1976) that often there are symmetry relations between the old and new phases. Landau has shown that for a second-order transition the two phases must be structurally related, and their symmetry groups must conform to certain group–subgroup relations. Symmetry plays a fundamental role in phase transitions, and better classifications must consider it; such a classification is proposed below[3]. Before considering this new classification, we briefly review some of the main theoretical ideas related to phase transitions.

Theoretical framework

First-order phase transitions

Equilibrium first-order phase transitions always involve activation, inhibiting their kinetics. The major factors determining the kinetics are defects, surface effects, and external fields. Polymorphs of Al_2SiO_5 (minerals kyanite, andalusite, and sillimanite, see Kerrick, 1990) are a classical example. All the transitions between these minerals are first-order reconstructive and require substantial activation energies; therefore, all the three minerals can coexist for millions of years in nature.

A very interesting example is the case of HgI_2 polymorphs (Hostettler *et al.*, 2001). At $p < 10$ GPa and $T < 600$ K, there are nine polymorphs in this system! Crystallisation from solution (which contains HgI_2 molecules) first produces a metastable yellow form, whose structure also consists of linear HgI_2 molecules, and only then metastable orange (containing supertetrahedra $[Hg_4I_{10}]$) and stable red (containing HgI_4 tetrahedra) forms. Metastable yellow and orange crystals begin to transform into the red phase when touched by a needle. Defects thus introduced serve as nucleation centres of the stable phase. The 'orange phase' is in fact represented by three distinct, but similar, structures, which can transform into one another. This system is a good example of the Ostwald rule,

---

[3] Another classification based on symmetry was proposed by Christy (1993).

which states that often crystallisation produces a metastable phase, which transforms into the stable one via a sequence of metastable phases. At high temperatures, another yellow molecular form is stable (it can crystallise from the vapour). At 1.3 GPa a new phase is formed, in which Hg atoms are fivefold coordinated. A somewhat similar kinetical behaviour was reported for dimethyl 3,6-dichloro-2,5-dihydroxyterephthalate (Richardson *et al.*, 1990). For that compound, transformation kinetics and actual transition temperatures were found to depend on such factors as crystal size, degree of perfection, types of defects, and heating rate. Again, defects play the role of nucleation sites of the new phase.

Thermodynamics of first-order transitions are based on the Clapeyron relation:

$$\frac{dp}{dT} = \frac{\Delta S}{\Delta V}, \tag{83}$$

where ΔS and ΔV are the entropy and volume differences between the phases. This relation is valid only for first-order transitions, because for continuous transitions both ΔV and ΔS are zero. The transition temperatures and pressures can be found from accurate atomistic (*e.g.*, Parker & Price, 1989: magnesium silicates) or quantum-mechanical total energy calculations (*e.g.*, Oganov & Brodholt, 2000: aluminium silicates), but there is no simple analytical theory. Only when the two phases are structurally similar can one apply approximate analytical theories, such as Landau theory.

A relation for second-order transitions, analogous to (Eqn. 83), was derived by Ehrenfest:

$$\frac{dp}{dT} = \frac{\Delta C_{\mathrm{p}}}{TV(\Delta \alpha)}, \tag{84}$$

where ΔC_{p} and $\Delta \alpha$ are the jumps of the heat capacity and thermal expansion at the transition. However, precise experiments, computer simulations, and accurate theories indicate that the heat capacity does not have a finite jump, but logarithmically diverges to infinity on both sides of the transition. This 'λ-behaviour' invalidates the Ehrenfest relation.

Landau theory of first- and second-order transitions

When the structural changes occurring upon transition are small, *i.e.* the two phases are structurally related, it is usually possible to define an order parameter (or several order parameters), whose continuous change describes all the intermediate structures on the transition pathway. The order parameter Q takes the value that minimises the free energy $F(Q)$ at given p–T conditions. The simplest expression for the free energy is the Landau potential:

$$F(Q){=}F_0 + \frac{1}{2} A(T{-}T_{\mathrm{C}})Q^2 + \frac{1}{3} BQ^3 + \frac{1}{4} CQ^4 + \dots, \tag{85}$$

where T_{C} is the critical temperature, and F_0 the free energy of the high-temperature

high-symmetry phase. The entropy is simply $S(Q) = -\partial F(Q)/\partial T = S_0 - \dfrac{1}{2} AQ^2$. This

dependence of entropy on the order parameter is most appropriate for displacive phase

transitions[4]. The internal energy is then $E(Q) = E_0 - \dfrac{1}{2} AT_C Q^2 + \dfrac{1}{3} BQ^3 + \dfrac{1}{4} CQ^4 + \dots$

For second-order transitions the odd-order terms in Equation 85 must be zero (*e.g.*, $B = 0$), making the double well symmetric[5]. When $B \neq 0$, the double well is asymmetric, and the transition is the first order. First-order transitions appear whenever any odd-order term is present in Equation 85 and can also appear when $C < 0$ (even if only even-order terms are present). Landau theory correctly predicts mode softening for second-order transitions and hysteresis for first-order transitions.

    The power of Landau theory comes from the fact that simple symmetry considerations allow one to find which of the coefficients B, C, … must be zero, and therefore (see below) what is the order of the transition and what are the possible couplings of order parameters to strains (see below). Second-order phase transitions, corresponding to a symmetrical double-well potential $E(Q)$, are always characterised by group–subgroup relations: the symmetry group of one (often called 'ordered', usually low-temperature) phase is a subgroup of the symmetry group of the other ('disordered', usually high-temperature) phase. The two symmetrically equivalent minima then correspond to the same ordered phase, and can be considered as 'twin domains', related by a symmetry element present in the disordered phase, but absent in the ordered one (Fig. 7).

    The potential (85) is often complicated by the presence of several order parameters or by couplings of the order parameter to lattice strains. Such coupling can cause a first-order behaviour even for a symmetric $E(Q)$ potential. For details see the review by Carpenter (2002), this volume. For a detailed general account of Landau theory see Landau & Lifshitz (1976) and Dove (1993, 1997).

    Finally, we note that Landau theory is also applicable to pressure-driven phase transitions. The Landau potential analogous to Equation 85 is:

$$G(Q) = F_0 + \frac{1}{2} A(p - p_C)Q^2 + \frac{1}{3} BQ^3 + \frac{1}{4} CQ^4 + \dots \ , \qquad (86)$$

implying $V(Q) = \dfrac{\partial G(Q)}{\partial p} = V_0 + \dfrac{1}{2} AQ^2$. The Helmholtz free energy is then

$$F(Q) = F_0 - \frac{1}{2} Ap_C Q^2 + \frac{1}{3} BQ^3 + \frac{1}{4} CQ^4 + \dots$$

---

[4] For order–disorder transitions the entropy is more accurately expressed as $S(Q) = S_0 + k_B[(1 + Q)\ln(1 + Q) + (1 - Q)\ln(1 - Q)]$. The simplest mean-field model of order–disorder transitions with this expression for the entropy is called the Bragg–Williams model.

[5] Absence of odd-order terms is only one of the necessary conditions. Other conditions were formulated by Birman (1966) by using group theory. All these conditions are necessary, but not sufficient.
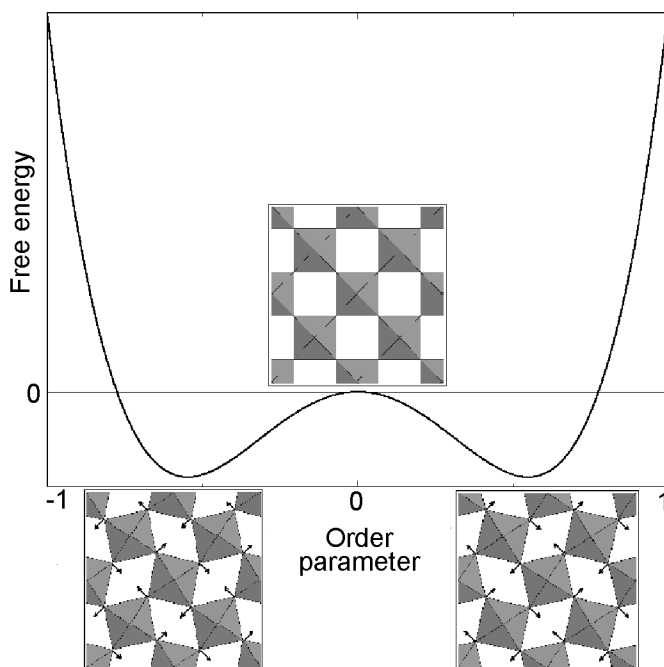
Fig. 7. Landau potential $F(Q) = F_0 + \frac{1}{2}A(T - T_C)Q^2 + \frac{1}{4}CQ^4$ at $T < T_C$, where distortions stabilise the structure. The two distorted perovskite-type structures shown on the bottom are equivalent (they are mirror reflections of each other) and differ only by the sense of the distortion. The arrows show the directions of octahedral rotations away from the cubic structure. The undistorted structure is shown in the centre. At temperatures higher than T_C, the stable structure will be locally distorted, but on average will have the symmetry of the undistorted structure.

Shortcomings of Landau theory

Landau theory is a mean field theory. Mean field treatment is a common way of approximately solving complex physical problems in many areas of science (*e.g.*, Hartree and Hartree–Fock methods for electronic structure calculations are mean field approximations). The main drawback of these methods is the neglect of fluctuations (in Landau theory the local structure and fluctuations of the order parameter are neglected). Landau theory assumes that all the neighbouring unit cells have the same configuration; domain structures and fluctuations of the order parameter in space and time are not treated properly. This problem becomes severe in the vicinity of T_C (so-called Ginzburg interval). For second-order transitions Landau theory predicts $Q \approx (T_C - T)^{1/2}$, while experiments show $Q \approx (T_C - T)^{1/3}$. The critical exponent of 1/3 has been many times confirmed by numerical computer simulations and renormalisation group theory[6].

---

[6] Experiments do give mean-field critical exponents far from T_C, but nearer T_C there is a crossover from the mean-field to critical behaviour, where the critical exponents depart significantly from the mean-field predictions.

Landau theory cannot explain the logarithmic divergence of the heat capacity near the critical point – instead, it yields a finite jump. Finally, it does not consider quantum effects at low temperatures and does not reproduce experimentally observed order parameter saturation at low temperatures, but predicts a steady increase of the order parameter with decreasing T. Nevertheless, it is generally believed that the symmetry apparatus of Landau theory is valid even where mean-field theory breaks down. In 1950, V.L. Ginzburg and L.D. Landau (see Landau & Lifshitz, 1976; Bowley & Sánchez, 1999) considered the case of an order parameter slowly varying in space, which led to the formulation of the simplest theory beyond the mean field. Ginzburg–Landau theory emphasises the role of critical fluctuations and defines criteria of validity of Landau theory. However, this theory is still approximate and does not reproduce experimental critical exponents.

Renormalisation group theory overcomes all these difficulties and serves as the modern basis of the theory of critical phenomena. It correctly predicts the critical exponents, explains their universal character, strong dependence on the dimensionality of the system and independence on the type of chemical bonding. The divergence of the heat capacity at T_C for second-order transitions is explained and some new phenomena are predicted, such as 'continuous lattice melting'. So far, only one example of continuous lattice melting has been found experimentally: Na_2CO_3 (Harris & Dove, 1995). Introductory texts on renormalisation group theory can be found in Chandler (1987), Rao & Rao (1978), and Wilson (1983); the latter reference is the Nobel lecture of Kenneth Wilson, the main inventor of this theory.

Ising spin model

This model is widely used to describe magnetic and atomic ordering processes in materials. In this model, a spin +1 or –1 is associated with each lattice site, depending on whether the magnetic moment on the site is 'up' or 'down', or whether the atom occupying the site is of the type '*A*' or '*B*'.

The total energy of the system is:

$$E = E_0 - J\sum_{i,j} S_i S_j - H\sum_i S_i, \tag{87}$$

where E_0 is the reference-state energy, and J is the interaction parameter between the sites: if $J < 0$, unlike spins prefer to group together, and there is a tendency to ordering at low temperatures; if $J > 0$, unmixing will occur at low temperatures. Disorder, although unfavourable energetically, is stabilised by the entropy at high temperatures. An external field H leads to a preferred orientation of the spins. The Ising model can be analytically solved only in one or two dimensions; in three dimensions it is solved numerically, usually by the Monte Carlo method. Peculiarly, one-dimensional Ising model exhibits no phase transitions, and at all temperatures above 0 K is in the disordered state.

Simulations based on the Ising model yield correct critical behaviour and critical exponents. The Ising model can be generalised for the case of more than two spins (see Yeomans, 1992). Another extension, designed to treat spin vectors, is called the Heisenberg model.

Ising-like models provide an interesting route for theoretical studies of polytypism and polysomatism. Figure 8 shows the structures of ringwoodite and wadsleyite, which can be considered as polytypic, differing in the stacking sequence of layers. These differently oriented layers can be mapped onto the sites of one-dimensional Ising lattices. For details see papers by Price (1983), Price & Yeomans (1984), and references therein. Price *et al.* (1985) used atomistic simulations to determine the interaction parameters for the Ising models and predict the possible polytypes, stacking faults, and their energetics.
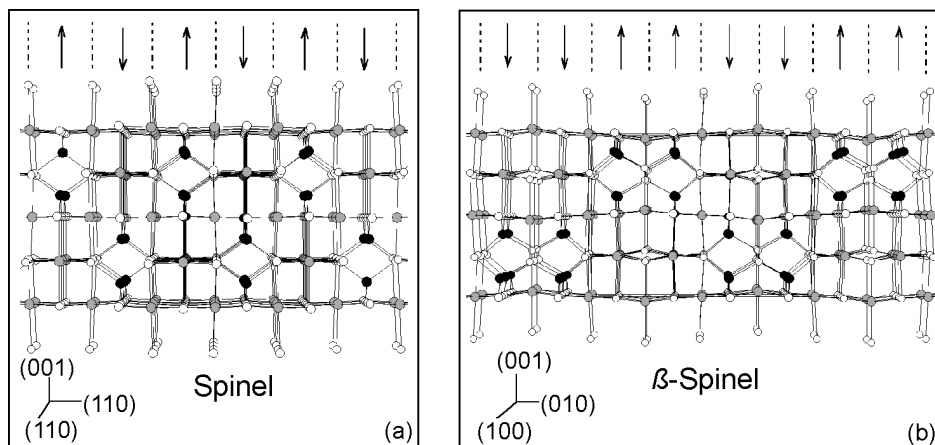
Fig. 8. Illustration of polytypism in spinelloids. *(a)* Spinel-like ringwoodite Mg_2SiO_4, *(b)* wadsleyite (*β*-spinel phase) Mg_2SiO_4. Intermediate polytypic sequences can also occur in stacking faults. Black circles are Si atoms, grey circles are Mg atoms, and empty circles are O atoms. Layers in two different orientations are denoted by 'up' and 'down' arrows.

New classification of phase transitions

Taking into account the fundamental role of symmetry, we suggest a new classification of phase transitions, given in Table 3. We divide all phase transitions into '*global*' (there are no structural relations between the old and new phases and the new phase is always the thermodynamically stable one, *i.e.* corresponds to the *global* free energy minimum), and '*local*' (the new phase corresponds to a *nearest* free energy minimum, which implies clear structural and symmetry relations; the new phase can be either thermodynamically stable or metastable).

Global transitions can involve any structural and symmetry changes. These transitions are strongly first-order and reconstructive, require substantial activation and cannot proceed at low temperatures. They occur via the nucleation-and-growth mechanism, and the nuclei of the new phase are usually formed near defects or surfaces.

Local transitions, on the contrary, can be continuous as well as first-order (often the first-order component is small). They can be reconstructive, displacive, order–disorder, or electronic. Often there are clear orientational relations (twinning, topotaxy) between the

Table 3. Classification of phase transitions. See Figure 9 for illustrations and text for some references

I. GLOBAL TRANSITIONS.	II. LOCAL TRANSITIONS.
I.1. With changes of the first co-ordination numbers	**II.1. Isosymmetric transitions**
Al_2SiO_5: kyanite–andalusite–silli-manite;	*electronic* – Ce
SiO_2: coesite–stishovite	*structural* – $KTiOPO_4$, PbF_2, KNO_3, sillimanite (Al_2SiO_5)
I.2. Without changes of the first coordination numbers	*structural with ordering change* – Ni–Al alloys, albite ($NaAlSi_3O_8$)
SiO_2: quartz–coesite	*structural & electronic* – Na_3MnF_6
	II.2. Group-subgroup relations.
	translationengleich – α – β-quartz (SiO_2)
	klassengleich – $P2_1/c$–$C2/c$ pyroxenes (*e.g.*, $MgSiO_3$)
	general – $Pm\overline{3}m$–$Pbnm$ perovskites (*e.g.*, $MgSiO_3$)
	indirect relations – rhombohedral–tetragonal $BaTiO_3$ perovskite; bcc–fcc Fe
	II.3. Incommensurate transitions
	1-dimensional – sillimanite (Al_2SiO_5)
	2-dimensional – quartz (SiO_2)
	3-dimensional – wüstite ($Fe_{1-x}O$)
	II.4. Crystal–quasicrystal transitions
	II.5 Pressure-induced amorphisation
	SiO_2 (quartz, coesite), berlinite ($AlPO_4$), andalusite (Al_2SiO_5)

old and new phases. These transitions may require no activation and, therefore, in some cases can proceed at very low (even absolute zero) temperatures. According to the degree of symmetry changes, there are several possible cases (Table 3).

Isosymmetric transitions
In this simplest case, there is no symmetry change upon transition. Using Landau theory, it has been shown (Bruce & Cowley, 1981; Christy, 1995) that isosymmetric transitions must be first-order, but disappear (become fully continuous, infinite-order transitions) above the critical point: there is a complete analogy here with the liquid–gas transitions (which are also isosymmetric). Isosymmetric transitions in crystals can be electronic (*e.g.*, Ce and SmS), structural (coordination numbers of atoms change – *e.g.*, $KTiOPO_4$) or intermediate (both electronic and structural changes are involved – *e.g.*, Na_3MnF_6). Another kind of structural isosymmetric transitions associated to the change of the ordering of atoms or defects was recently predicted for the Al-rich NiAl alloy (Alavi *et al.*, 1999) and albite $NaAlSi_3O_8$ (Oganov *et al.*, 2001c).

Metallic Ce undergoes an isosymmetric electronic phase transition Ce(I)–Ce(IV) (see Liu & Bassett, 1986, and references therein). Both Ce(I) and Ce(IV) have the fcc structure (space group $Fm\overline{3}m$). The volume change at the transion is very large (13%) at room temperature, but it rapidly decreases along the Ce(I)-Ce(IV) equilibrium line until it disappears at the critical point (2.15 GPa and 613 K). Another example of an electronic transition is SmS, which transforms from the low-pressure insulating phase to the high-pressure metallic phase; both have a cubic NaCl-type structure (see Fig. 10).

$KTiOPO_4$ (KTP) undergoes a structural isosymmetric transition at 5.8 GPa with a volume decrease of 2.7% and preservation of space group $Pna2_1$ (Alan & Nelmes, 1996). $KNO_3(II)$–$KNO_3(IV)$ isosymmetric (*Pnma*) transition, which occurs at 0.3 GPa, is accompanied by a volume decrease of 11.5% (Adams *et al.*, 1988) and drastic changes
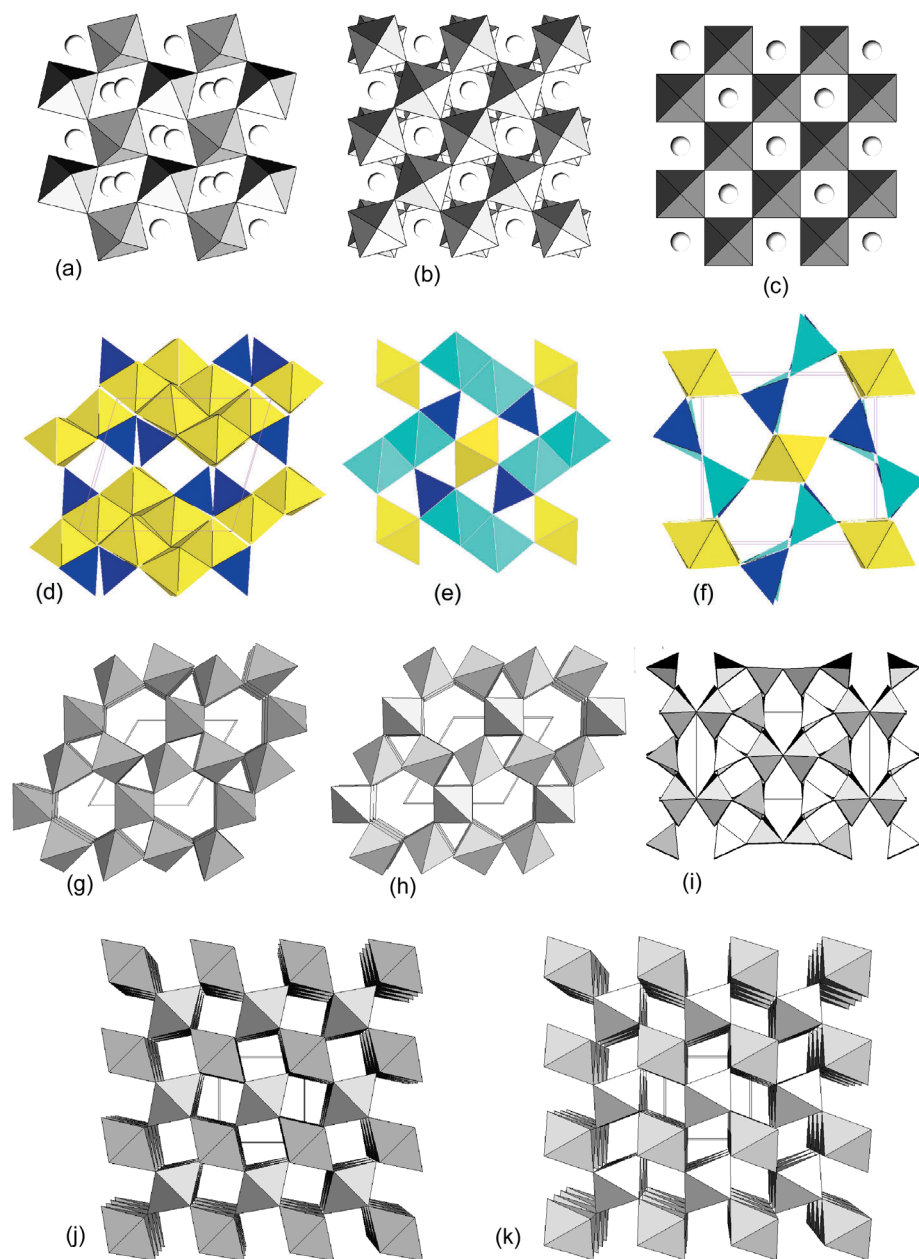
Fig. 9. Examples of different phase transitions. *(a–c)* perovskite phases (*ABX*₃): *(a) Pbnm*, *(b) I*4/*mcm*, *(c) Pm*3̄*m*. Spheres are *A* cations, polyhedra are BX_6 octahedra. *(d–f)* Al₂SiO₅ polymorphs: *(d)* kyanite, *(e)* andalusite, *(f)* sillimanite. Al polyhedra are shown in yellow and turquoise, Si polyhedra are blue. *(g–k)* silica polymorphs: *(g)* α-quartz, *(h)* β-quartz, *(i)* coesite, *(j)* stishovite, *(k)* CaCl₂-type structure (drawn with an exaggerated orthorhombic distortion).
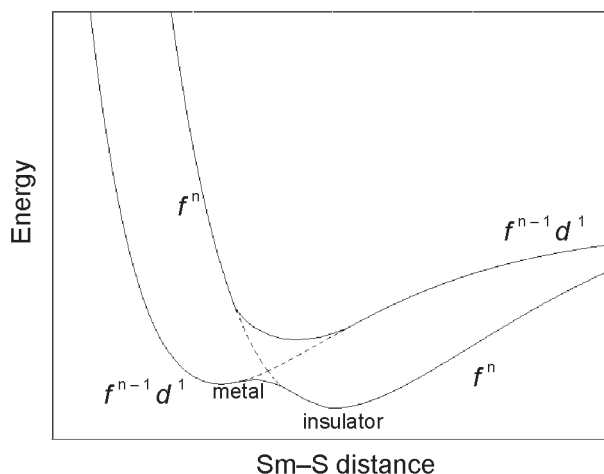
Fig. 10. Illustration of the isosymmetric insulator–metal transition in SmS. After Burdett (1995), with modifications. Mixing of two configurations (metallic $f^{n-1}d$ and insulating f^n) produces a double-well energy curve for the ground state, where the minimum with a smaller interatomic distance corresponds to a metal. Compression triggers the insulator–metal transition.

in the Raman spectra. K atom coordination number changes from 9 to 11. A first-order isosymmetric (space group *Pnam*) phase change has been observed at 9.8 GPa for PbF_2 (Haines *et al.*, 1998). In their molecular dynamics simulations, Badro *et al.* (1996, 1997) found that quartz (SiO_2) transformed into an isosymmetric (space group $P3_221$) phase with all silicon atoms in a 5-fold coordination when subjected to a uniaxial stress (20 GPa along the *a* and *b* axes and 27 GPa along the *c* direction). Na_3MnF_6 (space group $P2_1/n$) is an example of a phase transition with a simultaneous change in the atomic and electronic structure. At 2.2 GPa this compound undergoes a first-order isosymmetric phase transformation, which is associated with a change of orientation of the Jahn–Teller elongation of MnF_6 octahedra (Carlson *et al.*, 1998). This transformation shows a small hysteresis of 0.5 GPa.

