Calculation of the vacancy formation energy in aluminium

Calculation of the vacancy formation energy in aluminium

M J Gillan

Physics Department, University of Keele, Keele, Staffordshire ST5 5BG, UK and
Theoretical Physics Division, Harwell Laboratory, Oxfordshire OX11 0RA, UK

Abstract. This paper addresses the problem of calculating the vacancy formation energy in simple metals and presents results for aluminium. The calculations are based on the pseudopotential method, the local-density approximation for exchange and correlation, and periodically repeating geometry. The approach used is similar to that proposed by Car and Parrinello, and allows the simultaneous relaxation of the electrons and the ionic positions. Problems peculiar to metals in this approach are discussed and a way of overcoming them is presented. The calculated vacancy energy (0.56 eV) in aluminium is in quite good agreement with experiment (0.66 eV). Comparisons with perturbation theory show that the calculated value is subject to a technical error of only ~0.03 eV and that corrections due to periodic boundary conditions are also of this order. The contribution to the vacancy energy from non-linear effects is similar to the jellium estimate of Evans and Finnis.

1. Introduction

The last few years have seen great advances in the ability to calculate the structure and energy of solids from first principles (for reviews, see e.g. Heine 1984, Phariseau and Temmerman 1984, Cohen 1986, Srivastava and Weaire 1987, Ihm 1988). These advances have been made possible by the systematic use of density-functional theory, by improvements in pseudopotential technique and by the steady increase in the power of computers. It has been shown that basic properties such as the lattice parameter, the bulk modulus and the stable crystal structure can be calculated with great precision by solving Schrödinger's equation for the electrons using well defined approximations. An impressive feature of the new methods is that they can be applied to a wide range of different materials: they have been successful for metals, semiconductors and ionic materials alike. The new methods are now being extended to treat problems in material physics such as surface reconstruction (Needels *et al* 1987), grain-boundary structure (Payne *et al* 1987) and point-defect energetics (Louie *et al* 1976, Baraff and Schlüter 1979, Baraff *et al* 1983, Car *et al* 1984, Scheffler 1987) in semiconductors. The time seems ripe for a re-examination of the problem of calculating point-defect energies in metals, and it is the purpose of this paper to describe a new approach to this problem. We have chosen to focus on the calculation of the vacancy formation energy in aluminium, both because of the technological importance of this metal and because it is a case where previous calculations are in serious disagreement with experiment.

There has been a large amount of previous theoretical work on defect energies in the simple metals (those having a free-electron-like band structure). In the early days of pseudopotential theory, it seemed possible that the properties of these metals might be well accounted for by perturbation theory (Harrison 1966, Heine 1970, Heine and Weaire 1970). The thought was that the electron–ion pseudopotential might be weak enough to cause only a small disturbance in the otherwise uniform gas of valence electrons. This disturbance, and the associated electronic relaxation energy, could then be calculated by linear response theory, using a realistic description for the dielectric function of the uniform electron gas. It was realised that linear response theory was equivalent to a representation of the total energy as a volume term plus an inter-ionic pair potential (Harrison 1966, Heine and Weaire 1970). It has turned out, though, that this method is satisfactory only in a few special cases. In polyvalent metals like aluminium, the vacancy energy cannot be calculated in this way, because the pseudopotential is too strong. The essential difficulties were highlighted by Evans and Finnis (1976), by means of simple calculations on the formation energy of a vacancy in jellium. They showed that for an electron density corresponding to aluminium, non-linear effects would be expected to raise the formation energy by about 1 eV. Given that the formation energy in aluminium is 0.66 eV (Triftshäuser 1975, Berger *et al* 1978, Fluss *et al* 1978), the conclusion was that linear response theory is completely inadequate. Previous linear response calculations which had appeared to give good agreement with experiment (e.g. Popovic *et al* 1974) were therefore untrustworthy. This conclusion was amply reinforced in the work of Jacucci *et al* (1981), which showed that with a carefully constructed pseudopotential and an accurate form for the dielectric function, linear response theory gave a vacancy energy in aluminium that was close to zero. Subsequently, attempts have been made to go to higher order in perturbation theory (e.g. So and Woo 1981), but the results have been inconclusive. Part of the problem here has been that one has no idea of the convergence properties of the perturbation series.

It seems clear that the only way of resolving these problems is through a full *ab initio* calculation that does not depend on perturbation theory. Even now, this is a substantial undertaking, but we shall show in this paper that the task is well within the range of modern techniques. There have been two previous attempts to perform such a calculation for the vacancy in aluminium (Chakraborty *et al* 1981, Chakraborty and Siegel 1982), neither of which gave satisfactory results. This suggests that there are difficult technical problems in this kind of calculation. One of the main purposes of this paper is to explore these problems in some detail and to show how they may be overcome.

Apart from the question of achieving electronic self-consistency, which does not present a serious difficulty, one of the major problems in metals comes from Fermi-surface effects. Another difficulty is that in a complete treatment one needs to account for the lattice distortion surrounding the defect. Recently, Car and Parrinello (1985) have described general techniques which allow the simultaneous relaxation of the electrons to the self-consistent ground state and the relaxation of the ions to their equilibrium positions. Although we shall not follow their technique in any detail, the methods we shall develop lean heavily on their general ideas.

An important theme that we shall develop in this paper is that the perturbation approach, in spite of its inadequacies, can provide powerful support to a full self-consistent calculation. The idea here is that the approximations adopted in the full calculation must be such as to yield agreement with perturbation theory in the case of a weak pseudopotential. This requirement provides a very important control on the technical adequacy of the full calculation. At the same time, by comparing the full results

with perturbation theory, we can directly examine the contributions omitted by the latter. We shall be able to demonstrate that for the aluminium vacancy the energy contribution lacking in linear response theory has about the size proposed by Evans and Finnis (1976).

The plan of the paper is as follows. Section 2 recalls briefly the main theoretical ideas concerning density-functional theory, supercell geometry, the vacancy formation energy and the perturbation expansion. In § 3, we describe the computational techniques used to calculate the total energy and to perform the energy minimisation. We then (§ 4) present our numerical results for the energy of the perfect lattice, and the energy of formation of the ideal and relaxed vacancy. Our conclusions are outlined in § 5. Note that throughout this paper we work in atomic units, with energies in hartrees.

2. Theoretical framework

We summarise here the general theoretical ideas on which the work is based.

2.1. *The total ground-state energy*

According to density-functional theory (Hohenberg and Kohn 1964, Kohn and Sham 1965, Levy 1982, Callaway and March 1984), the total ground-state energy E of the system of electrons and ionic cores can be expressed as a functional of the electronic density distribution $\rho(r)$:

$$E[\rho] = E_K[\rho] + E_c[\rho] + E_H[\rho] + E_{xc}[\rho] + E_i. \tag{1}$$

Here, E_K is the kinetic energy of the non-interacting electron gas whose density is $\rho(r)$, E_c is the electron–core interaction energy, E_H is the Hartree energy, E_{xc} is the exchange–correlation energy and E_i is the Coulomb interaction energy of the cores. We treat explicitly only the valence electrons, whose interaction with the ionic cores is represented by a pseudopotential $v(r)$. The distribution $\rho(r)$ is then the (pseudo-)density of the valence electrons. The electron–core energy is

$$E_c = \int dr \, V(r)\rho(r) \tag{2}$$

where $V(r)$ is the sum of the ionic pseudopotentials

$$V(r) = \sum_i v(|r - R_i|) \tag{3}$$

with R_i the core positions. In general, the ionic pseudopotential v is non-local, but in this paper we shall work with a local form. We take the usual local density approximation (LDA) for the exchange–correlation energy

$$E_{xc} \simeq \int dr \, \rho(r)\varepsilon_{xc}(\rho(r)) \tag{4}$$

where $\varepsilon_{xc}(\rho)$ is the exchange–correlation energy per electron in a uniform electron gas of density ρ.

The ground-state energy is the minimum value of $E[\rho]$ as a function of $\rho(r)$, with the total number of electrons and the 'external' potential $V(r)$ held fixed. The associated Euler equation is equivalent to the Euler equation for non-interacting electrons in an

effective single-particle potential $V_{eff}(r)$ given by

$$V_{eff}(r) = V(r) + \int dr' \, \rho(r')/|r - r'| + \mu_{xc}(\rho(r)) \tag{5}$$

where μ_{xc} is the exchange–correlation chemical potential

$$\mu_{xc}(\rho) = \frac{d}{d\rho}[\rho \varepsilon_{xc}(\rho)]. \tag{6}$$

Because of this equivalence, one can regard E as a functional of the wavefunctions $\psi_i(r)$ of the occupied orbitals:

$$E = 2 \sum_{i \in occ} \int dr \, \psi_i^*(r)(-\tfrac{1}{2}\nabla^2)\psi_i(r) + E_c[\rho] + E_H[\rho] + E_{xc}[\rho] + E_i \tag{7}$$

with $\rho(r)$ given by

$$\rho(r) = 2 \sum_{i \in occ} |\psi_i(r)|^2 \tag{8}$$

the sums being restricted to occupied orbitals. The Euler equation for the $\psi_i(r)$ must require that the orbitals be orthonormal. It has the form

$$[-\tfrac{1}{2}\nabla^2 + V_{eff}(r)]\psi_i(r) = \sum_{j \in occ} \Lambda_{ij} \psi_j(r) \tag{9}$$

for the occupied orbitals i. The matrix of undetermined multipliers Λ_{ij} arising from the orthonormality constraint has eigenvalues ε_i which are the effective single-particle energies of Kohn and Sham. In a practical calculation, we have to determine the $\psi_i(r)$ which give a self-consistent solution of equations (5), (8) and (9).