Transitions with group–subgroup relations
Next degree of symmetry change involves group–subgroup relations between two crystalline phases. This is the most common type of local transitions.

Different types of group–subgroup relations exist. According to the Hermann theorem (Hahn & Wondratshcek, 1994), a maximal subgroup of a space group is either *translationengleich*[7]-type (*i.e.*, all translations are preserved, but some of the point-group symmetry operators disappear) or *klassengleich*[8]-type (all point-group operators are preserved, but some of the lattice translations disappear). Subgroups which are neither

---

[7] German word meaning 'having the same translations'; accepted as an international term *(International Tables for Crystallography)*

[8] German word meaning 'having the same (point-group) class'; accepted as an international term *(International Tables for Crystallography)*

klassengleich, nor translationengleich are called *general*; they involve both cell multiplication and lowering of the point-group symmetry. Examples are $\alpha \rightarrow \beta$-quartz (*translationengleich*), $P2_1/c$–$C2/c$ pyroxenes (*klassengleich*; Arlt & Angel, 2000 and Arlt *et al.*, 1998), and $Pbnm \rightarrow Pm\bar{3}m$ transitions in perovskites (*general*).

The fourth type of group–subgroup transitions involves indirect symmetry relations between the two phases via an intermediate archetypal phase of a higher symmetry, which is a supergroup for symmetries of both phases. An example is $BaTiO_3$, where the transition between the rhombohedral and tetragonal phases can be described with reference to the higher-symmetry cubic phase. Such transitions are usually weakly first-order.

Yet another possibility is that of a transition state of a lower symmetry, which is a common subgroup of the symmetries of both phases. These transitions are usually strongly first-order; often they can be described as reconstructive (see Christy, 1993). The fcc $\rightarrow$ bcc transition in Fe can be described with reference to lower-symmetry tetragonal or rhombohedral configurations, whose symmetries are common subgroups of both symmetry groups of the bcc and fcc phases. On the basis of purely geometrical considerations, Sowa (2000 and personal communication) concluded that the transition from the zinc-blende to the NaCl-type is most likely to proceed via an orthorhombic transition state. *Ab initio* calculations of Catti (2001) for SiC support this conclusion.

Incommensurate transitions

Further degree of symmetry change involves incommensurate modulation of the structure and loss of translational symmetry along the modulation vector, while preserving both short- and long-range order. Incommensurate modulation (and the corresponding loss of translational symmetry) can occur in one, two or three dimensions. Incommensurability can occur via mode softening or modulation of atomic positional occupancies, or in metals via electronic effects (*e.g.*, Peierls transition). Among the examples are $NaNO_2$, some feldspars, nepheline $(K,Na)AlSiO_4$, åkermanite $Ca_2MgSi_2O_7$, calaverite $AuTe_2$, wüstite $Fe_{1-x}O$; there is an incommensurate phase in quartz, intermediate between α- and β-phases. Incommensurate modulation can occur in 1, 2, or 3 dimensions. A different type of incommensurate phases can be described as misfit structures. We redirect the reader to an excellent review (van Smaalen, 1995) and some interesting papers – on theory of incommensurate transitions in insulators (Heine & McConnell, 1981) and unusual crystal morphology of incommensurate phases (Dam *et al.*, 1985).

Crystal–quasicrystal transitions

Quasicrystals, like incommensurate phases, are solids lacking crystal periodicity (at least in one direction), and having both long- and short-range order, but also non-crystallographic (5-, 8-, 10-, 12-fold) symmetry axes. Several types of quasicrystals are known: icosahedral quasicrystals with no lattice periodicity at all ('3D quasicrystals'), and quasicrystals with 1-dimensional periodicity ('2D quasicrystals') and dodecagonal (with 12-fold symmetry axes), decagonal (5-fold and 10-fold axes), or octagonal (8-fold axes) symmetry. 1D quasicrystals are possible as well.

Fig. 11. Penrose tiling for a decagonal quasicrystal. White and black tiles are different 'unit cells'. The ratio of white tiles to the black ones is the golden mean $\tau = (\sqrt{5} + 1)/2$.
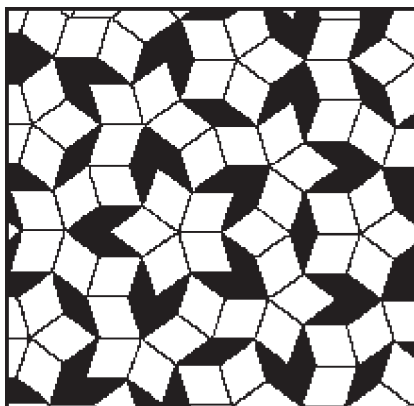
The first quasicrystal ($Al_{86}Mn_{14}$) was discovered quite recently (Shechtman *et al.*, 1984); now we know a large number of different quasicrystals, some theories of their structure and formation have been formulated, and even some computational studies performed. For a review, see Janot (1994).

It is likely that crystal–quasicrystal transitions are local. However, an ideal crystal and ideal quasicrystal are not very likely (although can in principle) to have the same composition. The structure of 2D and 3D quasicrystals can be described in terms of Penrose tilings with two or more types of 'unit cells'[9] (Fig. 11). Since the proportion of these 'unit cells' in a Penrose tiling is irrational, general quasicrystal composition will be characterised by an irrational chemical formula (see Janot, 1994). This explains both the extremely complicated chemical formulae of quasicrystals and their occurrence exclusively among metallic alloys (ionic or covalent solids cannot have an irrational chemical formula, as this would break charge neutrality).

Pressure-induced amorphisation

The ultimate degree of symmetry breaking is present in pressure-induced amorphisation. Lattice periodicity and long-range order are destroyed in one or more dimensions, while short-range order is retained. All symmetry elements are locally destroyed, but when viewed as a continuum, the amorphous phase has spherical point-group symmetry[10]. In the present context it is more appropriate to classify pressure-induced amorphisation as a symmetry-lowering process, as any process involving soft modes and modulations of the structure.

This phenomenon was discovered rather recently (Mishima *et al.*, 1984) in compression experiments on ice. It has attracted much attention, but many questions still remain open. See excellent reviews (Sharma & Sikka, 1996; Richet & Gillet, 1997) for more details.

---

[9] However, this description is not perfect and the best description of quasicrystals is given in terms of cuts of higher-dimensional lattices (see Janot, 1994).

[10] Pressure-induced amorphisation can produce anisotropic glasses, whose point-group symmetry is ellipsoidal (see McNeil & Grimsditch, 1991).

Many cases of pressure-induced amorphisation are known[11], as well as a few substances undergoing pressure-release amorphisation (for extensive reviews, see Sharma & Sikka, 1996; Richet & Gillet, 1997). In the latter phenomenon, some high-pressure phases become dynamically unstable and amorphise when decompressed to pressures below their stability fields[12].

Pressure-induced amorphisation is a metastable first-order transition. The amorphous phase must be denser than the original crystalline phase. This means that crystals with open structures should be easier to amorphise.

For ice, the two amorphous phases (one produced by cooling water, the other by pressurising ice) are separated by a first-order phase transition (Mishima *et al.*, 1985). This indicates clear structural differences between quenched and pressure-induced glasses.

Kruger & Jeanloz (1990) found that berlinite ($AlPO_4$) exhibits the so-called 'memory glass effect': the amorphous phase, appearing under compression of a single crystal of berlinite under 18 GPa, on decompression to 5 GPa turns back to the single crystal with the same crystallographic orientation. However, Gillet *et al.* (1995) argued that berlinite transforms into an intermediate strongly disordered crystalline phase at 15 GPa, and true amorphisation occurs at higher pressures and is irreversible.

Pressure-induced amorphisation is thought to be driven by the simultaneous or nearly simultaneous softening of a large part of a phonon branch; freezing in all these modes and their combinations destroys translational periodicity and creates random static atomic displacements. Simultaneous softening of a phonon branch along a direction in the Brillouin zone implies weak dispersion of this branch, which is most naturally achieved when the unit cell is large. Indeed, crystals with complicated structures and large unit cells are more prone to pressure-induced amorphisation.

Discussion of the new classification

The classification presented above is based on symmetry and naturally incorporates the notions of first-order and continuous, reconstructive, displacive, order–disorder, and electronic phase transitions. Its principles are very simple, but like any classification it has some elements of arbitrariness. The arbitrariness of this classification is in the distinction between local and global transitions; the symmetry classification of local transitions, on the other hand, is mathematically rigorous. The present classification emphasises the fundamental role of isosymmetric polymorphism and pressure-induced amorphisation as the two extreme cases of symmetry changes.

---

[11] Quartz SiO_2, coesite SiO_2, berlinite $AlPO_4$, GeO_2, zeolites scolecite $Ca_8Al_{16}Si_{24}O_{80} \cdot 24H_2O$ and mesolite $Na_{16}Ca_{16}Al_{48}Si_{72}O_{240} \cdot 64H_2O$, anorthite $CaAl_2Si_2O_8$, forsterite Mg_2SiO_4, fayalite Fe_2SiO_4, wollastonite $CaSiO_3$, enstatite $MgSiO_3$, muscovite $KAl_3Si_3O_{10}(OH)_2$, portlandite $Ca(OH)_2$ are just a few examples.

[12] This happens to the perovskite-structured modification of $CaSiO_3$, stable in the Earth's lower mantle, but at ambient conditions turning to a glass within a few hours.

Part IV: Simulation methods

In this section we mainly describe how to calculate the energy of a crystal quantum-mechanically. Being able to calculate the energy, one can calculate a number of properties straightforwardly. These include the equilibrium crystal structure, thermodynamic functions, equations of state, elastic constants, and many response functions. Semiclassical simulations, which (unlike *ab initio* simulations) are based on parameterised models of interatomic interactions, are discussed elsewhere in this volume and in a number of sources (*e.g.*, Catlow & Mackrodt, 1982; Allen & Tildesley, 1987; Urusov & Dubrovinsky, 1989). The advantage of *ab initio* simulations is that they do not involve any parameters that have to be fitted to experiment. *Ab initio* computational methods considered here represent major advances in modern theoretical physics.

General references for the interested reader include several books (Born & Huang, 1954; Venkataraman *et al.*, 1975; Catlow & Mackrodt, 1982; Allen & Tildesley, 1987; Dove, 1993; Sutton, 1993; Thijssen, 1999) and review articles (Catlow & Price, 1990; Catlow *et al.*, 1994; Bukowinski, 1994; Cohen, 1999; Hemley & Cohen, 1996; Gale, 1996; Stixrude *et al.*, 1998; Kohn, 1999a, 1999b).

General remarks

In equilibrium, any system attains the state with the lowest Gibbs free energy G at given p–T conditions. All equilibrium structural and electronic (wavefunction) parameters can be determined by minimising $G = E + pV - TS$. Many ground-state properties can then be obtained from G. This general task is very difficult and is solved with several reasonable approximations.

The first one is the Born–Oppenheimer approximation. Since atomic nuclei are much (thousands of times) heavier than electrons, nuclear motion will be much slower than electronic. In the Born–Oppenheimer approximation, the electrons instantaneously relax to the ground state every time an atom is displaced. This allows one to consider the quantum-mechanical problem in the framework of fixed nuclei, whose positions determine the ground-state electronic wavefunction. This approximation is usually very accurate.

The next step is to calculate the energy by solving the electronic Schrödinger equation. At this step, we usually neglect relativistic effects, which are, however, important for elements heavier than Kr[13]. Even the non-relativistic Schrödinger equation can be solved only approximately for solids.

---

[13] Relativistic effects appear when electrons are moving at speeds comparable to the speed of light. This occurs in heavy atoms, where the inner electrons move in a very strong nuclear potential. Two main types of relativistic effects are scalar relativistic effects (shrinking of the innermost atomic orbitals, which also affects the potential acting on the valence electrons) and spin–orbit interaction. To see how scalar relativistic effects occur (Pisani, 1996), note that the average velocity of a $1s$ electron in atomic units (a.u.) is Z (nuclear charge), and the speed of light is 137.036 a.u. For Hg ($Z = 80$) the relativistic mass of a $1s$ electron is 1.23 times its rest mass. This corresponds to a $\sim 23\%$ shrinking of the $1s$ orbital in Hg, since the Bohr radius ($a_0 = \hbar^2/me^2$) is inversely proportional to the mass of the electron. Shrunk core orbitals more effectively screen outer orbitals; valence s- and p-electrons become contracted and stabilised relative to the non-relativistic case, whereas valence d- and f-electrons become more expanded, destabilised, and thus more chemically active. Spin–orbit interaction becomes important in heavier atoms (roughly after Xe); in solids it can remove degeneracies of electron levels at high-symmetry points of the reciprocal space (Ashcroft & Mermin, 1976). For a fully relativistic treatment, one must solve the Dirac (rather than Schrödinger) equation. For a detailed review of relativistic effects in chemistry, see Pyykko (1988).

Having calculated the energy and its dependence on structural parameters, we can calculate the partition function Z (Eqn. 10) and the free energy F (Eqn. 11). To do that one needs to solve the Schrödinger equation for atomic motion and find all the energy levels in the system. This approach is impractical, and two most common ways of avoiding the direct solving of the nuclear Schrödinger equation: are (i) to use the classical approximation for nuclear motion (as in molecular dynamics and Monte Carlo simulations; this approach is valid at high temperatures) and (ii) to solve the Schrödinger equation assuming that atomic vibrations are small and the system is approximately harmonic (this approach is used in lattice dynamics simulations and is valid at relatively low temperatures). All these methods are discussed extensively in the references given above; lattice dynamics is discussed by Gramaccioli (2002) and Choudhury *et al.* (2002) in this volume.

Schrödinger equation

The total energy of a non-relativistic electron–nuclear system and all its energy levels can be calculated by solving the Schrödinger equation:

$$\hat{\mathbf{H}}\psi = E\psi, \tag{88}$$

where $\hat{\mathbf{H}}$ is the Hamilton operator (Hamiltonian) of the system. Writing $\hat{\mathbf{H}}$ more explicitly, in atomic units[14] we have:

$$\left[\sum_{i=1}^{N}\left(-\frac{1}{2}\nabla_i^2 \right) + \sum_{i=1}^{N} v_{e-n}(\mathbf{r}_i) + \sum_{i=1}^{N} \frac{1}{\left| \mathbf{r}_i - \mathbf{r}_j \right|} \right] \psi = E\psi, \tag{89}$$

where N is the number of electrons. The first term in Equation 89 is the kinetic energy operator and the next terms are the electron–nuclear and electron–electron potential energy operators, respectively; E is the electronic energy (to find the total energy, one needs to add to it the nuclear–nuclear interaction energy). Schematically, we can write Equation 89 as follows:

$$\left(-\frac{1}{2}\nabla^2 + v_{e-n} + v_{e-e} \right)\psi = E\psi. \tag{90}$$

The Hamiltonian is in principle very simple, having only Coulombic interactions. The main problem is that the wavefunction is a function of $3N$ electronic coordinates (where N is the number of electrons) – for many-electron systems it is exceedingly complicated! The common route is to use approximate Hamiltonians and simplified wavefunctions (in such radical density functional approaches as the Thomas–Fermi model, the wavefunction does not appear at all!). All modern computational methods are based on approximate one-electron theories, reducing Equation 89 to a set of equations

---

[14] In atomic units, $\hbar = m_e = e^2 = 4\pi\varepsilon_0 = 1$. The units of length and energy in this system are Bohr $a_0 = 0.529$ Å and Hartree (equal to minus twice the energy of the hydrogen atom): 1 Ha = 27.2116 eV. Atomic units are used in all quantum-mechanical equations here.

for one-electron orbitals. Density functional theory of Kohn & Sham (1965) is a formally exact one-electron theory.

Before moving on to approximate theories, we mention a few important exact results. The first result relates to the existence of a cusp of the electron density at nuclear positions, with the average derivative of the density at the nucleus ($r = r_0$):

$$\frac{\partial \overline{\rho}(r_0)}{\partial r_0} = -2Z\overline{\rho}(r_0),$$

(91)

where Z is the nuclear charge. Far from the nucleus, atomic electron density falls off exponentially:

$$\rho(r) \sim \exp[-2(2I)^{1/2}r],$$

(92)

where I is the ionisation potential. This relation is also true for molecules (Parr & Yang, 1989).

Electrons, as quantum particles, are never at rest – even at $T = 0$ K they have a large zero-point energy with a large kinetic component. For Coulombic systems (like atoms, molecules, crystals) the virial theorem (Landau & Lifshitz, 2001) states that the total electronic energy is equal to minus the kinetic energy of the electrons or to one half of the potential energy of the electrons:

$$E = -\langle E_{\text{kin}} \rangle = \frac{1}{2}\langle E_{\text{pot}} \rangle.$$

(93)

The virial theorem is exact only for an exact wavefunction.

If E is a function of some parameter q, then its derivative is:

$$\frac{dE}{dq} = \langle \psi \mid \frac{d\hat{\mathbf{H}}}{dq} \mid \psi \rangle + \left[\langle \frac{d\psi}{dq} \mid \hat{\mathbf{H}} \mid \psi \rangle + \langle \psi \mid \hat{\mathbf{H}} \mid \frac{d\psi}{dq} \rangle \right].$$

(94)

The most important applications of Equation 94 are the calculation of forces (in which case q is an atomic coordinate) and stresses, which are used for geometry optimisation in quantum-mechanical calculations. For an exact wavefunction, the terms in square brackets in Equation 94 sum to zero, and

$$\frac{dE}{dq} = \langle \psi \mid \frac{d\hat{\mathbf{H}}}{dq} \mid \psi \rangle.$$

(95)

Equation 95 is known as the Hellmann–Feynman theorem. A consequence of this theorem is that forces acting on nuclei can be calculated from the classical electrostatic expression involving nuclear positions and the electron density distribution. When the wavefunction is not exact (*e.g.*, due to basis set incompleteness), the terms in square brackets in Equation 94 do not sum to zero and must be taken into account. These terms (with a minus sign) are known as the Pulay force (or Pulay stress).

Hartree method

This method is not used in practical calculations; the only reason to consider it now is that it was the starting point of the Hartree–Fock and Kohn–Sham methods. Calculating the interelectronic interaction energy, Hartree considered only the classical Coulombic energy of the inhomogeneous electron gas (this term is known as the Hartree energy E_H):

$$E_{\mathrm{H}} = \frac{1}{2}\int \rho(\mathbf{r})v_{\mathrm{H}}(\mathbf{r})d\mathbf{r} = \frac{1}{2}\int \rho(\mathbf{r})d\mathbf{r}\int \frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}', \qquad (96)$$

The wavefunction was cast in the independent-particle approximation:

$$\psi = \phi_1\phi_2\phi_3\ldots\phi_{\mathrm{n}} = \Pi\phi_i, \qquad (97)$$

where ϕ_i are the one-electron orbitals. The electronic kinetic energy is the sum of the one-electron terms:

$$T_{\mathrm{s}} = \sum_i -\frac{1}{2}\nabla_i^2. \qquad (98)$$

The Schrödinger equation is transformed into a set of one-electron equations:

$$\varepsilon_i\phi_i(\mathbf{r}) = \left[-\frac{1}{2}\nabla^2 + v_{\mathrm{e-n}}(\mathbf{r}) + \int \frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}' \right]\phi_i(\mathbf{r}). \qquad (99)$$

Even though the wavefunction in Equation 97 is inconsistent with the Pauli principle, the Hartree method (surprisingly) gives many reasonable results. It reproduces the shell structure of atoms and gives semiquantitatively correct atomic electron densities. All this is due to the success of the independent-particle approximation to the kinetic energy (Eqn. 98): the Thomas–Fermi method, which differs only in that it uses the electron-gas expression for the kinetic energy, fails to reproduce the shell structure of atoms.

The electrons are totally independent and uncorrelated in the Hartree approximation and the probability of simultaneously finding electrons 1 and 2 in points $\mathbf{r}_1$ and $\mathbf{r}_2$, respectively, is:

$$P(\mathbf{r}_1, \mathbf{r}_2) = |\phi_1(\mathbf{r}_1)|^2 |\phi_2(\mathbf{r}_2)|^2. \qquad (100)$$

This means that there is a finite probability P for the two electrons to be at the same point at the same time. This is clearly wrong, as the electrostatic repulsion will keep the electrons apart. This repulsion introduces correlations between the motions of different electrons and lowers the interelectronic energy relative to the Hartree energy (Eqn. 96). The quantum effects beyond the Hartree approximation are called exchange and correlation. In the Hartree approximation, the electrons interact electrostatically with the average total electron density distribution. The energy includes the spurious interaction of each electron with itself – the self-interaction error.

Hartree–Fock (HF) method

Fock improved the Hartree method by taking into account the electronic exchange energy and antisymmetry of the wavefunction (Pauli principle). The Hartree–Fock equations take the form:

$$\hat{\mathbf{f}}\phi_i(\mathbf{r}) = \varepsilon_i \phi_i(\mathbf{r}), \tag{101}$$

where ε_i are the Hartree–Fock eigenvalues (in the approximation known as the Koopmans' theorem, ε_i are equal to minus the orbital ionisation energies). The operator $\hat{\mathbf{f}}$ is known as the Fock operator. Re-written explicitly, Equation 101 takes the form:

$$\varepsilon_i \phi_i(\mathbf{r}) = \left[-\frac{1}{2}\nabla^2 + v_{e-n}(\mathbf{r}) + \int \frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}' \right]\phi_i(\mathbf{r}) - \sum_j \delta_{\sigma i, \sigma j} \int d\mathbf{r}' \frac{\phi_j^*(\mathbf{r}')\phi_i(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\phi_j(\mathbf{r}), \tag{102}$$

which includes the Hartree potential and the (exact) exchange potential (in the last term). The exchange contribution dramatically improves the agreement with experiment and completely cancels the self-interaction error. The exchange potential is non-local (it is determined by the wavefunction in the whole space), which renders the HF method computationally expensive.

The HF wavefunction of an *N*-electron system is a single Slater determinant composed of one-electron wavefunctions:

$$\psi_{HF} = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(\mathbf{r}_1,\sigma_1) & \phi_1(\mathbf{r}_2,\sigma_2) & \dots & \phi_1(\mathbf{r}_N,\sigma_N) \\ \phi_2(\mathbf{r}_1,\sigma_1) & \phi_2(\mathbf{r}_2,\sigma_2) & \dots & \phi_2(\mathbf{r}_N,\sigma_N) \\ \dots & \dots & \dots & \dots \\ \phi_N(\mathbf{r}_1,\sigma_1) & \phi_N(\mathbf{r}_2,\sigma_2) & \dots & \phi_N(\mathbf{r}_N,\sigma_N) \end{vmatrix}, \tag{103}$$

where $\phi_i(\mathbf{r}_i,\sigma_i)$ denote spin-orbitals (represented as products of spatial orbitals and spin wavefunctions), σ_i – spins. Swapping particles (*i.e.* swapping the columns of the determinant) automatically changes the sign of the wavefunction in agreement with the Pauli principle. Single determinant is the simplest form for an antisymmetric wavefunction, and the HF method generates the best (*i.e.*, the lowest-energy) single-determinant wavefunction.

The probability of simultaneously finding like-spin electrons near each other is dramatically reduced relative to Equation 100 (see Thijssen, 1999). For the electrons with opposite spins, the HF method still (incorrectly) gives the uncorrelated probabilities (Eqn. 100). Effectively, around each electron there is an exchange hole, where electrons of the same spin have a reduced probability to enter. In reality, there also is a correlation hole, but correlation effects are neglected in the HF approach.

For atoms and molecules, HF calculations are a successful routine. Most HF studies of crystals have been performed with the CRYSTAL (Dovesi *et al.*, 1996) code implementing non-relativistic HF theory for periodic structures (crystals, polymers, surfaces). A number of impressive studies have been performed for minerals – *e.g.*, $MgSiO_3$ (ilmenite: Nada *et al.*, 1992, perovskite: D'Arco *et al.*, 1993b, 1994), sellaite MgF_2 (Catti *et al.*, 1991), SiO_2 polymorphs (Sherman, 1992), Si-mordenite $Si_{48}O_{96}$

(White & Hess, 1993), brucite $Mg(OH)_2$ (D'Arco *et al.*, 1993a). A very interesting review is given in Dovesi *et al.* (1992).

HF method predicts rather accurately the total energies and electron densities. However, binding energies in molecules and solids are always severely underestimated, typically by 20–40% (Dovesi *et al.*, 1992; Dovesi, 1996). This is due to the correlation error, which is not completely cancelled when calculating energy differences for processes involving breaking of electron pairs. Applying an *a posteriori* correction (using a correlation density functional to the HF electron density) allows one to recover typically ~ 80% of the correlation energy (Dovesi, 1996). This correction appears to give good predictions of chemical reaction energies (Catti *et al.*, 1994); atomisation energies are still ~ 5–10% underestimated.

HF eigenvalues give a reasonable approximation to the experimental ionisation potentials and electron affinities of atoms and ions, but turn out to be very poor in modelling band gaps and excitation spectra in solids (Pisani, 1996). Neglect of correlation deteriorates the predicted structures (bonds are usually ~ 1–3% too long; for the compounds of the 5th period elements they are ~ 10% too long! – Aprà *et al.*, 1991) and leads to a ~ 10% overestimation of the vibrational frequencies. Van der Waals forces, coming purely from correlation, are totally neglected in the HF approximation. Electron correlation is fundamental for the metallic state, for which HF method fails completely, as demonstrated for the simplest metal – homogeneous electron gas (Ashcroft & Mermin, 1976; Pisani, 1996).

By definition, the difference between the HF energy and the exact non-relativistic energy is called the correlation energy. There are two main types of correlation effects: (i) 'dynamical' correlation – in their instantaneous positions, electrons correlate their movement so as to avoid each other, irrespective of their spins. Van der Waals forces (coming from 'instantaneous dipoles' on atoms) are one of the manifestations of this effect, (ii) non-dynamical (or 'left-right') correlation – effect responsible for localisation of the exchange-correlation hole in molecules and solids (see below).

There are several possible ways to include correlation. Perturbation theory and configuration interaction methods are very popular, but both have serious problems (especially in application to condensed matter). Quantum Monte Carlo methods (see Foulkes *et al.*, 2001) are extremely promising. Currently, the most popular methods are based on density functional theory (DFT).

Introduction to density functional theory (DFT)

DFT is a revolutionary theory in quantum mechanics: it formulates quantum mechanics in terms of the observable and tractable electron density instead of the extremely complicated and unobservable wavefunction. Pre-DFT works include important developments in 1927 by Thomas and Fermi, and the Xα-method proposed by Slater in 1951. The rigorous foundation of DFT was laid in the work of Hohenberg & Kohn (1964) and their followers, who formulated and proved two remarkable theorems:

Theorem 1. The electron density $\rho(\mathbf{r})$ determines (within an additive constant) the external potential, *i.e.* the potential due to the nuclei. As a consequence, $\rho(\mathbf{r})$ determines the total energy and the wavefunction.

Theorem 2. The ground-state total energy of a system is variational with respect to the electron density, *i.e.* the exact $\rho(\mathbf{r})$ provides the minimum possible energy for the ground state.

The first theorem states that there exists a universal total energy functional $E[\rho]$, and that all ground-state properties are completely determined by the electron density. The first theorem applies to the ground and excited states alike; the variational principle of the second theorem is applicable only to the ground state and to the lowest-energy excited states of a given symmetry (see Jones & Gunnarsson, 1989).

The total electronic energy can be written as:

$$E[\rho] = T[\rho] + E_{e-n}[\rho] + E_{e-e}[\rho] = T[\rho] + \int d\mathbf{r}\rho(\mathbf{r})v_{e-n}(\mathbf{r}) + E_{e-e}[\rho], \quad (104)$$

where $T[\rho]$ is the kinetic energy functional and $v_{e-n}(\mathbf{r})$ is the potential due to the nuclei. The problem is that the non-local exact functional $E[\rho]$ is unknown and likely to be exceedingly complicated.

The Thomas–Fermi model is the lowest-order approximation. The kinetic energy functional in this model is based on the homogeneous electron gas expression:

$$T^{TF}[\rho] = \frac{3}{10}(3\pi^2)^{2/3}\int dr\rho^{5/3} \qquad (105)$$

and the electron–electron interactions are approximated by the Hartree expression (Eqn. 96). The Thomas–Fermi–Dirac model also includes approximate exchange energy:

$$E_x[\rho] = \int dr\rho(\mathbf{r})e_x(\rho(\mathbf{r})) = -\frac{3}{4\pi}(3\pi^2)^{1/3}\int dr\rho^{4/3}, \qquad (106)$$

where $e_x(\rho(\mathbf{r}))$ is the exchange energy of a homogeneous electron gas per 1 electron.

The errors in the kinetic energy associated with Equation 105 are too large: by the Teller theorem, no binding is permitted for molecules and solids in the Thomas–Fermi and Thomas–Fermi–Dirac models. Introduction of successively higher-order gradients of the density is not a way to eventually obtain an accurate method: the sixth-order gradient expansion of the kinetic energy diverges for atoms, and no further expansion can be constructed (Parr & Yang, 1989). In a way, Hohenberg-Kohn theorems provide an exactification of Thomas–Fermi theory (Kohn, 1999a), but only in principle, because known approximate kinetic energy functionals fail to give reasonable results.

Kohn–Sham method

Kohn & Sham (1965) devised a method, which enables accurate practical calculations. As Kohn (1999a) described it, this method is a formal exactification of the Hartree method. It starts with a fictitious system of non-interacting electrons having the same electron density as the real system of interacting electrons. We introduce electronic orbitals[15] and calculate the kinetic energy T_s of this fictitious non-interacting system, as

---

[15] Electron orbitals in the Kohn–Sham theory are only auxiliary constructions with no real physical meaning other than a tool to calculate the kinetic energy with a reasonable accuracy.

in Equation 98. T_s is a much better (though still not exact) estimate of the kinetic energy than Equation 105, as suggested by the comparison of the Hartree and Thomas–Fermi approximations (Kohn, 1999a). We also know that it gives a lower bound to the true kinetic energy (Parr & Yang, 1989):

$$T_s[\rho] \leq T[\rho]. \tag{107}$$

The Kohn–Sham method overcomes the difficulty in constructing a density functional for the kinetic energy; the price to be paid is the introduction of orbitals[16].

The total electronic energy is:

$$E[\rho] = T_s[\rho] + \int d\mathbf{r} \ \rho(\mathbf{r}) v_{e-n}(\mathbf{r}) + \frac{1}{2} \int d\mathbf{r} \ \rho(\mathbf{r}) v_H(\mathbf{r}) + E_{xc}[\rho], \tag{108}$$

i.e. it is a sum of the non-interacting kinetic energy, energy of electrons due to the interaction with the nuclei, Hartree energy and exchange-correlation energy. All the contributions in Equation 108 are known exactly, apart from the exchange-correlation energy functional $E_{xc}[\rho]$, which has to be somehow approximated. By definition (Eqn. 108), the difference between T_s and the exact kinetic energy T is included in the exchange-correlation energy. Because of this, the virial theorem (Eqn. 98) does not hold in the Kohn–Sham DFT (Parr & Yang, 1989).

The Kohn–Sham method reduces the formidable many-body problem (Eqn. 89) to a set of coupled one-electron equations mathematically similar to the Hartree–Fock equations:

$$\left\{ -\frac{1}{2} \nabla^2 + v_{e-n}(\mathbf{r}) + v_H\big[\rho(\mathbf{r})\big] + v_{xc}\big[\rho(\mathbf{r})\big] \right\} \phi_i(\mathbf{r}) = \varepsilon_i \phi(\mathbf{r}), \tag{109}$$

where the exchange-correlation potential $v_{xc}(\mathbf{r})$ is defined as:

$$v_{xc}(\mathbf{r}) = \frac{\delta E_{xc}[\rho(\mathbf{r})]}{\delta \rho(\mathbf{r})}. \tag{110}$$

Kohn–Sham orbitals

Although the Kohn–Sham orbitals were introduced only as a means to get the kinetic energy, they are often used in chemical studies; Baerends & Gritsenko (1997) and Stowasser & Hoffmann (1999) advocated their use. In all practical calculations an Aufbau principle is assumed to hold, where only the lowest-energy orbitals are occupied (Parr & Yang, 1989).

There are a few exact results; Janak theorem (Parr & Yang, 1989) states that the eigenvalues are the derivatives of the total energy with respect to the orbital occupancies:

$$\left(\frac{\partial E}{\partial n_i} \right) = \varepsilon_i. \tag{111}$$

---

[16] Note that in Thomas–Fermi and related models there is no wavefunction and no orbitals at all.

This contrasts with the Koopmans' theorem in HF theory. Only for the highest occupied orbital (see Jones & Gunnarsson, 1989) the eigenvalue is equal to minus the ionisation potential. The latter result, proved in the exact DFT, does not hold for DFT with approximate functionals.

For band gaps E_g in solids, an interesting result exists (Jones & Gunnarsson, 1989):

$$E_g = \Delta\varepsilon + \Delta, \tag{112}$$

where $\Delta\varepsilon$ is the eigenvalue gap, and Δ is a correction arising from exchange-correlation effects. This poorly known correction is present even in the case of the exact exchange-correlation functional. LDA and GGA approximate functionals give eigenvalue gaps that are ~ 30–50% smaller than experimental band gaps.

Exchange-correlation hole

The exchange-correlation energy is defined as:

$$E_{xc}[\rho] = \frac{1}{2}\int d\mathbf{r}\int \frac{\rho(\mathbf{r})\rho_{xc}(\mathbf{r},\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,d\mathbf{r}', \tag{113}$$

i.e. it is just the classical electrostatic energy of interaction between an electron and its exchange-correlation hole $\rho_{xc}(\mathbf{r}, \mathbf{r}')$. If the reference electron is sitting at the point $\mathbf{r}$, the exchange-correlation hole $\rho_{xc}(\mathbf{r}, \mathbf{r}')$ shows how much the probability of finding another electron at the point $\mathbf{r}'$ is decreased (compared to the time-averaged electron density $\rho(\mathbf{r}')$). A number of simple rules exist (see Perdew & Kurth, 1998).

The exchange-correlation hole can never exceed (by the absolute value) the charge density available at the point $\mathbf{r}'$:

$$-\rho_{xc}(\mathbf{r}, \mathbf{r}') \le \rho(\mathbf{r}'). \tag{114}$$

$\rho_{xc}(\mathbf{r}, \mathbf{r}')$ can be split into the exchange hole $\rho_x(\mathbf{r}, \mathbf{r}')$ and correlation hole $\rho_c(\mathbf{r}, \mathbf{r}')$:

$$\rho_{xc}(\mathbf{r}, \mathbf{r}') = \rho_x(\mathbf{r}, \mathbf{r}') + \rho_c(\mathbf{r}, \mathbf{r}'), \tag{115}$$

and the following important properties can be formulated:

$$\int d\mathbf{r}'\rho_{xc}(\mathbf{r},\mathbf{r}') = \int d\mathbf{r}'\rho_x(\mathbf{r},\mathbf{r}') = -1 \quad \text{(exchange hole sum rule)}, \tag{116}$$

$$\rho_x(\mathbf{r},\mathbf{r}') \le 0 \quad \text{(non-positiveness of the exchange hole)}, \tag{117}$$

$$\int d\mathbf{r}'\rho_{xc}(\mathbf{r},\mathbf{r}') = 0 \quad \text{(correlation hole sum rule)}. \tag{118}$$

While the exchange hole is always non-positive, the correlation hole is positive in some places and negative in others, integrating in total to zero. The effect of correlation is to make the exchange-correlation hole deeper and shorter-ranged.

In molecules and solids, both exchange and correlation holes can be delocalised over many centres, but the total exchange-correlation hole is usually well localised (see Fig. 12). The important property of 'short-sightedness' of electrons is one of the main justifications of the local density approximation (LDA). The long-range part of correlation, which cancels the long-range part of the exchange, is called the 'right-left'
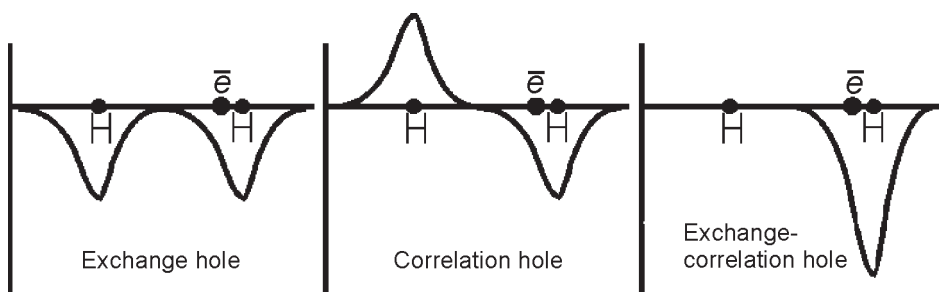
Fig. 12. Exchange, correlation, and total exchange-correlation holes in H_2 molecule. The nuclei (distance ~ 2.5 Å) are shown by circles. The reference electron is located 0.3 bohr to the left of the right nucleus. Re-drawn schematically after Baerends & Gritsenko (1997).

correlation. As is clear from the present discussion and as was first elucidated by Slater in 1950s, LDA and other local approximations (*e.g.*, GGA) automatically include 'right-left' correlation by assuming a localised exchange-correlation hole.

Formal properties of the exchange-correlation functional

Now let us consider the exchange-correlation functional. Even though the exact functional is unknown, a number of its exact properties are known (see Perdew & Kurth, 1998). These properties are essential in the derivation of good approximate functionals.

First of all, the exchange energy is negative and correlation energy non-positive:

$$E_x[\rho] < 0, \tag{119}$$

$$E_c[\rho] \leq 0. \tag{120}$$

Second, exact exchange completely cancels the self-interaction error. For one-electron systems (*e.g.*, H, He$^+$ *etc.*), exchange potential cancels the Hartree potential and the correlation energy is zero.

Another important piece of exact information is the scaling relations (see Perdew & Kurth, 1998), which determine how exchange and correlation functionals change under uniform coordinate scaling transformations.

A number of inequalities play an important role in DFT, of which the Lieb–Oxford bound is particularly important:

$$\left|E_{xc}[\rho]\right| \leq 2.273 \left|E_{x,LDA}[\rho]\right| = 2.273 \frac{3}{4\pi}(3\pi^2)^{1/3} \int d\mathbf{r} \rho^{4/3}. \tag{121}$$

We also know the exact spin scaling relations for the exchange:

$$E_x[\rho_\uparrow, \rho_\downarrow] = \frac{1}{2}E_x[2\rho_\uparrow, 0] + \frac{1}{2}E_x[0, 2\rho_\downarrow] \tag{122}$$

and non-interacting kinetic energy:

$$T_{\mathrm{s}}[\rho_\uparrow, \rho_\downarrow] = \frac{1}{2} T_{\mathrm{s}}[2\rho_\uparrow, 0] + \frac{1}{2} T_{\mathrm{s}}[0, 2\rho_\downarrow]. \tag{123}$$

These relations play the central role in DFT for spin-polarised systems. Unfortunately, there is no simple spin scaling relation for the correlation energy.

Other notations are useful, *e.g.* the spin density:

$$\rho_{\mathrm{spin}}(\mathbf{r}) = \rho_\uparrow(\mathbf{r}) - \rho_\downarrow(\mathbf{r}) \tag{124}$$

and dimensionless spin polarisation, ranging between 0 and 1:

$$\xi(\mathbf{r}) = \frac{|\rho_\uparrow(\mathbf{r}) - \rho_\downarrow(\mathbf{r})|}{\rho_\uparrow(\mathbf{r}) + \rho_\downarrow(\mathbf{r})}. \tag{125}$$

DFT has been extended to treat non-collinear magnetism (see Hobbs *et al.*, 2000 for a practical implementation and Sandratskii, 1998 for a theoretical review).

As it was noted before, in exact DFT the highest-occupied eigenvalue is equal to minus the ionisation potential of the system; it must change discontinuously at an integer number Z of electrons:

$$\varepsilon_{\mathrm{HO}} = \frac{\partial E}{\partial n_{\mathrm{HO}}} = \begin{array}{ll} -I & (Z-1 < N < Z) \\ -A & (Z-1 < N < Z+1), \end{array} \tag{126}$$

where I and A are the ionisation potential and electron affinity of the Z-electron system (*e.g.*, an atom). The exact exchange-correlation potential too is discontinuous at integer numbers of electrons. In approximate DFT the derivative discontinuity (Eqn. 126) cannot be reproduced; instead, we approximately get

$$\varepsilon_{\mathrm{HO}} \approx -\frac{I+A}{2}, \tag{127}$$

which confers the meaning of minus the Mulliken electronegativity on the highest-occupied eigenvalue $\varepsilon_{\mathrm{HO}}$.

DFT: Approximate functionals

Local density approximation (LDA)
The LDA is the simplest approximation to $E_{\mathrm{xc}}[\rho]$; although the local expression (Eqn. 105) for the kinetic energy is very poor for atoms, local exchange-correlation is rather accurate. In the LDA, the kinetic energy is calculated from the wavefunction rather than from the electron-gas formula (Eqn. 105).

The exchange-correlation energy is calculated as:

$$E_{\mathrm{xc}}[\rho] = \int d\mathbf{r}\,\rho(\mathbf{r}) e_{\mathrm{xc}}(\rho(\mathbf{r})), \tag{128}$$

where the exchange energy density of a homogeneous electron gas is known exactly from Equation 106:

$$e_{xc}(\rho) = -\frac{3}{4\pi}(3\pi^2)^{1/3}\rho^{1/3} = -\frac{3}{4\pi}\left(\frac{9\pi}{4}\right)^{1/3}r_s^{-1}, \qquad (129)$$

and the correlation energy is known very accurately from quantum Monte Carlo simulations (Ceperley & Alder, 1980; Ortiz & Ballone, 1994). An analytical representation satisfying the exactly known high- and low-density analytical limits and fitting the numerical quantum Monte Carlo results for the spin-unpolarised case, is (Perdew & Zunger, 1981):

$$\begin{aligned}
e_c(\rho) &= 0.0311\ln r_s - 0.048 + 0.0020 r_s \ln r_s - 0.0116 r_s \quad (r_s < 1) \\
e_c(\rho) &= -0.1423/(1 + 1.0529\sqrt{r_s} + 0.3334 r_s) \qquad\qquad (r_s > 1).
\end{aligned} \qquad (130)$$

Figure 13 shows different contributions to the total energy of the homogeneous electron gas and clearly indicates the importance of the correlation contribution.

Perdew & Wang (1992) expressed the correlation energy of the homogeneous electron gas more accurately by a single analytical formula, which also satisfies the high- and low-density limits and quantum Monte Carlo data:

$$e_c(\rho) = -2c_0(1 + \alpha_1 r_s)\ln\left[1 + \frac{1}{2c_0(\beta_1 r_s^{1/2} + \beta_2 r_s + \beta_3 r_s^{3/2} + \beta_4 r_s^2)}\right], \qquad (131)$$

where $c_0 = 0.031091$ and $c_1 = 0.046644$, $\beta_1 = \frac{1}{2c_0}\exp\left(-\frac{c_1}{2c_0}\right) = 7.5957$, $\alpha_1 = 0.21370$,
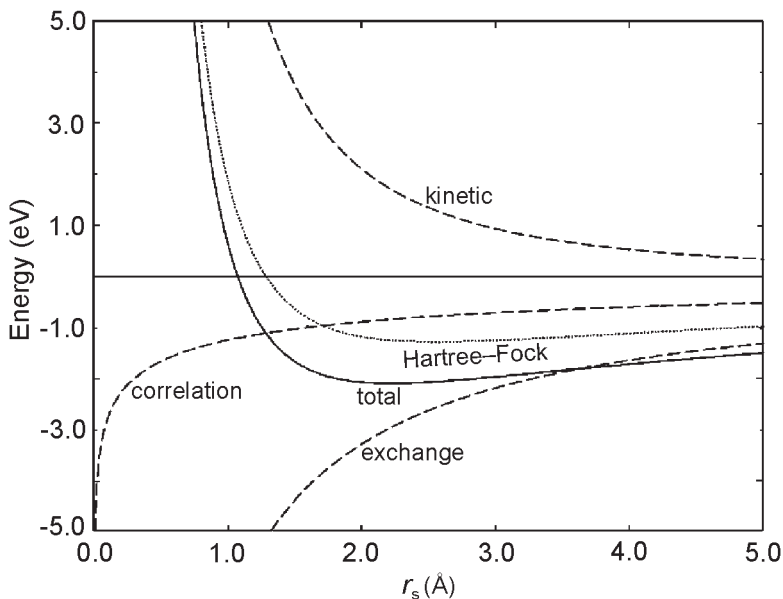
Fig. 13. Energy contributions for the homogeneous electron gas (per 1 electron). Average r_s values of the valence electrons of several metals are shown (redrawn with modifications from Pisani, 1996). Hartree–Fock curve includes all the contributions except correlation. It is clear that correlation is non-negligible.
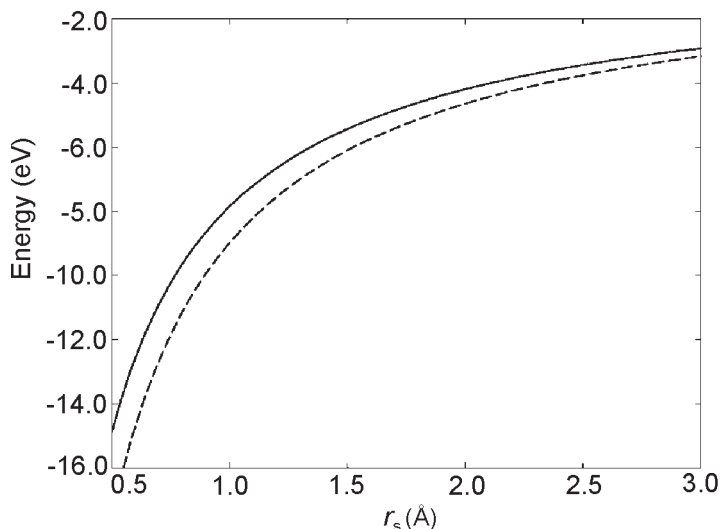
Fig. 14. Exchange-correlation energy of the spin-unpolarised (solid line) and fully spin-polarised (dashed line) electron gas.

$\beta_2 = 2c_0\beta_1^2 = 3.5876$, $\beta_3 = 1.6382$, and $\beta_4 = 0.49294$ for the spin-unpolarised ($\xi = 0$) electron gas. For a fully spin-polarised case ($\xi = 1$) $c_0 = 0.015545$ and $c_1 = 0.025599$, $\alpha_1 = 0.20548$, $\beta_3 = 3.3662$, and $\beta_4 = 0.62517$. Figure 14 shows that spin polarisation strongly affects exchange-correlation.

The LDA generalised for spin-polarised systems is called LSDA (local spin density approximation). For the exchange energy we have a simple exact spin scaling relation (Eqn. 122), which implies:

$$e_x(\rho_\uparrow, \rho_\downarrow) = e_x(\rho)\frac{(1+\xi)^{4/3} + (1-\xi)^{4/3}}{2} . \tag{132}$$

The exact spin scaling for correlation energy is unknown, but for the electron gas we have a very accurate formula proposed by Vosko *et al.* (1980).