For reasons that will become clear later, it will be convenient to work with the formal generalisation of this scheme to non-zero temperature (Mermin 1965, Callaway and March 1984). This can be done by working with a free energy functional A which depends on both the wavefunctions and the single-particle occupation numbers, which we denote by f_i. We write

$$A = E - TS \tag{10}$$

where E is the natural finite-temperature generalisation of equation (7):

$$E = 2 \sum_i f_i \int dr \, \psi_i^*(r)(-\tfrac{1}{2}\nabla^2)\psi_i(r) + E_c[\rho] + E_H[\rho] + E_{xc}[\rho] + E_i \tag{11}$$

and the entropy S is given by

$$S = -2k_B \sum_i [f_i \ln f_i + (1 - f_i) \ln(1 - f_i)]. \tag{12}$$

The density is now given by

$$\rho(r) = 2 \sum_i f_i |\psi_i(r)|^2. \tag{13}$$

In principle, the sums now go over a complete set of orthonormal single-particle states.

When we minimise A with respect to the $\psi_i(r)$, with the constraint of orthonormality, the Euler equation is

$$f_i[-\tfrac{1}{2}\nabla^2 + V_{eff}(r)]\psi_i(r) = \sum_j \Lambda_{ij} \psi_j(r) \tag{14}$$

where $V_{eff}(r)$ is, as before, given by equation (5). On minimising with respect to the f_i

with the total number of electrons fixed, we find

$$\int \mathrm{d}\boldsymbol{r}\,\psi_i^*(\boldsymbol{r})[-\tfrac{1}{2}\nabla^2 + V_{\mathrm{eff}}(\boldsymbol{r})]\psi_i(\boldsymbol{r}) + k_\mathrm{B}T[\ln f_i - \ln(1 - f_i)] = \mu \qquad (15)$$

where μ is the chemical potential. If we define

$$\eta_i = \int \mathrm{d}\boldsymbol{r}\,\psi_i^*(\boldsymbol{r})[-\tfrac{1}{2}\nabla^2 + V_{\mathrm{eff}}(\boldsymbol{r})]\psi_i(\boldsymbol{r}) \qquad (16)$$

then equation (15) gives

$$\eta_i - \mu = -k_\mathrm{B}T[\ln f_i - \ln(1 - f_i)] \qquad (17)$$

which is equivalent to

$$f_i = 1/\{\exp[(\eta_i - \mu)/k_\mathrm{B}T] + 1\}. \qquad (18)$$

Equations (14) are satisfied when the $\psi_i(\boldsymbol{r})$ are eigenfunctions of the single-particle Hamiltonian, so that

$$\Lambda_{ij} = \delta_{ij} f_i \varepsilon_i. \qquad (19)$$

Then $\eta_i = \varepsilon_i$, so that f_i are the Fermi occupation numbers for thermal equilibrium at temperature T. Clearly, as $T \to 0$, the entropy goes to zero, and we recover the usual zero-temperature scheme. There is one point to notice though, which will be important later. In the $T \to 0$ limit, the equations are satisfied just as well if, instead of taking the occupied wavefunctions to be eigenfunctions of the single-particle Hamiltonian, we take a unitary transformation of these eigenfunctions. But for $T \neq 0$, the ψ_i must be eigenfunctions, because of the presence of the f_i in equation (14).

When we apply this scheme, we shall of course take the temperature to be small compared with other relevant energies.

2.2. Supercell geometry

We have chosen to perform our defect calculations using supercell (periodically repeating) geometry: rather than attempting to treat a single isolated defect in an infinite crystal, we treat a periodically repeating array of defects. This has the advantage of allowing us to use rather standard momentum-space techniques (e.g. Ihm *et al* 1979) in calculating the total energy. To get useful results, we must of course take our repeating cell big enough to make the interaction between the defects negligible, and we shall pay particular attention to this problem.