LDA has large errors in exchange (about –5%) and correlation (about +100%), but these largely cancel each other, explaining the successes of the LDA. The LDA does not give accurate exchange-correlation holes for atoms, but gives reasonable spherical averages for these holes (luckily, only these are energetically important). Most importantly, LDA exactly satisfies all of the sum rules in Equations 116 and 118. The self-interaction error is largely cancelled by the exchange-correlation potential; *e.g.*, in the H atom ∼ 95% of this error is cancelled at the LSDA level of theory (Thijssen, 1999). This error becomes important for highly localised electronic states (such as *d*-states in transition metal atoms). Perdew & Zunger (1981) devised a simple method to incorporate the self-interaction correction in the LDA approach (LDA + SIC), but the method is difficult to apply to crystals.

The shortcomings of the LDA include the following: (i) underestimation of the total energies – LDA treats well the valence electrons, but underbinds the core electrons, (ii) overestimation of the atomisation energies (by ~ 20%) of molecules and solids, (iii) bond lengths are usually ~ 1% too short, (iv) reaction barriers are usually much too low compared to experiment, (v) large errors for weakly bonded systems (*e.g.*, hydrogen bonds), (vi) van der Waals bonding cannot be treated, (vii) sometimes the energy differences between polymorphs are inaccurate (*e.g.*, for quartz–stishovite the huge energy difference of ~ 0.5 eV is not reproduced: even the sign is wrong!), (viii) often a wrong ground state is predicted (*e.g.*, paramagnetic fcc instead of ferromagnetic bcc for Fe), (ix) serious errors for transition metal oxides, (x) in many cases, the dissociation products are not neutral atoms, but atoms with non-physical fractional charges – see a very interesting paper by Becke (2000).

These serious shortcomings inspire further developments of functionals. The impressive successes of the LDA suggest it as a reference point in these developments. Some of these developments will be discussed below Tables 4–6 illustrale the performance of different approximate functionals.

Generalised gradient approximation (GGA)

An obvious way to construct an improved functional is to expand the exchange-correlation energy in powers of density gradients. Including the first density gradient in this way produces an approximation called GEA (Gradient Expansion Approximation), which is exact for an electron gas with a slowly varying density, but turns out to be worse than the LDA for atoms, because GEA does not satisfy the sum rules in Equations 116 and 118. The short-range part of the exchange-correlation hole is improved over the LDA, but the long-range part is worsened and has spurious undamped oscillations, due to which GEA does not satisfy the sum rules and exchange hole sometimes becomes positive.

Unlike GEA, GGA is constructed in such a way as to preserve the correct features of the LDA and add some more. There are several popular GGA functionals, the best of which seem to be the PW91 functional (Wang & Perdew, 1991) and a very similar functional PBE (Perdew *et al.*, 1996). Wang & Perdew (1991) have constructed their functional by starting with the GEA exchange-correlation hole and setting it to zero everywhere the GEA exchange hole is positive and everywhere beyond a certain cut-off radius, chosen so as to enforce the sum rules.

The GGA exchange-correlation functional can be defined as follows:

$$E_{xc}[\rho, \nabla\rho] = \int d\mathbf{r} F_{xc}(\rho, s)\rho(\mathbf{r})e_x(\rho(\mathbf{r})),\qquad(133)$$

where $e_x(\rho(\mathbf{r}))$ is the exchange energy of the spin-unpolarised homogeneous electron gas per 1 electron, and the enhancement factor $F_{xc}(\rho, s)$ depends on both the density and its reduced gradient s:

$$s = \frac{|\nabla\rho|}{2(3\pi)^{1/3}\rho^{4/3}(\mathbf{r})}.\qquad(134)$$

Table 4. Total energies of atoms (in atomic units): comparison of experiment with several approximate methods.

Atom	HF	LSDA	LSDA + SIC	GGA (PW91)	Experimental
H	−0.500	−0.479	−0.500	−0.500	−0.500
He	−2.86	−2.835	−2.918	−2.900	−2.904
Ne	−128.55	−128.228	−129.268	−128.947	−128.937
Ar	−526.82	−525.938	−528.289	−527.539	−527.60

Data were taken from Perdew & Zunger (1981), Parr & Yang (1989), Lee & Martin (1997).

Table 5. Atomisation energies (in eV) of several molecules: experiment versus theory.

Molecule	HF	LSDA	GGA (PBE)	GGA (PW91)	Meta-GGA	Experiment
H_2	3.64	4.90	4.55	4.55	4.97	4.73
H_2O	6.72	11.58	10.15	10.19	9.98	10.06
HF	4.21	7.03	6.16	6.20	6.01	6.11
Li_2	0.09	1.00	0.82	0.87	0.98	1.04
O_2	1.43	7.59	6.24	6.20	5.70	5.25
F_2	−1.60	3.34	2.30	2.34	1.87	1.69
CH_4	14.22	20.03	18.21	18.26	18.26	18.17

Experimental values include a zero-point energy correction. Meta-GGA calculations used GGA electron densities. (Taken from Perdew *et al.*, 1996; meta-GGA results: from Perdew *et al.*, 1999).

Table 6. Comparison of HF, LDA and GGA for solids.

Property	HF (HF + c)	LSDA	PW91 (PBE)	Experiment
Periclase MgO [GGA = PW86]				
a_0 [Å]	4.191	4.160	4.244	4.20
E_a [eV]	−7.32 (−9.69)	–	–	−10.28
K_0 [GPa]	186	198	157	167
Ferromagnetic bcc-Fe				
V_0 [Å³]	–	10.44	(11.34)	11.77
K [GPa]	–	260	(200)	172
K'_0	–	4.6	(4.5)	5.0
Diamond C				
a_0 [Å]	3.58	3.53	3.57 (3.57)	3.567
E_a [eV]	−5.2 (−7.4)	−8.87	−7.72 (−7.72)	−7.55
K_0 [GPa]	471	455	438 (439)	442

Data are compiled from Lee & Martin (1997), Lichanot (2000), Dovesi (1996), Zupan *et al.* (1998). HF + c means Hartree–Fock calculation with an *a posteriori* correlation correction.

The Lieb–Oxford bound (Eqn. 121) can be rewritten in terms of $F_{xc}(\rho, s)$:

$$F_{xc}(\rho, s) \leq 2.273. \tag{135}$$

Enhancement factors of the LDA, GEA and GGA are given in Figure 15.

GGA significantly improves the description of the core (and to some extent the valence) electrons. Total energies are much better than in the LSDA and HF. LSDA overbinding is corrected by the GGA. Energy differences and especially reaction barriers are improved, as well as the description of magnetic systems. Exchange energies of atoms are reproduced with a typical error of only 0.5%; the typical error for the correlaton energy is 5% (Perdew & Kurth, 1998). GGA exchange-correlation is not self-interaction-free, but for the H atom the self-interaction error is almost completely cancelled within the GGA. GGA, like LSDA, cannot reproduce the derivative discontinuity (Eqn. 126) on passing an integer number of electrons.

Some of the failings of the GGA include: (i) overestimation of bond lengths by ~ 1% – the GGA 'overcorrects' the LDA, (ii) as in the LSDA, but less often, electronic ground states of atoms, molecules and solids are sometimes incorrect, (iii) sometimes the energy differences are incorrect. While the LDA tends to overestimate the stability of dense structures, gradient corrections of the GGA stabilise open structures (see Zupan *et al.*, 1998).
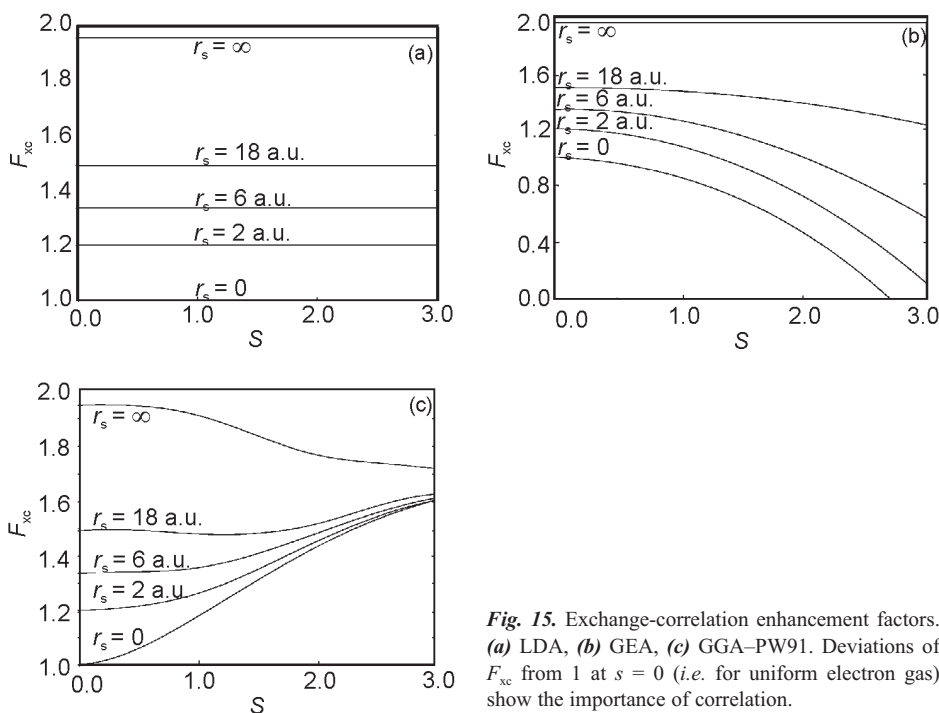
Fig. 15. Exchange-correlation enhancement factors. *(a)* LDA, *(b)* GEA, *(c)* GGA–PW91. Deviations of F_{xc} from 1 at $s = 0$ (*i.e.* for uniform electron gas) show the importance of correlation.

Meta-GGA

The next step after GGA is to include the Laplacian of the electron density $\nabla^2\rho$. This step is very promising, since the Laplacian of the density proved to be very important in the Bader analysis of chemical bonding (Bader, 1990; Tsirelson, 1993; Coppens, 1997). Quantum Monte Carlo simulations (Nekovee *et al.*, 2001) demonstrated that often GGA worsens the LDA exchange-correlation hole and inclusion of the Laplacian of the density is essential in modelling the exact hole. It is also useful to include the kinetic energy density, defined as:

$$\tau(\mathbf{r}) = \frac{1}{2}\sum_{occ}\nabla^2\phi_i(\mathbf{r}). \tag{136}$$

Becke (2000) showed that $\tau(\mathbf{r})$ is a useful indicator of delocalisation of the exchange hole. Conventional functionals fail for systems with delocalised exchange holes (*e.g.*, stretched H_2^+). Including $\tau(\mathbf{r})$, it is possible to accurately model the exchange energy of such systems. Becke (2000) has constructed a meta-GGA exchange functional, which accurately simulated the exact exchange.

Perdew *et al.* (1999) have constructed a meta-GGA exchange-correlation functional including both $\nabla^2\rho$ and $\tau(\mathbf{r})$. This functional is not fully non-empirical, but has two fitted parameters. Among its good properties are self-interaction-free correlation and reduction to LDA for the uniform electron gas. This meta-GGA outperforms the GGA for atomisation energies, but bond lengths (and lattice parameters) are not significantly improved. There are other meta-GGA functionals and much ongoing work. See Perdew & Burke (1996) and Kurth *et al.* (1999) for a very interesting discussion of different functionals.

Hybrid functionals

Combining local (LDA or GGA) correlation and exact (HF) exchange turns out to give worse results than pure GGA exchange-correlation. Nevertheless, as emphasised in Becke's seminal paper (Becke, 1993), exact exchange is important in constructing accurate functionals. Becke proposed a hybrid functional, whose exchange part is a mixture of local and exact exchange and whose correlation part is a local functional, schematically:

$$E_{xc} = E_{xc}^{local} + a_{mix}(E_x^{exact} - E_x^{local}), \tag{137}$$

where a_{mix} is a fitted parameter (~ 0.20). Hybrid functionals generally are highly accurate in predicting structures and atomisation energies, as well as total energies. There are a host of hybrid functionals, the most popular of which is called B3LYP.

Van der Waals bonding. One of the most difficult problems for quantum-mechanical simulations is to reproduce van der Waals bonding. Originating from dynamical correlation, van der Waals forces are by definition absent in the HF theory. Even at the DFT level there are problems: these long-range forces cannot be adequately reproduced by any of the local functionals. LDA is too crude for correlation generally; GGA (PBE functional) can give some dispersion forces and even give reasonable bond lengths for noble-gas dimers. However, due to its local nature, GGA gives an exponential, rather

than R^{-6} dependence for the van der Waals energy. Kohn *et al.* (1998) have proposed a practical recipe to get accurate dispersion forces and their long-range behaviour within the DFT framework.

Technical details of *ab initio* simulations

Brillouin zone sampling

An infinite crystal contains an infinite number of electrons. Exploiting translational symmetry of crystals and introducing the reciprocal lattice, it is possible to consider only the number of electrons contained within one unit cell. However, the HF or Kohn–Sham equations must be solved for each **k**-point of the reciprocal space. The wavevector **k** becomes a new quantum number.

By Bloch theorem, each crystal orbital with the wavevector **k** is a product of a periodic function $w_{\mathbf{k}}(\mathbf{r})$ and a plane wave:

$$\phi_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\mathbf{r}} w_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\mathbf{r}} \sum_{\mathbf{K}} C_{\mathbf{k}+\mathbf{K}} e^{i\mathbf{K}\mathbf{r}}, \tag{138}$$

where the periodic function $w_{\mathbf{k}}(\mathbf{r})$ is usually expressed as a Fourier series with coefficients $C_{\mathbf{k}+\mathbf{K}}$ (**K** is a reciprocal lattice vector).

The orbitals and orbital energies generally depend on the wavevector **k**, much like phonon frequencies depend on it. One has to solve the one-electron equations at each **k**-point and integrate over the Brillouin zone; in practice only a finite set of **k**-points is used and convergence must always be checked. For metals one needs many more **k**-points than for insulators or semiconductors. Smaller unit cells also require more **k**-points to be included. Monkhorst & Pack (1976) defined a uniform set of special sampling **k**-points as:

$$\mathbf{k} = u_{\mathrm{p}}\mathbf{b_1} + u_{\mathrm{r}}\mathbf{b_2} + u_{\mathrm{s}}\mathbf{b_3}, \tag{139}$$

where $\mathbf{b_1}$, $\mathbf{b_2}$, and $\mathbf{b_3}$ are the reciprocal lattice basis vectors and u_{p}, u_{r}, u_{s} are numbers from the sequence:

$$u_{\mathrm{r}} = \frac{2r - q - 1}{2q} \qquad (r = 1, 2, 3, \ldots, q). \tag{140}$$

The total number of **k**-points is q^3, but due to symmetry the number of independent **k**-points can be much smaller.

Basis sets

There are many different basis set schemes, of which we discuss only the traditional atomic orbital and plane wave basis sets. In the LCAO (Linear Combination of Atomic Orbitals) scheme one defines atom-centred orbitals as a product of the angular (giving the shape to *s*-, *p*-, *d*-, and *f*-orbitals) and radial ($\chi(r)$) parts, where the radial part is a linear combination of either Slater-type functions ($r^m e^{-\xi r}$) or Gaussian-type functions ($e^{-\alpha r^2}$), where m, ξ, and α are parameters. Slater functions better approximate the atomic orbitals, but are much more computationally expensive, so Gaussian functions are usually preferred.

In practice, LCAO basis sets are incomplete. In molecules and crystals, parameters of the LCAO wavefunction (*e.g.*, orbital exponents) depend on the atomic positions, but are only optimised for one structure. Dependence of the basis set on atomic positions implies the presence of Pulay forces, which must be evaluated when optimising crystal structures. Very economical for insulators, LCAO basis requires very many Gaussian functions for studies of metals. There is another problem with the LCAO basis, called the basis set superposition error. It is also difficult to systematically increase the basis set expressed in local orbitals.

Plane wave basis set is the most natural and general basis set for crystals, following directly from the Bloch theorem. A single plane wave is:

$$\phi_{\mathbf{k}+\mathbf{K}}^{\mathrm{PW}}(\mathbf{r}) = e^{i(\mathbf{k}+\mathbf{K})\mathbf{r}}. \tag{141}$$

This basis set is complete and very convenient for many applications, but a huge number of plane waves are needed to describe rapidly changing wavefunctions in solids. *E.g.*, for the valence electrons in Al, an estimated 10^6 plane waves are needed to reproduce oscillations of the valence electron wavefunction in the core region. For the core electrons this problem is extremely serious – the core electron wavefunctions are more rapidly changing and there is a cusp of the density (Eqn. 91) at the nucleus. There are several ways to overcome this problem – *e.g.*, the LAPW method, PAW (projector augmented wave) method, linear muffin-tin method, and the pseudopotential method, which we shall describe below. For details of these methods see Singh (1994), Blöchl (1994), Thijssen (1999).

By construction, plane wave basis set can be used only in conjunction with periodic boundary conditions. Atoms, molecules, and surfaces can be treated in an approximate fashion by using sufficiently large unit cells preventing significant interaction between their periodic images.

Pseudopotentials

Payne *et al.* (1992) have given a very clear and comprehensive practical review of the plane wave pseudopotential method. The main ideas of the pseudopotential approach are: (i) to exclude chemically inactive core electrons from explicit consideration and (ii) to replace (within the core) the true Coulombic potential due to the core by a smoother effective potential acting on the valence electrons. This approach is based on the frozen core approximation, *i.e.* core orbitals are assumed to be the same in a free atom and in any chemical environment. To improve the accuracy, one can use 'small core' pseudopotentials with some of the core orbitals treated as valence orbitals.

Outside the 'core' radius r_c the potential and the wavefunction are correct by construction. Within the core region the wavefunction differs from the exact one: it has no radial nodes and is smoother (Figs. 16 and 17), but gives the correct number of electrons. Nodes and oscillations of the exact valence functions in the core region are required by orthogonality with the core orbitals[17]. Pseudopotentials are constructed so as

---

[17] If orbitals of a given angular momentum appear only in the valence shell, they have no radial nodes, no oscillations, better penetrate the core and experience very strong potentials. As a consequence, for elements where this occurs (1st row elements, $3d$ and $4f$ elements) conventional pseudopotentials are inefficient. A way out was found in the formulation of ultrasoft pseudopotentials (Vanderbilt, 1990), which allow one to use relatively large r_c and reduce the number of plane waves by a factor of ~ 2 without any loss in accuracy.
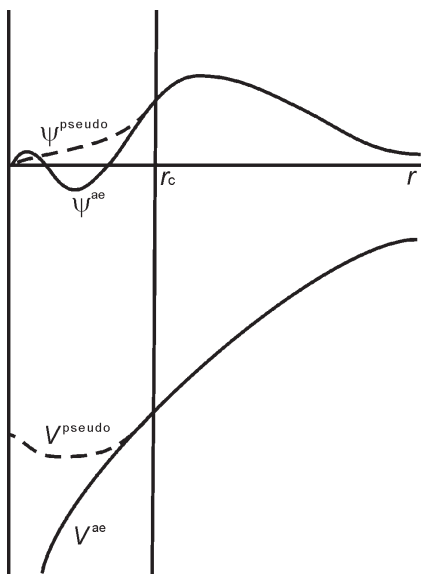
Fig. 16. Construction of a pseudopotential. Beyond r_c the wavefunction and potential match the true all-electron ones (with modifications after Payne *et al.*, 1992).

to match the all-electron eigenvalues in as many different atomic configurations as possible, ensuring transferability to chemically different systems.

The construction of pseudopotentials is inherently non-unique: it depends on r_c and other technical details. Larger r_c result in smoother potentials, which require a smaller number of plane waves, but degrade the accuracy.

Milman *et al.* (2000) analysed the performance of plane wave pseudopotential calculations for compounds of almost all elements. Comparison between pseudopotential and all-electron methods was made in many works, *e.g.* Holzwarth *et al.* (1997). The general conclusion is that pseudopotential calculations are very accurate, except in cases where core polarisation effects are significant (*e.g.*, Ca atom in CaF_2). Another source of errors of pseudopotential calculations is significant overlap of the valence and core orbitals for some atoms (*e.g.*, Na). In such cases, non-linear core corrections (Louie *et al.*, 1982) significantly improve pseudopotentials.

In conjunction with pseudopotentials, plane wave basis sets become extremely useful. Using ultrasoft pseudopotentials, one can satisfactorily model solids by using ~ 100 plane waves per atom. The number of plane waves is controlled by the kinetic energy cut-off parameter E_{cut}; only plane waves with the kinetic energy below E_{cut} are included:

$$\frac{1}{2}|\mathbf{k} + \mathbf{K}|^2 \leq E_{cut} \quad \text{(in atomic units).} \tag{142}$$

The number of plane waves is roughly proportional to the volume of the unit cell: $N_{PW} \approx \dfrac{V}{6\pi^2}(2E_{cut})^{3/2}$. A plane wave basis set does not depend on atomic positions; therefore, there are no Pulay forces. However, it does depend on volume for a finite plane wave basis, resulting in the spurious Pulay stress. The origin of the Pulay stress is in the
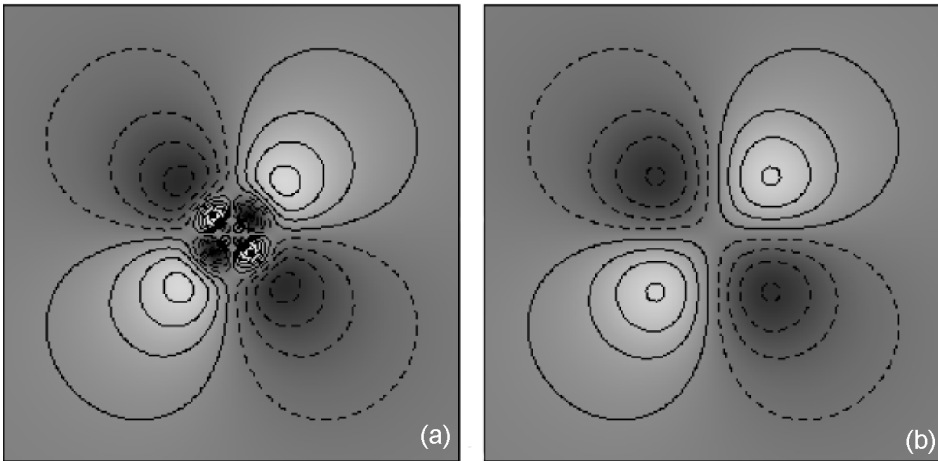
Fig. 17. All-electron *(a)* and pseudowavefunction *(b)* of a 5*d* orbital in Au. Solid contours = positive, dashed contours = negative wavefunction.

basis set incompleteness due to the presence of only a finite number of plane waves. Increasing the basis set, it is possible to reduce the Pulay stress and errors in the total energy to arbitrarily small values.

Figure 18 shows valence electron distributions calculated using this method. Analysis of the charge density is a powerful tool for investigating chemical bonding and interatomic interactions in crystals (Coppens, 1997; Tsirelson, 1986, 1993).

Existing programs for crystals

A large number of *ab initio* codes exist. Some of the most popular codes and their web pages are listed below:

VASP: http://cms.mpi.univie.ac.at/vasp/
CASTEP: http://www.cse.clrc.ac.uk/Activity/UKCP
ABINIT: http://www.abinit.org
PWSCF and **PHONON**: http://www.sissa.it/cm/PWcodes/
CPMD: http://www.cpmd.org
WIEN: http://www.wien2k.at/
SIESTA: http://www.uam.es/departamentos/ciencias/fismateriac/siesta/
CRYSTAL: http://www.cse.dl.ac.uk/Activity/CRYSTAL
LmtART: http://physics.njit.edu/~savrasov/Programs/index_lmtart.htm
TB-LMTO-ASA:
        http://www.mpi-stuttgart.mpg.de/andersen/LMTODOC/LMTODOC.html

Treatment of thermal effects

Temperature is a difficult parameter to handle, because of the difficulty in calculating the partition function (Eqn. 10). Among the approximate methods, lattice dynamics, molecular dynamics, and Monte Carlo methods are the most popular ones.
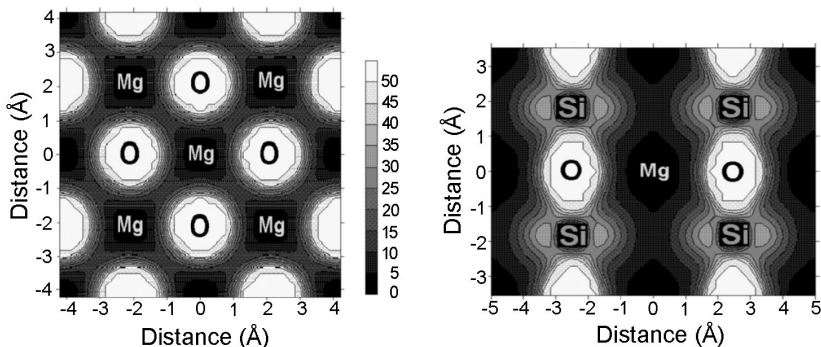
Fig. 18. Theoretical valence electron distributions in minerals. **(a)** MgO ([100] plane), **(b)** hypothetical cubic MgSiO$_3$ perovskite ([110] plane). The density units are 0.01 e/Å^3. Regions with values > 0.5 e/Å^3 are shown in white. MgO: most valence electrons are localised on O atoms, leaving Mg almost completely ionised. MgSiO$_3$: Mg atoms are almost fully ionised, but there is significant electron density on Si atoms, indicating partial covalency of the Si–O bonds. Calculations were performed using the VASP package (Kresse & Furthmüller, 1996). Lev00 code (Kantorovich, 1996–2001) was used for visualisation.

Lattice dynamics takes into account the quantum aspects of lattice vibrations, but is usually restricted to the harmonic or quasiharmonic approximation. Consequently, such simulations are very accurate at low temperatures, but become invalid on approaching the melting curve.

On the other hand, molecular dynamics and Monte Carlo simulations (for further reading see Allen & Tildesley, 1987), which include the full treatment of the anharmonic as well as harmonic effects, but treat the atomic motion classically. Consequently, these simulations are exact only at high temperatures.

Lattice dynamics provides a simple way to calculate the free energy and optimise the structure by minimising the free energy within the quasiharmonic approximation. It has recently become possible to calculate the free energy derivatives analytically (Kantorovich, 1995; for a particular implementation for semiclassical calculations see Gale, 1998), which is extremely useful in predicting the temperature evolution of crystal structures. Lattice dynamics simulations are now possible in conjunction with DFT by means of density functional perturbation theory (Baroni *et al.*, 2001). Figure 19 shows an example of such calculations. Lattice dynamics is discussed in depth by Gramaccioli (2002) [this volume] and Choudhury *et al.* (2002) [this volume] and by Born & Huang (1954), Venkataraman *et al.* (1975), and Dove (1993). Below we briefly describe only the molecular dynamics technique.

Molecular dynamics (MD)

In MD, we describe the classical system by a set of positions, $\mathbf{r}_i(t)$, and velocities, $\mathbf{v}_i(t)$, of all particles. For a system of *n* atoms there are $(3n - 3)$ degrees of freedom, and the classical kinetic energy is:

$$E_{\text{kin}} = \sum_i \frac{m_i |\mathbf{v}_i|^2}{2} = \frac{(3n-3)k_{\text{B}}T}{2}. \tag{143}$$
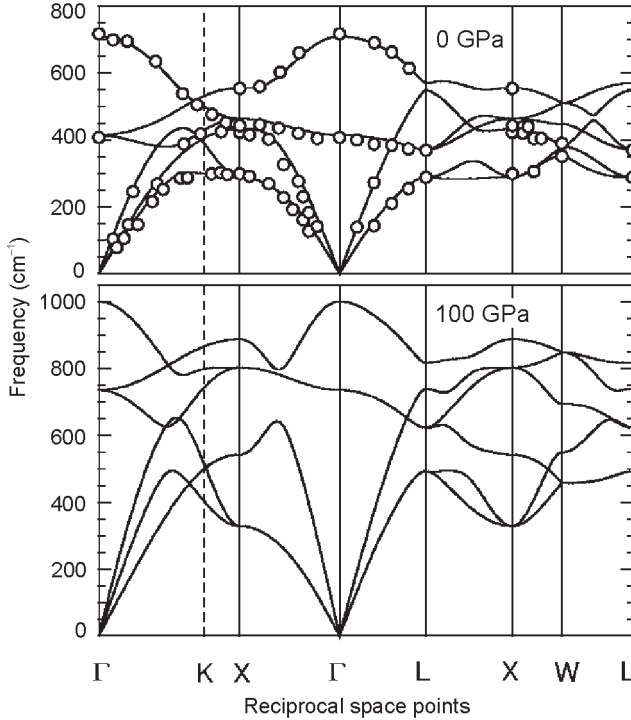
Fig. 19. Phonon dispersion curves of MgO at zero pressure and 100 GPa. Experiment (circles) and results of density functional perturbation theory within the LDA (after Karki *et al.*, 2000a).

The initial kinetic energy (or temperature) is specified as an input, and the initial velocities are assigned to the atoms randomly, according to the Maxwell distribution. The atomic trajectories are constructed by solving Newtonian equations of motion. The initial conditions are likely to be non-equilibrium, so the simulation is divided into two stages: 'equilibration' and 'production' (properties are calculated only from the results of the production stage). Newton's equations of motion are conservative (*i.e.* there is no energy dissipation), the volume and the number of particles are conserved as well, so here we deal with the microcanonical, or *NVE*, ensemble. Some properties can be computed as time averages: one can introduce instantaneous values of properties; *e.g.* for the instantaneous values of the pressure we have:

$$p = \frac{Nk_{\mathrm{B}}T}{V}\delta_{ij} - \frac{1}{3V}\sum_i \mathbf{F}_i\mathbf{r}_i, \tag{144}$$

where $\mathbf{F}_i$ is the force acting on the *i*-th atom. Technical notes: (i) Trajectories can be calculated only by using a finite timestep Δt, rather than infinitesimal dt. The most reliable algorithm for integrating the equation of motion is the Verlet leapfrog algorithm:

$$\mathbf{r}_i(t + \Delta t) = 2\mathbf{r}_i(t) - \mathbf{r}_i(t - \Delta t) \frac{\mathbf{F}_i(t)}{m_i}(\Delta t)^2. \tag{145}$$

The errors in the positions are of the order $(\Delta t)^4$. Δt must be sufficiently small (typically ~ 1 fs). (ii) For simulating bulk materials, large supercells should be used in conjunction with periodic boundary conditions. The larger the system size, the more phonons are included and the better description of long-wavelength fluctuations. Large supercell size allows for more couplings between phonons with different **k**-vectors and therefore improves the description of anharmonicity. The effect of the system size must be always checked. For Lennard–Jones systems, the largest systems studied include millions of atoms, with *ab initio* MD the current limit is a few hundred atoms in the supercell. (iii) To explore all possible states, simulations should run infinitely long. Usually, good statistical averages can be collected over a reasonable time, ~ 1–10 ps.

It is possible to extend MD to other ensembles – *NVT*, *NPT*, *etc.* (letters indicate the conversed parameters: e.g., in the *NVT* ensemble, the number of atoms, volume, and temperature are conserved.) In the *NVT* ensemble, it is possible to keep the temperature constant by simple rescaling of velocities, but this would not result in correct canonical distributions. The most popular and correct way is to use extended Lagrangian formulations.

The Lagrangian function defined (Landau & Lifshitz, 2001) as:

$$L = E_{kin}(\mathbf{v}) - E_{pot}(\mathbf{r}) = \sum_i \frac{m_i|\mathbf{v}_i|^2}{2} - E_{pot}(\mathbf{r}) \tag{146}$$

leads to the following equation of motion:

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \mathbf{v}_i}\right) = \left(\frac{\partial L}{\partial \mathbf{r}_i}\right), \tag{147}$$

from which Newton's equations of motion ($m_i \partial \mathbf{v}_i/\partial t = -\partial E_{pot}(\mathbf{r})/\partial \mathbf{r}_i = \mathbf{F}_i$) follow. The total energy is a constant of motion, *i.e.* is time-independent.

In the constant-temperature method of Nosé (1984), the Lagrangian function (Eqn. 146) is augmented by two other terms and takes the form:

$$L = \sum_i \frac{m_i|\mathbf{v}_i|^2}{2} - E_{pot}(\mathbf{r}) + \frac{1}{2}Q\dot{s}^2 - (f+1)k_B T_0 \ln s, \tag{148}$$

where s is the new dynamical variable, and Q is the associated mass parameter. T_0 is the desired temperature, and f is the number of degrees of freedom in the system. This Lagrangian ensures that temperature fluctuates around T_0 and does not drift. The most important constant of motion is the Hamiltonian, the sum of the real energy and the fictitious Nosé terms. New equations of motion, with modified forces, are:

$$m_i\ddot{\mathbf{r}}_i = \frac{1}{s^2}\mathbf{F}_i - \frac{2}{s}\dot{s}\mathbf{p}_i, \tag{149}$$

$$Q\ddot{s} = \sum_i m_i s |\dot{\mathbf{r}}_i|^2 - \frac{(f+1)k_{\mathrm{B}}T_0}{s}. \tag{150}$$

The choice of the mass parameter Q does not affect the canonical averages in principle, but in order to approach these averages in reasonably short time, it is better to choose Q so that the period of oscillation of the temperature (or s) is similar to the average period of atomic vibrations. The period of oscillations of s is:

$$t_0 = 2\pi \left(\frac{Q\langle s \rangle^2}{2fk_{\mathrm{B}}T_0} \right)^{1/2} \approx 2\pi \left(\frac{Q}{2fk_{\mathrm{B}}T_0} \right)^{1/2}. \tag{151}$$

In the same spirit, Parrinello & Rahman (1981) devised a constant-pressure method. They added into the Lagrangian an extra potential term (pV) and kinetic term ($\frac{1}{2}Q\sum_\alpha \sum_\beta \mathbf{H}_{\alpha\beta}^2$), where $\mathbf{H}_{\alpha\beta}$ is the matrix of lattice vectors).

Car & Parrinello (1985) applied the same trick to construct the first scheme of *ab initio* MD simulations, where both atomic and electronic coordinates (plane wave coefficients $C_{\mathbf{k}+\mathbf{K}}$) evolve simultaneously with time. The Car–Parrinello method is discussed in detail by Remler & Madden (1990). More modern versions of *ab initio* MD are based on the determination of the ground state at each timestep ('Born–Oppenheimer dynamics'); this method is more stable and more suitable for metals. Significant advances in the Earth (Alfè *et al.*, 1999; Oganov *et al.*, 2001b) and planetary (Ancilotto *et al.*, 1997; Cavazzoni *et al.*, 1999) sciences have been obtained with *ab initio* MD simulations. This is the method of choice for accurate simulations of high-temperature phenomena that are significantly anharmonic, such as melting (Sugino & Car, 1995; Alfè *et al.*, 1999), ionic conductivity, displacive phase transitions, thermal expansion (Buda *et al.*, 1990) and elastic (Oganov *et al.*, 2001b) properties.

Most properties can be obtained from MD in one of three ways: (i) by calculating time averages directly, (ii) from fluctuations, (iii) from correlation functions.

Fluctuations provide a convenient route to calculate numerous response properties (heat capacity, elastic constants, Grüneisen parameter *etc.*). Generally, one can write:

$$\frac{dS}{k_{\mathrm{B}}} = \frac{dE}{k_{\mathrm{B}}T} + \xi dX, \tag{152}$$

where X is some extensive variable, and ξ is the associated intensive variable (divided by $k_{\mathrm{B}}T$). A general equation for fluctuations is then (Chandler, 1987):

$$-\left(\frac{\partial \langle X \rangle}{\partial \xi} \right) = \langle (\Delta X)^2 \rangle. \tag{153}$$

Let us consider some particular equations for different ensembles (Allen & Tildesley, 1987). In the *NVT* ensemble one has:

$$\langle (\Delta E)^2 \rangle_{\mathrm{NVT}} = k_{\mathrm{B}}T^2 C_{\mathrm{V}}. \tag{154}$$

In the *NVE* ensemble:

$$\langle(\Delta E_{pot})^2\rangle_{NVE} = \langle(\Delta E_{kin})^2\rangle_{NVE} = \frac{3}{2}Nk_B^2T^2\left(1 - \frac{3Nk_B}{2C_V}\right). \tag{155}$$

In the *NPT* ensemble:

$$\langle(\Delta V)^2\rangle_{NPT} = \frac{Vk_BT}{K_T}, \tag{156}$$

$$\langle[\Delta(E+pV)]^2\rangle_{NPT} = k_BT^2C_p, \tag{157}$$

$$\langle\Delta V\Delta(E+pV)\rangle_{NPT} = k_BT^2V\alpha_p. \tag{158}$$

Parrinello & Rahman (1982) have derived equations for the determination of the elastic constants from strain fluctuations in the *NPT* ensemble:

$$\langle\Delta e_{ij}\Delta e_{kl}\rangle_{NPT} = \frac{k_BT}{V}S_{ijkl}^T. \tag{159}$$

For more extended treatments of fluctuations, see Landau & Lifshitz (1976), Allen & Tildesley (1987), Dove (1988), Cheung (1977). Landau & Lifshitz (1976) derived the general theory of fluctuations from the equilibrium distribution functions.

Correlation functions provide an important tool to study relaxation processes and transport properties (*e.g.*, viscosity) and power spectra, the most important of which is the phonon density of states. The velocity autocorrelation function for an *i*-th atom is defined as:

$$C_i(t) = \frac{\langle \mathbf{v}_i(0)\mathbf{v}_i(t)\rangle}{\langle|\mathbf{v}_i(0)|^2\rangle} = \frac{(\lim\tau\to\infty)\dfrac{1}{\tau}\displaystyle\int_0^\tau \mathbf{v}_i(t')\mathbf{v}_i(t'+t)dt'}{\langle\mathbf{v}_i^2\rangle}. \tag{160}$$

The phonon density of states is just a Fourier transform of the mass-weighted sum of atomic velocity autocorrelation functions (Dove, 1993):

$$g(\omega) \sim \frac{1}{T}\int\left[\sum_i m_iC_i(t)\right]\cos(\omega t)dt. \tag{161}$$

The free energy cannot be determined from MD directly; such methods as thermodynamic integration (Allen & Tildesley, 1987; Sugino & Car, 1995) are used for that purpose.

Standard MD uses the classical approximation and cannot be applied at low temperatures, where quantum effects are essential[18]. Quantum effects can be incorporated by (i) path integral MD (see Allen & Tildesley, 1987), (ii) using the phonon spectrum $g(\omega)$ calculated in MD, (iii) applying quantum corrections. The quantum

---

[18] Some properties (*e.g.*, mode-average Grüneisen parameter) can be derived from classical MD at any temperature. However, there are difficulties in equilibrating the system at low temperatures, where the vibrations are nearly harmonic and there is practically no energy transfer between vibrational modes.

correction for the Helmholtz free energy per atom in the lowest order is (Landau & Lifshitz, 1976):

$$\Delta F = F(\text{quantum}) - F(\text{classical}) = \frac{\hbar^2}{24 k_B T} \left\langle \sum_i \frac{\nabla_i^2 E_{\text{pot}}}{m_i} \right\rangle, \qquad (162)$$

where ∇_i^2 is the Laplacian with respect to the coordinates of the i-th atom. Higher-order ($\hbar^3$ and higher) corrections are needed only at temperatures below $\sim \theta_D/2$. Quantum corrections to other properties can be worked out by differentiating Equation 162 – see Matsui (1989).

Part V: Examples from recent studies

Here we present results of recent quantum-mechanical calculations performed in our group. These results illustrate the theoretical concepts outlined above and the enormous potential of *ab initio* simulations for studies of minerals and planetary materials. All the calculations discussed below were performed using the PW91 GGA (Wang & Perdew, 1991) and pseudopotential plane wave technique. Some calculations were also performed using the projector augmented-wave method (Blöchl, 1994; Kresse & Joubert, 1999), which is an all-electron frozen-core method. All our calculations were performed with the VASP (*Vienna Ab Initio Simulation Package*, Kresse & Furthmüller, 1996) code.

Phase diagram of Al_2SiO_5

This phase diagram is very important in metamorphic petrology and has been a topic of a large number of papers over the last 60 years. The three known polymorphs – kyanite, andalusite, and sillimanite – are common minerals in the Earth's crust and (only kyanite) the upper mantle. Recent experiments suggested that this system can be important for the lower mantle as well: Ahmed-Zaid & Madon (1991, 1995) found that at high temperatures and pressures of the lower mantle kyanite transforms into what these authors interpreted as a V_3O_5-like phase of Al_2SiO_5 (where both Al and Si are in the octahedral coordination). These authors suggested that this new phase is the main Al-containing mineral in the lower mantle. However, most experiments (*e.g.*, Schmidt *et al.*, 1997) indicate that the Al_2SiO_5 polymorphs decompose into the mixture of oxides at high pressures.

Our simulations (Oganov & Brodholt, 2000) indicate that the latter is indeed the case. Neither the known polymorphs (kyanite, andalusite, sillimanite), nor the hypothetical V_3O_5-like or pseudobrookite-like phases (where Al and Si are also octahedrally coordinated) are thermodynamically stable above 11 GPa and therefore cannot exist in the lower mantle. From 0 K calculations, we determined the stable and metastable transition pressures; using experimetal dp/dT slopes we were able to construct a phase diagram (Fig. 20), whose main difference from experiment is the overestimation of all transition pressures by ~ 2 GPa. The GGA calculations were also successful in predicting crystal structures, atomisation energies, and EOSs of these minerals.
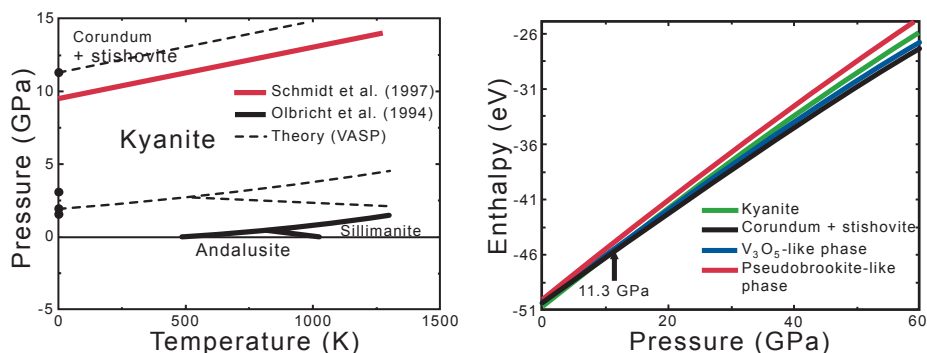
Fig. 20. (a) Predicted (dashed lines) and experimental (solid lines) phase diagrams of Al_2SiO_5. **(b)** Enthalpy as a function of pressure for the candidate high-pressure phases.

Metastable Al_2SiO_5 phases

Although Al_2SiO_5 polymorphs are thermodynamically unstable above 11 GPa, they can exist as metastable phases at much higher pressures, if temperature is low enough to make the transition to the stable state (*i.e.* decomposition) kinetically hindered. Calculations predict extremely interesting behaviour of these polymorphs at high pressures (Oganov *et al.*, 2001d).

Ab initio simulations show that while dense polymorphs (kyanite and the hypothetical phases) remain in the metastable state up to at least 80 GPa, the low-density phases (andalusite and sillimanite) sponateously transform into other metastable phases at ~ 40 GPa. It is interesting to note that semiclassical simulations based on the ionic shell model produce semiquantitatively similar results. To study the nature of these transitions in more detail, we also calculated phonon dispersion curves with the ionic shell model (to see if there are any soft modes) and introduced small random atomic
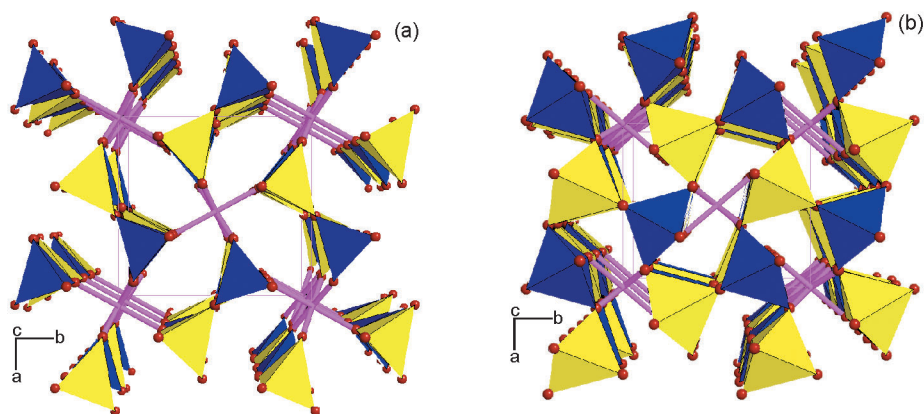
Fig. 21. (a) Crystal structure of sillimanite and **(b)** post-sillimanite metastable high-pressure phase. Blue = Si polyhedra, yellow = Al polyhedra, purple sticks = octahedral Al–O bonds.

Fig. 22. Geometry of the SiO_5 polyhedra in the post-sillimanite structure (prediction at 50 GPa).
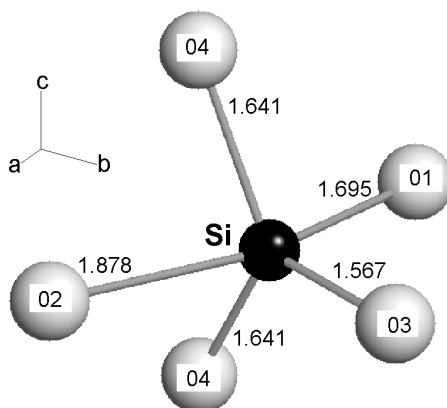
displacements and deformations of the unit cell in GGA calculations (also to test dynamical stability, but with *ab initio* simulations).

As these calculations show, sillimanite undergoes a transition, upon which Si and Al atoms transform from tetrahedrally coordinated to 5-fold coordinated (Fig. 21). The geometry of the SiO_5 polyhedron in the post-sillimanite phase is shown in Figure 22. The transition is isosymmetric (both phases have *Pbnm* symmetry) and, in accordance with Landau theory, is first order. It has a pronounced hysteresis between 28 GPa and 38 GPa and peculiar non-monotonic variation of the *c* parameter (Fig. 23). Simulations also indicate that sillimanite can be incommensurately modulated at pressures within the hysteresis loop.

Pressure-induced amorphisation is predicted for andalusite at 52 GPa (at zero temperature, or between 52 GPa and 34 GPa at non-zero temperatures). The post-andalusite structure is dynamically unstable and the predicted type of dynamical instability (Fig. 24) indicates that pressure-induced amorphisation should occur along the *c* and, possibly, *a* axes.

Equation of state of MgO

We studied periclase (MgO) by using both pseudopotential and all-electron frozen-core PAW calculations (Oganov & Dorogokupets, submitted). Both small-core ('Be core', *i.e.* $1s^2 2s^2$ core and $2p^6 3s^2$ valence configuration) and large-core ('Ne core', with $3s^2$ valence configuration) potentials were used for Mg, while O in all cases was described with $1s^2$ core and $2s^2 2p^4$ valence configuration. The calculated EOSs and the experimental results are shown in Figure 25. Mg (as well as other alkali earth elements) is known to have rather diffuse 'semicore' $2p$ orbitals, which significantly overlap with the valence $3s$ orbitals and (at high pressures) also with the orbitals of the neighbouring atoms; this circumstance creates problems for the pseudopotential approach. Inclusion of nonlinear core corrections (Louie *et al.*, 1982) is known to substantially improve pseudopotentials, but as we see in Figure 25 results of such calculations still differ visibly from more accurate PAW and small-core pseudopotential calculations. Since PAW is an all-electron method, it fully includes the above-mentioned effects, and even large-core PAW
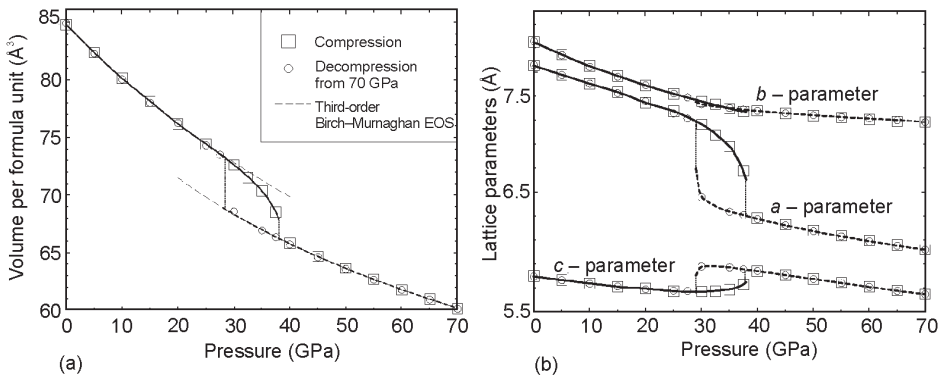
Fig. 23. Pressure evolution of *(a)* unit cell volume and *(b)* lattice parameters of sillimanite.
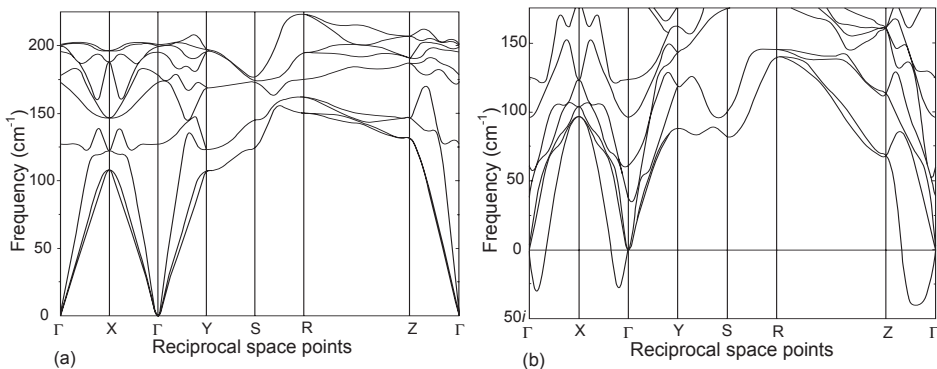

Fig. 24. Phonon dispersion curves of *(a)* andalusite at room pressure and *(b)* of the high-pressure dynamically unstable post-andalusite phase. Only several lowest-frequency phonon branches are shown.

calculations give very accurate results. The excellent performance of large-core PAW calculations also suggests that the effects of Mg core polarisation are small.

We find that the transition to the CsCl structure occurs at pressures too high for the Earth's mantle – the calculated transition pressure is 509 GPa with the most accurate small-core PAW calculations (Fig. 26). This value exceeds by 50% the pressure in the centre of the Earth! Large-core pseudopotential calculations give the pressure of 488 GPa. Among the most accurate theoretical results, LDA pseudopotential calculations (Karki *et al.*, 1997) give 451 GPa and all-electron (LAPW), LDA calculations of Mehl *et al.* (1988) give 510 GPa, whereas both LDA and GGA all-electron calculations of Jaffe *et al.* (2000) give 515 GPa. These results are consistent with the experimental observation of the remarkable stability of the NaCl structure for MgO: no phase transitions were observed up to at least 227 GPa (Duffy *et al.*, 1995). This stability makes MgO very convenient as a pressure standard in ultrahigh-pressure experiments.

Using *ab initio* MD, we have calculated the thermal EOS, Grüneisen parameters, and intrinsic anharmonicity parameters of MgO. The results (Fig. 27) suggest that the temperature variation of the Grüneisen parameter is mild and the Mie–Grüneisen

Fig. 25. Equation of state of MgO from experiment and GGA calculations. PAW = Projector Augmented-Wave calculations, ECP = Effective Core Pseudopotential calculations.
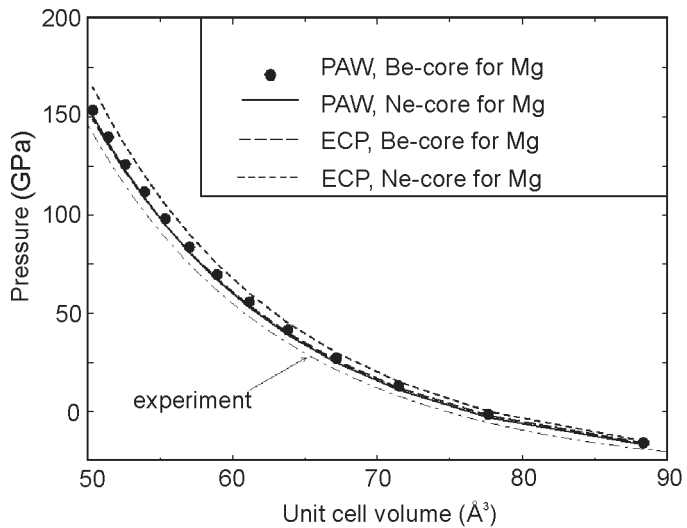
approximation is adequate. Different contributions to the total pressure at 3000 K are shown in Figure 28. The anharmonic pressure is negative and small, –3 GPa at its utmost, which is quantitatively similar to the experimental data (Dorogokupets & Oganov, in prep.). However, thermal expansion is more sensitive to intrinsic anharmonicity than pressure. Karki *et al.* (1999, 2000a) recently studied lattice dynamics, thermodynamic properties, and EOS of MgO within the quasiharmonic approximation in their pseudopotential LDA calculations based on density functional perturbation theory.
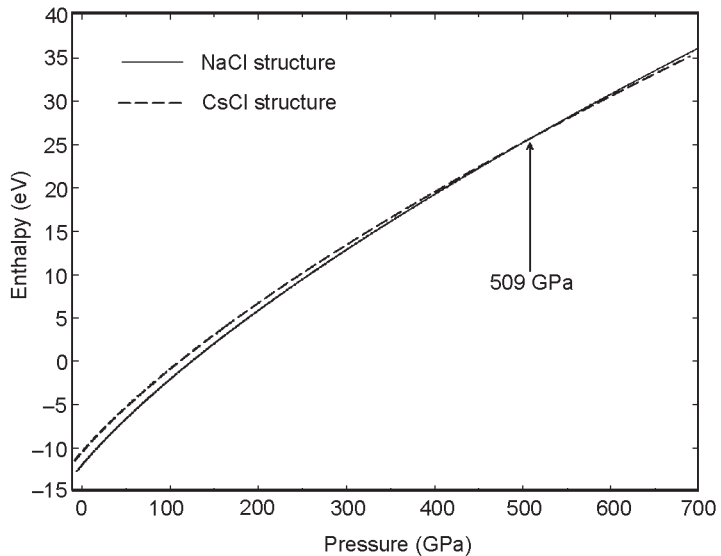
Fig. 26. Phase transition of MgO from the NaCl to the CsCl structure from PAW (Be-core) calculations.
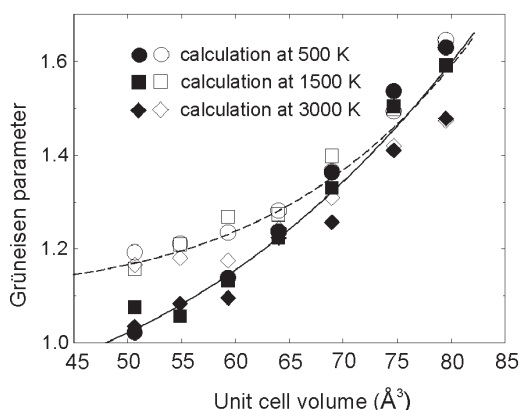
Fig. 27. Effective Grüneisen parameter of MgO from *ab initio* MD. Solid symbols = Be-core PAW results; open symbols = Ne-core pseudopotential calculations. Lines = best fits.

Equation of state and elasticity of MgSiO₃ perovskite

MgSiO$_3$ perovskite is a mineral of extreme geophysical importance. This importance and recent progress in *ab initio* calculations explain the simultaneous appearance of several recent theoretical papers on this mineral (Oganov *et al.*, 2001a, 2001b, Oganov *et al.*, in prep.; Karki, 2000b, 2001). Our calculations suggest that in MgSiO$_3$ perovskite large-core pseudopotentials perform very well, unlike in MgO. This is probably related to longer Mg–O bonds in perovskite and, consequently, much smaller overlap between the 2p orbitals of Mg and valence orbitals of O atoms.

  Ab initio MD calculations (Oganov *et al.*, 2001a, 2001b) show that MgSiO$_3$ perovskite remains orthorhombic (Fig. 29) throughout the lower mantle. This agrees with our static simulations (Fig. 30; Oganov *et al.*, in prep.), which show that higher-symmetry perovskite structures are much less dense and very much higher in energy than
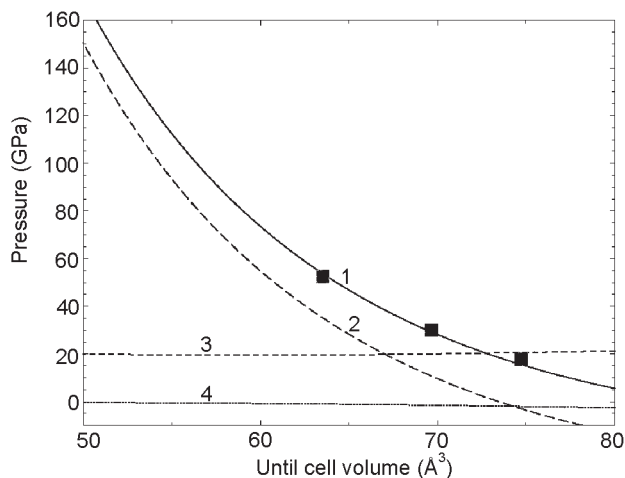

Fig. 28. Theoretical equation of state of MgO at 3000 K (1) and different pressure contributions: (2) static, (3) quasiharmonic and (4) anharmonic. Large grey squares = experiment-based extrapolations of Speziale *et al.* (2001) at 3000 K. Theoretical results (lines 1–4) are based on Be-core PAW calculations.

Fig. 29. 'Snapshot' of the crystal structure of $MgSiO_3$ perovskite at 88 GPa and 3500 K from *ab initio* MD simulations.
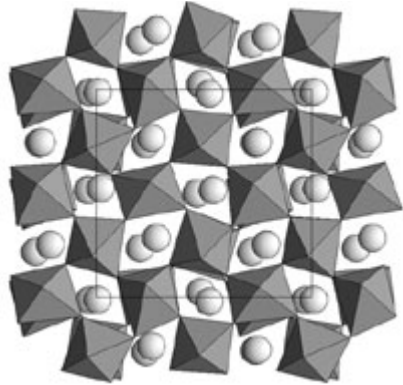
the observed *Pbnm* phase. The same result has been obtained in previous works (Stixrude & Cohen, 1993; Wentzcovitch *et al.*, 1993; D'Arco *et al.*, 1993b; Warren *et al.*, 1998). Earlier, the hypothetical transition to a cubic ($Pm\overline{3}m$) perovskite phase, which is expected to be a superionic conductor of electricity (O'Keeffe & Bovin, 1979; Matsui & Price, 1991) due to the diffusion of O^{2-} ions, was invoked to explain the high measured electrical conductivity of the lower mantle.

To construct the EOS, we calculated the phonon pressure using the Grüneisen parameters from *ab initio* MD and vibrational energy from the Debye model. The calculated thermal expansion (Fig. 31) agrees well with the results of Karki *et al.* (2001) at high pressures. At low pressures and high temperatures, as expected, the quasiharmonic approximation used by Karki *et al.* (2001) significantly overestimates thermal expansion. Our thermal expansion values at 0 GPa strongly support the most recent experiments, which indicate thermal expansion that is too low to be consistent with a pure-perovskite lower mantle. Karki *et al.* (2001) found that pure-perovskite model would agree with seismological profiles only for unrealistically high temperatures, thus ruling out this model. Probably, the lower mantle composition is pyrolitic or similar.

Our work (Oganov *et al.*, 2001b) reported on what seem to be the first calculations of the elastic constants of any material at finite temperatures using *ab initio* MD. These calculations for $MgSiO_3$ perovskite at the temperatures and pressures of the lower mantle give $R_T \approx 0.7$ in agreement with geophysical observations. The calculated R_p increases from 1.5 at the depth of 1000 km to 1.9 at 2000 km. The remaining deficit of R_p can be explained by significant anelasticity (Karato, 1993) throughout the lower mantle, as well as significant compositional heterogeneity (*e.g.*, Masters *et al.*, 2000) below 2000–2500 km.

With our results on temperature variation of the elastic constants, we were able to find (Oganov *et al.*, 2001b) from seismic tomography maps of Masters *et al.* (2000) that the maximum temperature contrast between the hot and cold streams increases from 900 K at 1000 km depth to 1500 K at 2000 km and possibly to ~ 2000 K at the core–mantle boundary. These values should play an important role in the future models of the Earth's dynamics and can be used as important constraints in numerical models of mantle convection.
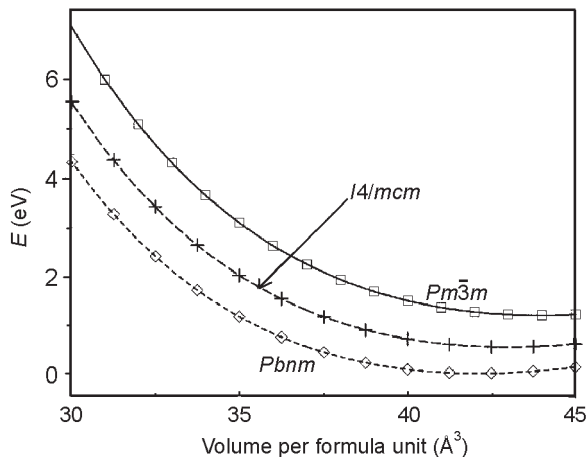
Fig. 30. $E(V)$ curves for the orthorhombic, tetragonal, and cubic phases of $MgSiO_3$ perovskite: GGA calculations. Minimum energy of the *Pbnm* phase is set as the reference zero. Energies are per formula unit.

Towards an *ab initio* thermal model of the Earth

Here we show (Oganov & Price, in prep.) that, having accurate *ab initio* results and geophysical measurements, we are not far away from establishing an accurate reference thermal model for the Earth. Let us again consider the lower mantle.
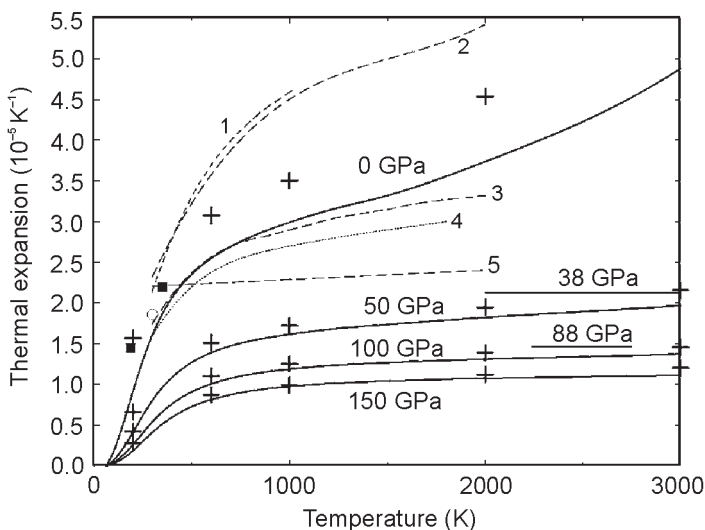

Fig. 31. Thermal expansion of $MgSiO_3$ perovskite. Solid lines – results of Oganov *et al.* (2001a). Solid horizontal lines – direct *ab initio* MD results at 38 GPa and 88 GPa between 1500 K and 3500 K. Crosses – calculations of Karki *et al.* (2001). Experimental data at 0 GPa: squares (Ross & Hazen, 1989); lines (1) Knittle & Jeanloz (1986), (2) Mao *et al.* (1991), (3) Wang *et al.* (1994), (4) Funamori *et al.* (1996), (5) Fiquet *et al.* (2000).
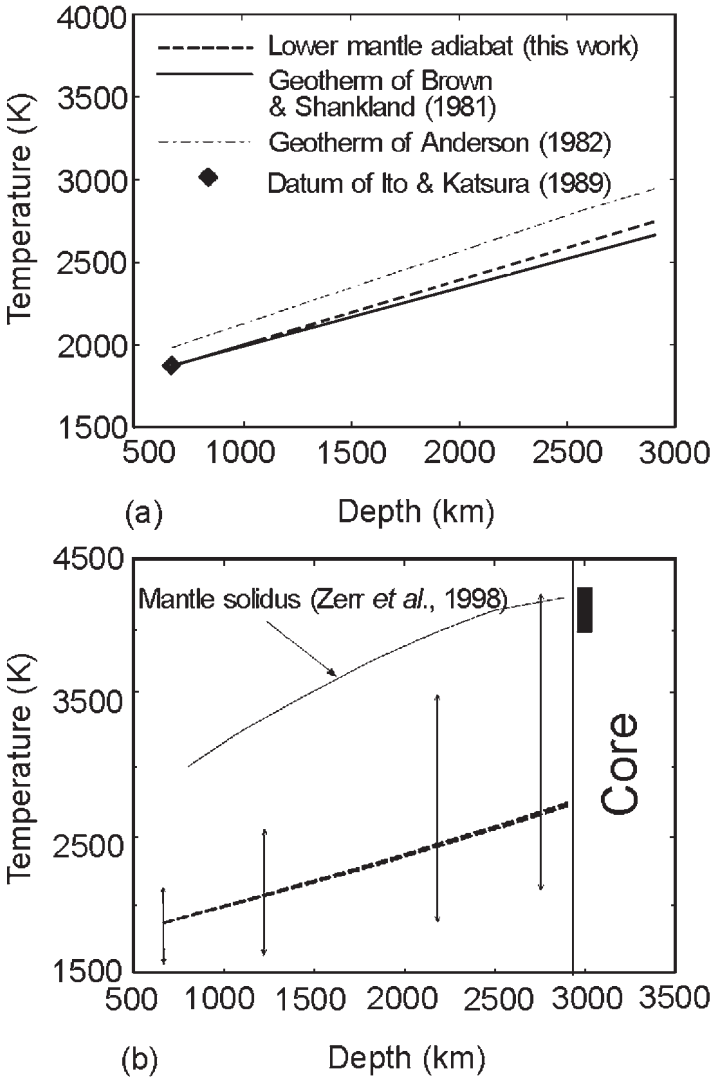
Fig. 32. (a) Adiabatic geotherm (this work) in comparison with the geotherms of Brown & Shankland (1981) and Anderson (1982). **(b)** Schematic temperature distribution in the lower mantle. Two-headed arrows indicate the maximum temperature contrasts (the lower arrow gives the lowest possible temperature at a given depth; the upper arrow gives the maximum temperature). Thick rectangle in (b) corresponds to current estimates of temperature at the core–mantle boundary.

Since the lower mantle is a convecting system, its temperature distribution should be not far from adiabatic:

$$T(z) = \int_{z_1 = 670 \, \text{km}}^{z} T(z)\gamma(V)\left(\frac{d \ln \rho}{dz}\right) dz \,, \tag{163}$$

where we can use $T(670 \text{ km}) = 1873 \pm 100$ K determined from phase equilibria (Ito & Katsura, 1989) and geophysical (PREM) density profile $\rho(z)$. Figure 32 shows that two adiabats obtained with *ab initio* $\gamma(V)$ of MgSiO$_3$ and MgO (which together comprise ~ 90–95% of the lower mantle) are practically indistinguishable; therefore, we have an accurately constrained adiabatic geotherm (between these two adiabats). This adiabat is quite similar to the classical geotherms of Brown & Shankland (1981) and Anderson (1982).

For the core side of the core-mantle boundary, experiments of Boehler (1996) suggest $T \approx 4000$ K. In good agreement with this, *ab initio* simulations (D. Alfè, personal communication) give ~ 4200 K. This gives the highest temperature of the mantle. From this and from our temperature contrasts we see that the minimum temperature of the mantle near the core–mantle boundary is 2100 K. Using the starting temperature $T(670 \text{ km}) = 1873$ K, the core–mantle boundary temperature (4000–4400 K), our mantle adiabat and temperature contrasts, one can obtain a very interesting general picture (Fig. 32). Comparing this picture with experimental solidus of pyrolite determined by Zerr *et al.* (1998), we expect some degree of partial melting, restricted only to the hottest regions in the very bottom of the lower mantle. This agrees with seismological observations (Williams & Garnero, 1996) and would have important implications for the anomalous properties of the core–mantle D″ layer (anomalous seismic signatures such as ultralow-velocity zones, ionic conductivity, kinetics of chemical equilibration with the core *etc.*).

Acknowledgements

References

Adams, D.M., Hatton, P.D., Heath, A.E. & Russell, D.R. (1988): X-ray diffraction measurements on potassium nitrate under high pressure using synchrotron radiation. *J. Phys. C. Solid State Phys.*, **21**:505–515.

Agoshkov, V.M. (1985): Calculation of thermodynamic properties of minerals based on a model of the vibrational spectrum for silicates and oxides. *Geokhimiya*, **(10)**:1441–1454. (in Russian)

Agoshkov, V.M., Kieffer, S.W. & McMillan, P.F. (1994): Lattice dynamics and thermodynamic properties of minerals. *In* Marfunin, A.S. (ed.): *Advanced mineralogy*, **1**:419–430. Berlin: Springer-Verlag.

Ahmed-Zaid, I. & Madon, M. (1991): A high-pressure form of Al$_2$SiO$_5$ as a possible host of aluminium in the lower mantle. *Nature*, **353**:426–428.

Ahmed-Zaid, I. & Madon, M. (1995): Electron microscopy of high-pressure phases synthesized from natural garnets in a diamond anvil cell: Implications for the mineralogy of the lower mantle. *Earth Planet. Sci. Lett.*, **129**:233–247.

Alan, D.R. & Nelmes, R.J. (1996): The structural pressure dependence of potassium titanyl phosphate (KTP) to 8 GPa. *J. Phys., Condens. Matter*, **8**:2337–2363.

Alavi, A., Lozovoi, A. & Finnis, M.W. (1999): Pressure-induced isostructural phase transition in Al-rich NiAl alloys. *Phys. Rev. Lett.*, **83**:979–982.

Alexandrov, K.S. & Prodaivoda, G.T. (1993): Elastic properties of minerals. *Kristallografiya*, **38**:214–234. (in Russian)

Alfé, D., Gillan, M.J. & Price, G.D. (1999): The melting curve of iron at the pressures of the Earth's core from ab initio calculations. *Nature*, **401**:462–464.

Alfé, D., Gillan, M.J. & Price, G.D. (2002): Composition and temperature of the Earth's core constrained by combining ad initio calculations and seismic data. *Earth Planet. Sci. Lett.* **195**:91–98.

Allègre, C.J., Poirier, J.-P., Humler, E. & Hofmann, A.W. (1995): The chemical composition of the Earth. *Earth Planet. Sci. Lett.*, **134**:515–526.

Allen, M.P. & Tildesley, D.J. (1987): *Computer simulation of liquids.* Oxford: Clarendon Press, 385 p.

Ancilotto, F., Chiarotti, G.L., Scandolo, S. & Tosatti, E. (1997): Dissociation of methane into hydrocarbons at extreme (planetary) pressure and temperature. *Science*, **275**:1288–1290.

Anders, E. & Ebihara, M. (1982): Solar system abundances of the elements. *Geochim. Cosmochim. Acta*, **46**:2363–2380.

Anderson, D.L. (1989): *Theory of the Earth*. Boston: Blackwell Sci. Publ., 366 p.

Anderson, O.L. (1982): The Earth's core and the phase diagram of iron. *Philos. Trans. R. Soc. Lond.*, **A306**:21–35.

Angel, R.J. & Ross, N.L. (1996): Compression mechanisms and equations of state. *Philos. Trans. R. Soc. Lond.*, **A354**:1449–1459.

Angel, R.J., Ross, N.L., Seifert, F. & Fliervoet, T.F. (1996): Structural characterization of pentacoordinate silicon in a calcium silicate. *Nature*, **384**:441–444.

Angel, R.J., Allan, D.R., Miletich, R. & Finger, L.W. (1997): The use of quartz as an internal pressure standard in high-pressure crystallography. *J. Appl. Crystallogr.*, **30**:461–466.

Aprà, E., Stefanovich, E., Dovesi, R. & Roetti, C. (1991): An ab initio Hartree–Fock study of silver chloride. *Chem. Phys. Lett.*, **186**:329–335.

Arlt, T. & Angel, R.J. (2000): Displacive phase transitions in C-centred clinopyroxenes: spodumene, $LiScSi_2O_6$ and $ZnSiO_3$. *Phys. Chem. Miner.*, **27**:719–731.

Arlt, T., Angel, R.J., Miletich, R., Armbruster, T. & Peters, T. (1998): High-pressure $P2_1/c$–$C2/c$ phase transitions in clinopyroxenes: influence of cation size and electronic structure. *Am. Mineral.*, **83**:1176–1181.

Ashcroft, N.W. & Mermin, N.D. (1976): *Solid state physics*. New York (N.Y.): Holt, Rinehart & Winston.

Bader, R.F.W. (1990): *Atoms in molecules. A quantum theory.* Oxford: Oxford Univ. Press, 438 p.

Badro, J., Barrat, J.-L. & Gillet, P. (1996): Numerical simulation of α-quartz under nonhydrostatic compression: memory glass and five-coordinated crystalline phases. *Phys. Rev. Lett.*, **76**:772–775.

Badro, J., Teter, D.M., Downs, R.T., Gillet, P., Hemley, R. & Barrat, J.-L. (1997): Theoretical study of a five-coordinated silica polymorph. *Phys. Rev., B, Condens. Matter*, **56**:5797–5806.

Badro, J., Struzhkin, V.V., Shu, J., Hemley, R.J., Mao, H.-K., Kao, C.-C., Rueff, J.-P. & Shen, G. (1999): Magnetism in FeO at megabar pressures from X-ray emission spectroscopy. *Phys. Rev. Lett.*, **83**:4101–4104.

Baerends, E.J. & Gritsenko, O.V. (1997): A quantum chemical view of density functional theory. *J. Phys. Chem., A, Mol. Spectrosc. Kinet. Environ. Gen. Theory*, **101**:5383–5403.

Baroni, S., Gianozzi, P. & Testa, A. (1987): Green-function approach to linear response in solids. *Phys. Rev. Lett.*, **58**:1861-1864.

Baroni, S., de Gironcoli, S., Dal Corso, A. & Gianozzi, P. (2001): Phonons and related crystal properties from density-functional perturbation theory. *Rev. Mod. Phys.*, **73**:515–562.

Barron, T.H.K. & Klein, M.L. (1965): Second-order elastic constants of a solid under stress. *Proc. Phys. Soc.*, **85**:523–532.

Becke, A.D. (1993): Density-functional thermochemistry. 3. The role of exact exchange. *J. Chem. Phys.*, **98**:5648–5652.

Becke, A.D. (2000): Simulation of delocalised exchange by local density functionals. *J. Chem. Phys.*, **112**:4020–4026.

Belikov, B.P., Aleksandrov, K.S. & Ryzhova, T.V. (1970): *Elastic constants of rock-forming minerals.* Moscow: Nauka, 276 p (in Russian).

Bina, C.R. & Navrotsky, A. (2000): Possible presence of high-pressure ice in cold subducting slabs. *Nature*, **408**:844–847.

Birman, J.L. (1966): Simplified theory of symmetry change in second-order phase transitions: application to V_3Si. *Phys. Rev. Lett.*, **17**:1216–1219.

Blöchl, P.E. (1994): Projector augmented-wave method. *Phys. Rev., B, Condens. matter*, **50**:17,953–17,979.

Boehler, R. (1996): Melting temperatures of the Earth's mantle and core: Earth's thermal structure. *Annu. Rev. Earth Planet. Sci.*, **24**:15–40.

Born, M. & Huang, K. (1954): *Dynamical theory of crystal lattices.* Oxford: Clarendon Press, 420 p.

Bowley, R. & Sánchez, M. (1999): *Introductory statistical mechanics.* Second edition. Oxford: Oxford Univ. Press, 352 p.

Brown, J.M. & Shankland, T.J. (1981): Thermodynamic properties in the Earth as determined from seismic profiles. *Geophys. J. R. Astr. Soc.*, **66**:579–596.

Bruce, A.D. & Cowley, R.A. (1981): *Structural phase transitions.* London: Francis and Taylor, 326 p.

Buda, F., Car, R. & Parrinello, M. (1990): Thermal expansion of c-Si via ab initio molecular dynamics. *Phys. Rev., B, Condens. Matter*, **41**:1680–1683.

Buerger, M.J. (1961): Polymorphism and phase transformations. *Fortschr. Mineral.*, **39**:9–24.

Bukowinski, M.S.T. (1994): Quantum geophysics. *Ann. Rev. Earth Planet. Sci.*, **22**:167–205.

Burdett, J.K. (1995): *Chemical bonding in solids.* New York (N.Y.): Oxford Univ. Press.

Car, R. & Parrinello, M. (1985): Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett.*, **55**:2471–2474.

Carlson, S., Xu, Y.Q., Hålenius, U. & Norrestam, R. (1998): A reversible, isosymmetric, high-pressure phase transition in Na_3MnF_6. *Inorg. Chem.*, **37**:1486–1492.

Carpenter, M.A. (2002): Microscopic strain, macroscopic strain and the thermodynamics of phase transitions in minerals. *In* Gramaccioli, C.M. (ed.): Energy modelling in minerals /*EMU Notes Mineral.,* **4**/. Budapest: Eötvös Univ. Press, 311–346.

Catlow, C.R.A. & Mackrodt, W.C. (eds.) (1982): *Computer simulation of solids /Lect. Notes Phys.,* **166**/. Berlin: Springer-Verlag, 320 p.

Catlow, C.R.A. & Price, G.D. (1990): Computer modelling of solid-state inorganic materials. *Nature*, **347**:243–248.

Catlow, C.R.A., Bell, R.G. & Gale, J.D. (1994): Computer modelling as a technique in materials chemistry. *J. Mater. Chem.*, **4**:781–792.

Catti, M. (1989): Crystal elasticity and inner strain – a computational model. *Acta Crystallogr.*, **A45**:20–25.

Catti, M. (2001): Orthorhombic intermediate state in the zinc blende to rocksalt transformation path of SiC at high pressure. *Phys. Rev. Lett.*, **87**: paper 035504.

Catti, M., Pavese, A., Dovesi, R., Roetti, C. & Causà, M. (1991): Quantum-mechanical Hartree–Fock self-consistent field study of the elastic constants and chemical bonding of MgF_2 (sellaite). *Phys. Rev., B, Condens. Matter*, **44**:3509–3517.

Catti, M., Valerio, G., Dovesi, R. & Causà, M. (1994): Quantum-mechanical calculation of solid-state equilibrium $MgO + Al_2O_3 = MgAl_2O_4$ (spinel) versus pressure. *Phys. Rev., B, Condens. Matter*, **49**:14,179–14,187.

Cavazzoni, C., Chiarotti, G.L., Scandolo, S., Tosatti, E., Bernasconi, M. & Parrinello, M. (1999): Superionic and metallic states of water and ammonia at giant planet conditions. *Science*, **283**:44–46.

Ceperley, D.M. & Alder, B.J. (1980): Ground state of the electron gas by a stochastic method. *Phys. Rev. Lett.*, **45**:566–569.

Chandler, D. (1987): *Introduction to modern statistical mechanics.* New York (N.Y.): Oxford Univ. Press, 274 p.

Cheung, P.S.Y. (1977): On the calculation of specific heats, thermal pressure coefficients and compressibilities in molecular dynamics simulations. *Mol. Phys.*, **33**:519–526.

Chizmeshya, A.V.G., Wolf, G.H. & McMillan, P.F. (1996): First-principles calculation of the equation-of-state, stability, and polar optic modes of $CaSiO_3$ perovskite. *Geophys. Res. Lett.*, **23**:2725–2728. (correction: *ibid.* (1998), **25**:711).

Choudhury, N., Chaplot, S.L., Ghose, S., Rao, H.N. & Mittal, R. (2002): Lattice dynamics, inelastic, neutron scattering and thermodynamic properties of minerals. *In* Gramaccioli, C.M. (ed.): Energy modelling in minerals /*EMU Notes Mineral.,* **4**/. Budapest: Eötvös Univ. Press, 211–243.

Christy, A.G. (1993): Multistage diffusionless pathways for reconstructive phase transitions: application to binary compounds and calcium carbonate. *Acta Crystallogr.*, **B49**:987–996.

Christy, A.G. (1995): Isosymmetric structural phase transitions: Phenomenology and examples. *Acta Crystallogr.*, **B51**:753–757.

Chudinovskikh, L. & Boehler, R. (2001): High-pressure polymorphs of olivine and the 660-km seismic discontinuity. *Nature*, **411**:574–577.

Cohen, R.E. (1991): Bonding and elasticity of stishovite SiO_2 at high pressure: Linearized augmented plane wave calculations. *Am. Mineral.*, **76**:733–742.

Cohen, R.E. (1999): Bonding and electronic structure of minerals. *In* Catlow, C.R.A. & Wright, K. (ed.): *Microscopic properties and processes in minerals /NATO ASI Ser. C*, **543**/. Dordrecht: Kluwer, 201–264.

Cohen, R.E., Mazin, I.I. & Isaak, D.G. (1997): Magnetic collapse in transition metal oxides at high pressure: implications for the Earth. *Science*, **275**:654–657.

Cohen, R.E., Gulseren, O. & Hemley, R.J. (2000): Accuracy of equation-of-state formulations. *Am. Mineral.*, **85**:338–344.

Coppens, P. (1997): *X-ray charge densities and chemical bonding /Int. Union Crystallogr. Texts*, **4**/. Oxford: Oxford Univ. Press, 358 p.

Cowley, R.A. (1976): Acoustic phonon instabilities and structural phase transitions. *Phys. Rev., B, Condens. Matter*, **13**:4877–4885.

D'Arco, Ph., Causà, M., Roetti, C. & Silvi, B. (1993a): Periodic Hartree–Fock study of a weakly bonded layer structure: brucite $Mg(OH)_2$. *Phys. Rev., B, Condens. Matter*, **47**:3522–3529.

D'Arco, Ph., Sandrone, G., Dovesi, R., Orlando, R. & Saunders, V.R. (1993b): A quantum mechanical study of the perovskite structure type of $MgSiO_3$. *Phys. Chem. Miner.*, **20**:407–414.

D'Arco, Ph., Sandrone, G., Dovesi, R., Aprà, E. & Saunders, V.R. (1994): A quantum-mechanical study of the relative stability under pressure of $MgSiO_3$-ilmenite, $MgSiO_3$-perovskite, and MgO-periclase + SiO_2-stishovite assemblage. *Phys. Chem. Miner.*, **21**:285–293.

Dam, B., Janner, A. & Donnay, J.D.H. (1985): Incommensurate morphology of calaverite $(AuTe_2)$ crystals. *Phys. Rev. Lett.*, **55**:2301–2304.

Deuss, A. & Woodhouse, J. (2001): Seismic observations of splitting of the mid-transition zone discontinuity in Earth's mantle. *Science*, **294**:354–357.

Dorogokupets, P.I. (2000): Self-consistent thermodynamic functions in equations of state for minerals. *Dokl. Earth Sci.*, **375A**, 1459–1462.

Dorogokupets, P.I. & Oganov, A.R.: Self-consistent equation of state of periclase up to 200 GPa and 3500 K. (in prep.)

Dove, M.T. (1988): Molecular dynamics simulations in the solid state sciences. *In* Salje, E.K.H. (ed): *Physical properties and thermodynamic behaviour of minerals /NATO ASI Ser. C*, **225**/. Dordrecht: Reidel, 501–590.

Dove, M.T. (1993): *Introduction to lattice dynamics.* Cambridge: Cambridge Univ. Press, 258 p.

Dove, M.T. (1997): Theory of displacive phase transitions in minerals. *Am. Mineral.*, **82**:213–244.

Dovesi, R. (1996): Total energy and related properties. *In* Pisani, C. (ed.): *Quantum-mechanical ab intio calculation of the properties of crystalline materials /Lect. Notes Chem.*, **67**/ Berlin: Springer-Verlag, 179–207.

Dovesi, R., Roetti, C., Freyria-Fava, C., Aprà, E., Saunders, V. & Harrison, N.M. (1992): Ab intio Hartree–Fock treatment of ionic and semi-ionic compounds: state of the art. *Philos. Trans. R. Soc. Lond.*, **A341**:203–210.

Dovesi, R., Saunders, V.R., Roetti, C., Causà, M., Harrison, N.M., Orlando, R. & Aprà, E. (1996): *CRYSTAL95. User's manual.* Torino: Univ. of Torino.

Duffy, T.S., Hemley, R.J. & Mao, H.K. (1995): Equation of state and shear strength at multimegabar pressures: magnesium oxide to 227 GPa. *Phys. Rev. Lett.*, **74**:1371–1374.

Dziewonski, A.M. & Anderson, D.L. (1981): Preliminary reference Earth model. *Phys. Earth Planet. Inter.*, **25**:297–356.

Fang, Z., Solovyev, I.V., Sawada, H. & Terakura, K. (1999): First-principles study on electronic structures and phase stability of MnO and FeO under high pressure. *Phys. Rev., B, Condens. Matter*, **59**:762–774.

Fedorov, F.I. (1968): *Theory of elastic waves in crystals.* New York (N.Y.): Plenum Press, 375 p.

Fiquet, G. (2001): Mineral phases of the Earth's mantle. *Z. Kristallogr.*, **216**:248–271.

Fiquet, G., Dewaele, A., Andrault, D., Kunz, M. & Le Bihan, T. (2000): Thermoelastic properties and crystal structure of MgSiO₃ perovskite at lower mantle pressure and temperature conditions. *Geophys. Res. Lett.*, **27**:21–24.

Foulkes, W.M.C., Mitas, L., Needs, R.J. & Rajagopal, G. (2001): Quantum Monte Carlo simulations of solids. *Rev. Mod. Phys.*, **73**:33–83.

Funamori, N., Yagi, T., Utsumi, W., Kondo, T. & Uchida, T. (1996): Thermoelastic properties of MgSiO₃ perovskite determined by in situ X ray observations up to 30 GPa and 2000 K. *J. Geophys. Res.*, **101**:8257–8269.

Gale, J.D. (1996): Empirical potential derivation for ionic materials. *Philos. Mag.*, **B73**:3–19.

Gale, J.D. (1998): Analytical free energy minimization of silica polymorphs. *J. Phys. Chem., B, Condens. Matter Mater. Surf. Interfaces Biophys.*, **102**:5423–5431.

Garland, C.W. & Weiner, B.B. (1971): Changes in the thermodynamic character of the NH₄Cl order-disorder transition at high pressures. *Phys. Rev., B, Condens. Matter*, **3**:1634–1637.

Gillet, P., Badro, J., Varrel, B. & McMillan, P.F. (1995): High-pressure behaviour in α-AlPO₄: Amorphization and the memory-glass effect. *Phys. Rev., B, Condens. Matter*, **51**:11,262–11,269.

Gillet, P., Matas, J., Guyot, F. & Ricard, Y. (1999): Thermodynamic properties of minerals at high pressures and temperatures from vibrational spectroscopic data. *In* Catlow, C.R.A. & Wright, K. (eds.): *Microscopic properties and processes in minerals /NATO ASI Ser. C*, **543**/. Dordrecht: Kluwer, 71–92.

Gillet, P., Daniel, I., Guyot, F., Matas, J. & Chervin, J.C. (2000): A thermodynamic model for MgSiO₃-perovskite derived from pressure, temperature and volume dependence of the Raman mode frequencies. *Phys. Earth Planet. Inter.*, **117**:361–384.

Gramaccioli, C.M. (2002): Lattice dynamics: Theory and application to minerals. *In* Gramaccioli, C.M. (ed.): Energy modelling in minerals /*EMU Notes Mineral.*, **4**/. Budapest: Eötvös Univ. Press, 245–270.

Hahn, T. (ed.) (1994): *International tables for crystallography. Vol. A, Space-group symmetry.* Fourth, revised edition. Dordrecht: Kluwer.

Hahn, T. & Wondratschek, H. (1994): *Symmetry of crystals. Introduction to International tables for crystallography Vol. A.* Sofia: Heron Press, 134 p.

Haines, J., Leger, J.M. & Schulte, O. (1998): High-pressure isosymmetric phase transition in orthorhombic lead fluoride. *Phys. Rev., B, Condens. Matter*, **57**:7551–7555.

Hama, J. & Suito, K. (1996): The search for a universal equation of state correct up to very high pressures. *J. Phys., Condens. Matter*, **8**:67–81.

Harris, M.J. & Dove, M.T. (1995): Lattice melting at structural phase transitions. *Mod. Phys. Lett., B, Condens. Matter Phys. Stat. Phys. Appl. Phys.*, **9**:67–85.

Harte, B., Harris, J.W., Hutchison, M.T., Watt, G.R. & Wilding, M.C. (1999): Lower mantle mineral associations in diamonds from São Luiz, Brazil. *In* Fei, Y., Bertka, C.M. & Mysen, B.O. (eds.), *Mantle petrology: Field observations and high-pressure experimentation: A Tribute to Francis R. (Joe) Boyd /Geochem. Soc. Spec. Publ.*, **6**/. St. Louis (Mo.): Geochem. Soc., 125–153.

Hatch, D.M. & Merrill, L. (1981): Landau description of the calcite-CaCO₃(II) phase transition. *Phys. Rev., B, Condens. Matter*, **23**:368–374.

Heine, V. & McConnell, J.D.C. (1981): Origin of modulated incommensurate phases in insulators. *Phys. Rev. Lett.*, **46**:1092–1095.

Helffrich, G. (2000): Topography of the transition zone seismic discontinuities. *Rev. Geophys.*, **38**:141–158.

Hemley, R.J. & Cohen, R.E. (1996): Structure and bonding in the deep mantle and core. *Philos. Trans. R. Soc. Lond.*, **A354**:1461–1479.

Hobbs, D., Kresse, G. & Hafner, J. (2000): Fully unconstrained noncollinear magnetism within the PAW method. *Psi-k Newslett.*, **41** (October 2000): 135–146.

Hohenberg, P. & Kohn, W. (1964): Inhomogeneous electron gas. *Phys. Rev.*, **136**:B864–B871.

Holzapfel, W.B. (1996): Physics of solids under strong compression. *Rep. Prog. Phys.*, **59**:29–90.

Holzwarth, N.A.W., Matthews, G.E., Dunning, R.B., Tackett, A.R. & Zeng, Y. (1997): Comparison of the projector augmented-wave, pseudopotential, and linearized augmented-plane-wave formalisms for density-functional calculations of solids. *Phys. Rev., B, Condens. Matter*, **55**:2005–2017.

Hostettler, M., Birkedal, H. & Schwarzenbach, D. (2001): Polymorphs and structures of mercuric iodide. *Chimia*, **55**:541–545.

Iitaka, T. & Ebisuzaki, T. (2001): First-principles calculation of elastic properties of solid argon at high pressures. *Phys. Rev.*, **B65**:art. 012103.

Isaak, D.G., Cohen, R.E., Mehl, M.J. & Singh, D.J. (1993): Phase stability of wüstite at high pressure from first-principles linearized augmented plane-wave calculations. *Phys. Rev., B, Condens. Matter*, **47**:7720–7731.

Ito, E. & Katsura, T. (1989): A temperature profile of the mantle transition zone. *Geophys. Res. Lett.*, **16**:425–428.

Jackson, I. (1998): Elasticity, composition and temperature of the Earth's lower mantle: a reappraisal. *Geophys. J. Int.*, **134**:291–311.

Jaeger, G. (1998): The Ehrenfest classification of phase transitions: introduction and evolution. *Arch. Hist. Exact Sci.*, **53**:51–81.

Jaffe, J.E., Snyder, J.A., Lin, Z. & Hess, A.C. (2000): LDA and GGA calculations for high-pressure phase transitions in ZnO and MgO. *Phys. Rev., B, Condens. Matter*, **62**:1660–1665.

Janot, C. (1994): *Quasicrystals: A primer.* Oxford: Oxford Univ. Press, 409 p.

Jones, R.O. & Gunnarsson, O. (1989): The density functional formalism, its applications and prospects. *Rev. Mod. Phys.*, **61**:689–746.

Kantorovich, L.N. (1995): Thermoelastic properties of perfect crystals with nonprimitive lattices. 1. General theory. *Phys. Rev., B, Condens. Matter*, **51**:3520–3534.

Kantorovich, L.N. (1996–2001): *User-friendly visualisation program for plane-wave ab initio DFT codes CASTEP/CETEP/VASP*. Unpublished, see http://www.cmmp.ucl.ac.uk /~lev/codes/lev00

Karato, S-I. (1993): Importance of anelasticity in the interpretation of seismic tomography. *Geophys. Res. Lett.*, **20**:1623–1626.

Karki, B.B. (1997): *High-pressure structure and elasticity of the major silicate and oxide minerals of the Earth's lower mantle.* Ph.D. Thesis, Univ. of Edinburgh, 170 p.

Karki, B.B., Stixrude, L., Clark, S.J., Warren, M.C., Ackland, G.J. & Crain, J. (1997): Structure and elasticity of MgO at high pressure. *Am. Mineral.*, **82**:51–60.

Karki, B.B., Wentzcovitch, R.M., de Gironcoli, S. & Baroni, S. (1999): First-principles determination of elastic anisotropy and wave velocities of MgO at lower mantle conditions. *Science*, **286**:1705–1707.

Karki, B.B., Wentzcovitch, R.M., de Gironcoli, S. & Baroni, S. (2000a): High-pressure lattice dynamics and thermoelasticity of MgO. *Phys. Rev., B, Condens. Matter*, **61**:8793–8800.

Karki, B.B., Wentzcovitch, R.M., de Gironcoli, S. & Baroni, S. (2000b): Ab initio lattice dynamics of $MgSiO_3$ perovskite at high pressure. *Phys. Rev., B, Condens. Matter*, **62**:14,750–14,756.

Karki, B.B., Wentzcovitch, R.M., de Gironcoli, S. & Baroni, S. (2001): First principles termoelasticity of $MgSiO_3$-perovskite: consequences for the inferred properties of the lower mantle. *Geophys. Res. Lett.*, **28**:2699–2702.

Kennett, B.L.N., Engdahl, E.R. & Buland, R. (1995): Constraints on seismic velocities in the Earth from traveltimes. *Geophys. J. Int.*, **122**:108–124.

Kennett, B.L.N., Widiyantoro, S. & van der Hilst, R.D. (1998): Joint seismic tomography for bulk sound and shear wave speed in the Earth's mantle. *J. Geophys. Res.*, **103**:12,469–12,493.

Kerrick, D.M. (1990) (ed.): *The Al_2SiO_5 polymorphs /Rev. Mineral.*, **22**/. Washington (D.C.): Mineral. Soc. Am., 406 p.

Kesson, S.E., Fitz Gerald, J.D. & Shelley, J.M.G. (1994): Mineral chemistry and density of subducted basaltic crust at lower-mantle pressures. *Nature*, **372**:767–769.

Kesson, S.E., Fitz Gerald, J.D. & Shelley, J.M.G. (1998): Mineralogy and dynamics of a pyrolite mantle. *Nature*, **393**:252–255.

Kieffer, S.W. (1979a): Thermodynamics and lattice vibrations of minerals: 1. Mineral heat capacities and their relationship to simple lattice vibrational models. *Rev. Geophys. Space Phys.*, **17**:1–19.

Kieffer, S.W. (1979b): Thermodynamics and lattice vibrations of minerals: 2. Vibrational characteristics of silicates. *Rev. Geophys. Space Phys.*, **17**:20–34.

Kieffer, S.W. (1979c): Thermodynamics and lattice vibrations of minerals: 3. Lattice dynamics and an approximation for minerals with application to simple substances and framework silicates. *Rev. Geophys. Space Phys.*, **17**:35–59.

Kieffer, S.W. (1980): Thermodynamics and lattice vibrations of minerals: 4. Application to chain and sheet silicates and orthosilicates. *Rev. Geophys. Space Phys.*, **18**:862–886.

Kieffer, S.W. (1982): Thermodynamics and lattice vibrations of minerals: 5. Application to phase equilibria, isotopic fractionation, and high-pressure thermodynamic properties. *Rev. Geophys. Space Phys.*, **20**:827–849.

Knittle, E. & Jeanloz, R. (1986): Thermal expansion of silicate perovskite and stratification of the Earth's mantle. *Nature*, **319**:214–216.

Knittle, E. & Jeanloz, R. (1991): Earth's core-mantle boundary: results of experiments at high pressures and temperatures. *Science*, **251**:1438–1443.

Kohn, W. (1999a): Nobel Lecture: Electronic structure of matter – wave functions and density functionals. *Rev. Mod. Phys.*, **71**:1253–1266.

Kohn, W. (1999b): An essay on condensed matter physics in the twentieth century. *Rev. Mod. Phys.*, **71**:S59–S77.

Kohn, W. & Sham, L.J. (1965): Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, **140**:A1133–A1138.

Kohn, W., Meir, Y. & Makarov, D.E. (1998): Van der Waals energies in density functional theory. *Phys. Rev. Lett.*, **80**:4153–4156.

Komada, N. & Westrum, E.F. (1997): Modeling lattice heat-capacity contributions by a single-parametric phonon dispersion approach. *J. Chem. Thermodyn.*, **29**:311–336.

Kresse, G. & Furthmüller, J. (1996): Efficiency of ab initio total-energy caclulations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.*, **6**:15–50.

Kresse, G. & Joubert, D. (1999): From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev., B, Condens. Matter*, **59**:1758–1775.

Kruger, M.B. & Jeanloz, R. (1990): Memory glass: an amorphous material formed from $AlPO_4$. *Science*, **249**:647–649.

Kurth, S., Perdew, J.P. & Blaha, P. (1999): Molecular and solid-state tests of density functional approximations: LSD, GGAs, and meta-GGAs. *Int. J. Quant. Chem.*, **75**:889–909.

Landau, L.D. & Lifshitz, E.M. (1976): *Statistical physics.* Part I. /*Course of Theoretical Physics*, **5**/. Third edition. Moscow: Nauka, 584 p. (In Russian)

Landau, L.D. & Lifshitz, E.M. (1980, 2001): *Mechanics.* /*Course of Theoretical Physics*, **1**/. Fourth edition. Moscow: Nauka, 217 p. (In Russian)

Le Stunff, Y., Wicks, C.W., Jr. & Romanowicz, B. (1995): P'P' precursors under Africa: evidence for mid-mantle reflectors. *Science*, **270**:74–77.

Lee, I-H. & Martin, R.M. (1997): Applications of the generalised-gradient approximation to atoms, clusters, and solids. *Phys. Rev., B, Condens. Matter*, **56**:7197–7205.

Lichanot, A. (2000): Hartree–Fock and density functional calculations of the elastic constants of the alkaline-earth oxides: comparison with experiment. *Solid State Commun.*, **116**:543–546.

Liu, L.-G. & Bassett, W.A. (1986): Elements, oxides, and silicates. High-pressure phases with implications for the Earth's interior. New York (N.Y.): Oxford Univ. Press, 250 p.

Loubeyre, P., LeToullec, R., Hausermann, D., Hanfland, M., Hemley, R.J., Mao, H.K. &, Finger, L.W. (1996): X-ray diffraction and equation of state of hydrogen at megabar pressures. *Nature*, **383**:702–704.

Louie, S.G., Froyen, S. & Cohen, M.L. (1982): Nonlinear ionic pseudopotentials in spin-density functional calculations. *Phys. Rev., B, Condens. Matter*, **26**:1738–1742.

Mao, H.K., Hemley, R.J., Fei, Y., Shu, J.F., Chen, L.C., Jephcoat, A.P. & Wu, Y. (1991): Effect of pressure, temperature, and composition on lattice parameters and density of $(Mg,Fe)SiO_3$-perovskites to 30 GPa, *J. Geophys. Res.*, **96**:8069–8079.

Masters, G., Laske, G., Bolton, H. & Dziewonski, A. (2000): The relative behaviour of shear velocity, bulk sound velocity, bulk sound speed, and compressional velocity in the mantle: implications for chemical and thermal structure. *In* Karato, S.I. *et al.* (eds.) *Earth's deep interior: Mineral physics and tomography from the atomic to the global scale* /*AGU Geophys. Monogr.*, **117**/. Washington (D.C.): Am. Geophys. Union, 63–87.

Matsui, M. (1989): Molecular dynamics study of the structural and thermodynamic properties of MgO crystal with quantum correction. *J. Chem. Phys.*, **91**:489–494.

Matsui, M. & Price, G.D. (1991): Simulation of the pre-melting behaviour of MgSiO₃ perovskite at high pressures and temperatures. *Nature*, **351**:735–737.

McNeil, L.E. & Grimsditch, M. (1991): Pressure-amorphized SiO₂-quartz: An anisotropic amorphous solid. *Phys. Rev. Lett.*, **68**:83–85.

Mehl, M.J., Cohen, R.E. & Krakauer, H. (1988): Linearized augmented plane wave electronic structure calculations for MgO and CaO. *J. Geophys. Res.*, **93**:8009–8022.

Mendelssohn, M.J. & Price, G.D. (1997): Computer modelling of a pressure induced phase change in clinoenstatite pyroxenes. *Phys. Chem. Miner.*, **25**:55–62.

Milman, V., Winkler, B., White, J.A., Pickard, C.J., Payne, M.C., Akhmatskaya, E.V. & Nobes, R.H. (2000): Electronic structure, properties, and phase stability of inorganic crystals: a pseudopotential plane wave study. *Int. J. Quant. Chem.*, **77**:895–910.

Mishima, O., Calvert, L.D. & Whalley, E. (1984): Melting of ice I at 77 K and 10 kbar: a new method for making amorphous solids. *Nature*, **310**:393–394.

Mishima, O., Calvert, L.D. & Whalley, E. (1985): An apparently first-order transition between two amorphous phases of ice induced by pressure. *Nature*, **314**:76–78.

Monkhorst, H.J. & Pack, J.D. (1976): Special points for Brillouin-zone integrations. *Phys. Rev., B, Condens. Matter*, **13**:5188–5192.

Nada, R., Catlow, C.R.A., Dovesi, R. & Saunders, V. (1992): An ab initio Hartree–Fock study of the ilmenite-structured MgSiO₃ *Proc. R. Soc. Lond.*, **A436**:499–509.

Nekovee, M., Foulkes, W.M.C. & Needs, R.J. (2001): Quantum Monte Carlo analysis of exchange and correlation in the strongly inhomogeneous electron gas. *Phys. Rev. Lett.*, **87**: paper 036401.

Nosé, S. (1984): A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.*, **52**:255–268.

Nye, J.F. (1998): *Physical properties of crystals. Their representation by tensors and matrices.* Oxford: Oxford Univ. Press, 329 p.

O'Keeffe, M. & Bovin, J.O. (1979): Solid electrolyte behavior of NaMgF₃: geophysical implications. *Science*, **206**:599–600.

Oganov, A.R. (2002): *Computer simulation studies of minerals.* Ph.D. Thesis, Univ. of London, 290 p.

Oganov, A.R. & Brodholt, J.P. (2000): High-pressure phases in the Al₂SiO₅ system and the problem of Al-phase in Earth's lower mantle: ab initio pseudopotential calculations. *Phys. Chem. Miner.*, **27**:430–439.

Oganov, A.R. & Dorogokupets, P.I.: All-electron and pseudopotential study of MgO: equation of state, anharmonicity, stability. *Phys. Rev., B, Condens. Matter*, submitted.

Oganov, A.R. & Price, G.D.: Towards an ab initio thermal model of the Earth. (in prep.)

Oganov, A.R., Brodholt, J.P. & Price, G.D. (2000): Comparative study of quasiharmonic lattice dynamics, molecular dynamics and Debye model in application to MgSiO₃ perovskite. *Phys. Earth Planet. Inter.*, **122**:277–288.

Oganov, A.R., Brodholt, J.P. & Price, G.D. (2001a): Ab initio elasticity and thermal equation of state of MgSiO₃ perovskite, *Earth Planet. Sci. Lett.*, **184**:555–560.

Oganov, A.R., Brodholt, J.P. & Price, G.D. (2001b): The elastic constants of MgSiO₃ perovskite at pressures and temperatures of the Earth's mantle. *Nature*, **411**:934–937.

Oganov, A.R., Organova, N.I. & Urusov, V.S. (2001c): Nature of Al-Si anti-ordering in a two-phase feldspar from Pektusan volcano. *Geochem. Int.*, **39**:1160–1171. (Transl. from *Geokhimiya*, (**12**):1265–1276)

Oganov, A.R., Price, G.D. & Brodholt, J.P. (2001d): Theoretical investigation of metastable Al₂SiO₅ polymorphs. *Acta Crystallogr.*, **A57**:548–557.

Oganov, A.R., Price, G.D. & Brodholt, J.P.: Theory of MgSiO₃ perovskite: quantum-mechanical simulations and geophysical implications. (in prep.)

Olbricht, W., Chatterjee, N.D. & Miller, K. (1994): Bayes estimation – a novel approach to derivation of internally consistent thermodynamic data for minerals, their uncertainties, and correlations. 1.Theory. *Phys. Chem. Miner.*, **21**:36–49.

Ortiz, G. & Ballone, P. (1994): Correlation energy, structure factor, radial distribution function, and momentum distribution of the spin-polarised uniform electron gas. *Phys. Rev., B, Condens. Matter*, **50**:1391–1405. (erratum: *ibid.* (1997), **56**:970.)

Parker, S.C. & Price, G.D. (1989): Computer modelling of phase transitions in minerals. *Adv. Solid State Chem.*, **1**:295–327.

Parker, S.C. & Wall, A. (1991): Calculations of the radial seismic velocity/density ratio for MgO and MgSiO$_3$ perovskite at high pressure. *Geophys. Res. Lett.*, **18**:2185–2188.

Parr, R.G. & Yang, W. (1989): *Density-functional theory of atoms and molecules.* Oxford: Oxford Univ. Press, 333 p.

Parrinello, M. & Rahman, A. (1981): Polymorphic transitions in single crystals: A new molecular dynamics method. *J. App. Phys.*, **52**:7182–7190.

Parrinello, M. & Rahman, A. (1982): Strain fluctuations and elastic constants. *J. Chem. Phys.*, **76**:2662–2666.

Pasternak, M.P., Taylor, R.D., Jeanloz, R., Li, X., Nguyen, J.H. & McCammon, C.A. (1997): High pressure collapse of magnetism in Fe$_{0.94}$O: Mössbauer spectroscopy beyond 100 GPa. *Phys. Rev. Lett.*, **79**:5046–5049.

Payne, M.C., Teter, M.P., Allan, D.C., Arias, T.A. & Joannopoulos, J.D. (1992): Iterative minimization techniques for *ab initio* total energy calculations: molecular dynamics and conjugate gradients. *Rev. Mod. Phys.*, **64**:1045–1097.

Perdew, J.P. & Burke, K. (1996): Comparison shopping for a gradient-corrected density functional. *Int. J. Quant. Chem.*, **57**:309–319.

Perdew, J.P. & Kurth, S. (1998): Density functionals for non-relativistic Coulomb systems. *In* Joubert, D.P. (ed.) *Density functionals: Theory and applications /Lect. Notes Phys.*, **500**/. Berlin: Springer-Verlag, 8–59.

Perdew, J.P. & Wang, Y. (1992): Accurate and simple analytic representation of the electron-gas correlation energy. *Phys. Rev., B, Condens. Matter*, **45**:13,244–13,249.

Perdew, J.P. & Zunger, A. (1981): Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev., B, Condens. Matter*, **23**:5048–5079.

Perdew, J.P., Burke, K. & Ernzerhof, M. (1996): Generalized gradient approximation made simple. *Phys. Rev. Lett.*, **77**:3865–3868.

Perdew, J.P., Kurth, S., Zupan, A. & Blaha, P. (1999): Accurate density functionals with correct formal properties: a step beyond the generalized gradient approximation. *Phys. Rev. Lett.*, **82**:2544–2547.

Pisani, C. (1996): Ab initio approaches to the quantum-mechanical treatment of periodic systems. *In* Pisani, C. (ed.): *Quantum-mechanical ab intio calculation of the properties of crystalline materials /Lect. Notes Chem.*, **67**/ Berlin: Springer-Verlag, 47–75.

Poirier, J.-P. (2000): *Introduction to the physics of the Earth's interior.* Second edition. Cambridge: Cambridge Univ. Press, 326 p.

Poirier, J.-P. & Tarantola, A. (1998): A logarithmic equation of state. *Phys. Earth Planet. Inter.*, **109**:1–8.

Price, G.D. (1983): Polytypism and the factors determining the stability of spinelloid structures. *Phys. Chem. Miner.*, **10**:77–83.

Price, G.D. & Yeomans, J. (1984): The application of the ANNNI model to polytypic behavior. *Acta Crystallogr.*, **B40**:448–454.

Price, G.D., Parker, S.C. & Yeomans, J. (1985): The energetics of polytypic structures – a computer simulation of magnesium silicate spinelloids. *Acta Crystallogr.*, **B41**:231–239.

Price, G.D., Parker, S.C. & Leslie, M. (1987): The lattice dynamics and thermodynamics of the Mg$_2$SiO$_4$ polymorphs. *Phys. Chem. Miner.*, **15**:181–190.

Pyykko, P. (1988): Relativistic effects in structural chemistry. *Chem. Rev.*, **88**:563–594.

Rao, C.N.R. & Rao, K.J. (1978): *Phase transitions in solids: an approach to the study of the chemistry and physics of solids*. New York (N.Y.): McGraw-Hill, 330 p.

Remler, D.K. & Madden, P.A. (1990): Molecular dynamics without effective potentials via the Car-Parrinello approach. *Mol. Phys.*, **70**:921–966.

Richardson, M.F., Yang, Q.-C., Novotny-Bregger, E. & Dunitz, J.D. (1990): Conformational polymorphism of dymethyl 3,6-dichloro-2,5-dihydroxyterephthalate. II. Structural, thermodynamic, kinetic and mechanistic aspects of phase transformations among the three crystal forms. *Acta Crystallogr.*, **B46**:653–660.

Richet, P. & Gillet, P. (1997): Pressure-induced amorphisation of minerals: a review. *Eur. J. Mineral.*, **9**:907–933.

Ringwood, A.E. (1991): Phase transformations and their bearing on the constitution and dynamics of the mantle. *Geochim. Cosmochim. Acta*, **55**:2083–2110.

Robertson, G.S. & Woodhouse, J.H. (1996): Constraints on lower mantle properties from seismology and mineral physics. *Earth Planet. Sci. Lett.*, **143**:197–205.

Robie, R.A. & Edwards, J.L. (1966): Some Debye temperatures from single crystal elastic constant data. *J. Appl. Phys.*, **37**:2659–2663.

Ross, N.L. & Hazen, R.M. (1989): Single crystal X-ray diffraction study of $MgSiO_3$ perovskite from 77 to 400 K. *Phys. Chem. Miner.*, **16**:415–420.

Sandratskii, L.M. (1998): Noncollinear magnetism in itinerant-electron systems: theory and applications. *Adv. Phys.*, **47**:91–160.

Schmidt, M.W., Poli, S., Comodi, P. & Zanazzi, P.F. (1997): High-pressure behavior of kyanite: Decomposition of kyanite into stishovite and corundum. *Am. Mineral.*, **82**:460–466.

Sharma, S.M. & Sikka, S.K. (1996): Pressure-induced amorphization of materials. *Prog. Mater. Sci.*, **40**:1–77.

Shearer, P. (1990): Seismic imaging of upper-mantle structure with new evidence for a 520-km discontinuity. *Nature*, **344**:121–126.

Shechtman, D., Blech, I., Gratias, D. & Cahn, J.W. (1984): Metallic phase with long-range orientational order and no translational symmetry. *Phys. Rev. Lett.*, **53**:1951–1953.

Sherman, D.M. (1991): The high-pressure electronic structure of magnesiowüstite (Mg,Fe)O – applications to the physics and chemistry of the lower mantle. *J. Geophys. Res.*, **96**:14,299–14,312.

Sherman, D.M. (1992). Equation of state and high-pressure phase transitions of stishovite (SiO2): ab initio (periodic Hartree-Fock) results. *J. Geopys. Res.*, **98**:11,865-11,873.

Singh, D.J. (1994): *Planewaves, pseudopotentials and the LAPW method.* Boston: Kluwer, 115 p.

Sirotin Yu.I. & Shaskolskaya, M.P. (1975): *Fundamentals of crystal physics.* Moscow: Nauka, 680 p. (In Russian)

Sowa, H. (2000): A transition path from the zinc-blende to the NaCl-type. *Z. Kristallogr.*, **215**:335–342.

Speziale, S., Zha, C.-S., Duffy, T.S., Hemley, R.J. & Mao, H.-K. (2001): Quasi-hydrostatic Compression of magnesium oxide to 52 GPa: Implications of the pressure-volume-temperature equation of state. *J. Geophys. Res.*, **B106**: 515–528.

Stixrude, L. & Cohen, R.E. (1993): Stability of orthorhombic $MgSiO_3$ perovskite in the Earth's lower mantle. *Nature*, **364**:613–616.

Stixrude, L., Hemley, R.J., Fei, Y. & Mao, H.K. (1992): Thermoelasticity of silicate perovskite and magnesiowüstite and stratification of the Earth's mantle. *Science*, **257**:1099–1101.

Stixrude, L., Cohen, R.E., Yu, R.C. & Krakauer, H. (1996): Prediction of phase transition in $CaSiO_3$ perovskite and implications for lower mantle structure. *Am. Mineral.*, **81**:1293–1296.

Stixrude, L., Cohen, R.E. & Hemley, R.J. (1998): Theory of minerals at high pressure. In Hemley, R.J. (ed.): Ultrahigh-pressure mineralogy /*Rev. Mineral.*, **37**/. Washington (D.C.): Mineral. Soc. Am., 639–671.

Stowasser, R. & Hoffmann, R. (1999): What do the Kohn–Sham orbitals and eigenvalues mean? *J. Am. Chem. Soc.*, **121**:3414–3420.

Sugino, O. & Car, R. (1995): Ab initio molecular dynamics study of first-order phase transitions: melting of silicon. *Phys. Rev. Lett.*, **74**:1823–1826.

Sutton, A.P. (1993): *Electronic structure of materials.* Oxford: Oxford Univ. Press, 260 p.

Terhune, R.W., Kushida, T. & Ford, G.W. (1985): Soft acoustic modes in trigonal crystals. *Phys. Rev., B, Condens. Matter*, **32**:8416–8419.

Thijssen, J.M. (1999): *Computational physics.* Cambridge: Cambridge Univ. Press, 546 p.

Tsirelson, V.G. (1986): *Electron crystal chemistry* /*Adv. Sci. Technol., Ser. Cryst. Chem.*, **20**/. Moscow: Russian Institute of Scientific and Technical Information, 261 p. (In Russian)

Tsirelson, V.G. (1993): *Chemical bonding and thermal motion of atoms in crystals.* /*Adv. Sci. Technol., Ser. Cryst. Chem.*, **27**/. Moscow: Russian Institute of Scientific and Technical Information, 270 p. (In Russian)

Urusov, V.S. & Dubrovinsky, L.S. (1989): *Computer modelling of structure and properties of minerals.* Moscow: Moscow State Univ. Press, 200 p. (In Russian)

van Smaalen, S. (1995): Incommensurate crystal structures. *Crystallogr. Rev.*, **4**:79–202.

Vanderbilt, D. (1990): Soft self-consistent pseudopotentials in a generalized eigenvalue problem. *Phys. Rev., B, Condens. Matter*, **41**:7892–7895.

Vaughan, M.T. & Weidner, D.J. (1978): The relationship of elasticity and crystal structure of andalusite and sillimanite. *Phys. Chem. Miner.*, **3**:133–144.

Venkataraman, G., Feldkamp, L.A. & Sahni, V.C. (1975): *Dynamics of perfect crystals.* Cambridge (Mass.): MIT Press, 517 p.

Verhoogen, J. (1980): *Energetics of the Earth.* Washington (D.C.): Natl. Acad. Press, 139 p.

Vinet, P., Ferrante, J., Smith, J.R. & Rose, J.H. (1986): A universal equation of state for solids. *J. Phys. C. Solid State Phys.*, **19**:L467–L473.

Vinet, P., Rose, J.H., Ferrante, J. & Smith, J.R. (1989): Universal features of the equation of state of solids. *J. Phys., Condens. Matter*, **1**:1941–1963.

Vinnik, L., Niu, F. & Kawakatsu, H. (1998): Broadband converted phases from midmantle discontinuities. *Earth Planets Space*, **50**:987–997.

Vočadlo, L., Brodholt, J., Alfé, D., Gillan, M.J. & Price, G.D. (2000): Ab initio free energy calculations on the polymorphs of iron at core conditions. *Phys. Earth Planet. Inter.*, **117**:123–137.

Vosko, S.H., Wilk, L. & Nusair, M. (1980): Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Can J. Phys.*, **58**:1200–1211.

Wallace, D.C. (1998): *Thermodynamics of crystals.* New York (N.Y.): Dover Publ., 484 p.

Wang, Y. & Perdew, J.P. (1991): Correlation hole of the spin-polarized electron gas, with exact small-vector and high-density scaling. *Phys. Rev., B, Condens. Matter*, **44**:13,298–13,307.

Wang, Y., Yip, S., Phillpot, S. & Wolf, D. (1993): Crystal instabilities at finite strain. *Phys. Rev. Lett.*, **71**:4182–4185.

Wang, Y., Weidner, D.J., Liebermann, R.C. & Zhao, Y. (1994): $P–V–T$ equation of state of state of (Mg,Fe)SiO_3 perovskite determined by in situ X-ray observations up to 30 GPa and 2000 K. *J. Geophys. Res.*, **101**:8257–8269.

Wang, Y., Li, J., Yip, S., Phillpot, S. & Wolf, D. (1995): Mechanical instabilities of homogeneous crystals. *Phys. Rev., B, Condens. Matter*, **52**:12,627–12,635.

Warren, M.C., Ackland, G.J., Karki, B.B. & Clark, S.J. (1998): Phase transitions in silicate perovskites from first principles. *Mineral. Mag.*, **62**:585–598.

Welch, D.O., Dienes, G.J. & Paskin, A. (1978): A molecular dynamical study of the equation of state of solids at high temperature and pressure. *J. Phys. Chem. Solids*, **39**:589–603.

Wentzcovitch, R.M., Martins, J.L. & Price, G.D. (1993): Ab initio molecular dynamics with variable cell shape: application to $MgSiO_3$. *Phys. Rev. Lett.*, **70**:3947–3950.

White, J.C. & Hess, A.C. (1993): Periodic Hartree–Fock study of siliceous mordenite. *J. Phys. Chem.*, **97**:6398–6404.

Williams, Q. & Garnero, E. (1996): Seismic endence for partial melt at the base of Earth's mantle. *Science*, **273**:1528–1530.

Wilson, K.G. (1983): The renormalization group and related phenomena. *Rev. Mod. Phys.*, **55**:583–600.

Winkler, B. & Dove, M.T. (1992): Thermodynamic properties of $MgSiO_3$ perovskite derived from large-scale molecular dynamics. *Phys. Chem. Miner.*, **18**:407–415.

Xu, Y., Shankland, T. & Poe, B.T. (2000): Laboratory-based electrical conductivity in the Earth's mantle. *J. Geophys. Res.*, **105**:27,865–27,875.

Yeomans, J.M. (1992): *Statistical mechanics of phase transitions.* Oxford: Oxford Univ. Press, 168 p.

Zerr, A., Diegler, A. & Boehler, R. (1998): Solidus of Earth's deep mantle. *Science*, **281**:243–246.

Zhang, L., Ahsbahs, H., Kutoglu, A. & Hafner, S.S. (1992): Compressibility of grunerite. *Am. Mineral.*, **77**:480–483.

Zharkov, V.N. & Kalinin, V.A. (1968): *Equations of state of solids at high pressures and temperatures.* Moscow: Nauka, 312 p. (In Russian)

Zupan, A., Blaha, P., Schwarz, K. & Perdew, J.P. (1998): Pressure-induced phase transitions in solid Si, SiO_2, and Fe: Performance of local-spin-density and generalized-gradient-approximation density functionals. *Phys. Rev., B, Condens. Matter*, **58**:11,266–11,272.