In supercell geometry, the ionic positions are periodically repeated, so that if there is an ion at position $\boldsymbol{R}$ then there are periodic images at $\boldsymbol{R} + \boldsymbol{t}$ for all translation vectors $\boldsymbol{t}$ of the superlattice. The electron–ion potential $V(\boldsymbol{r})$ has the symmetry of the superlattice:

$$V(\boldsymbol{r} + \boldsymbol{t}) = V(\boldsymbol{r}). \qquad (20)$$

The density $\rho(\boldsymbol{r})$ and hence the effective potential $V_{\mathrm{eff}}(\boldsymbol{r})$ have the same symmetry. Then by Bloch's theorem the wavefunctions ψ_i must satisfy

$$\psi_i(\boldsymbol{r} + \boldsymbol{t}) = \mathrm{e}^{\mathrm{i}\boldsymbol{q}\cdot\boldsymbol{t}}\psi_i(\boldsymbol{r}) \qquad (21)$$

where $\boldsymbol{q}$ is some wavevector, which can be taken to be in the first Brillouin zone of the superlattice. The wavefunctions can thus be classified according to the wavevector $\boldsymbol{q}$,

and we shall write $\psi_{qi}(r)$ for the ith wavefunction at wavevector q. We shall use a plane-wave basis set and expand:

$$\psi_{qi}(r) = (1/\Omega^{1/2}) \sum_G a_{qiG} \, e^{i(G+q)\cdot r} \tag{22}$$

where Ω is the volume of the supercell and G are the reciprocal lattice vectors of the superlattice. The density $\rho(r)$ in principle involves an integral of the $|\psi_{qi}(r)|^2$ over the Brillouin zone, but in practice this is replaced by a sum:

$$\rho(r) = 2 \sum_q w_q \sum_i f_{qi} |\psi_{qi}(r)|^2 \tag{23}$$

where q goes over a chosen set of points in the Brillouin zone, the w_q are a suitable set of weights (see § 3.8), and f_{qi} is the occupation number of the ith orbital at wavevector q. The self-consistent solution of the single-particle equations (5), (13) and (14) then consists of the determination of the amplitudes a_{qiG} for G's up to some cut-off.

2.3. Vacancy formation energy

We shall focus on the energy of formation of a vacancy at constant volume; this is the same as the formation energy at constant pressure, provided the volume is such that the pressure is zero (Chang and Falicov 1971). We start from a perfect lattice consisting of N atoms on N lattice sites, remove one of the atoms to form the vacancy, and replace it on a new bulk lattice site; finally we compress the system back to its original volume. Denoting by $E(N, \nu; \Omega)$ the energy of the system of N atoms and ν vacancies occupying $N + \nu$ lattice sites in volume Ω, the energy of formation E_f is

$$E_f = E(N, 1; \Omega) - E(N, 0; \Omega). \tag{24}$$

We are ultimately interested in the case where $E(N, 1; \Omega)$ refers to the fully relaxed positions of the ions around the vacancy, but we shall also discuss the formation energy E_f^0 of the 'ideal' unrelaxed vacancy. In supercell geometry, the energies $E(N, 1; \Omega)$ and $E(N, 0; \Omega)$ cannot both be calculated directly, since systems of $N + 1$ and N lattice sites cannot both satisfy periodic boundary conditions. This is not a problem, though, because the perfect lattice energy is just N times the energy per atom, so that E_f can be written as

$$E_f = E(N - 1, 1;(N - 1)\Omega/N) - [(N - 1)/N]E(N, 0; \Omega). \tag{25}$$

We can thus get what we want from calculations on two periodically repeating systems having the same number N of lattice sites, one being the defective system of $N - 1$ atoms and a vacancy, the other being the perfect lattice of N atoms.

When we analyse our numerical results, it will be helpful to separate the energies in equation (25) into various contributions. Consider first the energy of the perfect lattice of N atoms in volume Ω. This can be constructed starting from the uniform electron gas of NZ electrons in volume Ω (Z is the core charge). If $\varepsilon(\rho)$ is the total energy per electron in the uniform gas of density ρ, then the energy of this initial system is $NZ\varepsilon(NZ/\Omega)$. Now plant into this the N ionic cores on their regular lattice positions, but holding the electron gas uniform. The contribution to the energy from this step is the Madelung energy of the cores $E_{Mad}(N, 0; \Omega)$, plus N times the interaction energy of each core with the uniform electron gas of density ρ, which we denote by $e_0(\rho)$. Finally, we allow the electron gas to relax to the self-consistent ground state; we call this relaxation energy

$E_{rel}(N, 0; \Omega)$. Thus

$$E(N, 0; \Omega) = NZ\varepsilon(NZ/\Omega) + Ne_0(NZ/\Omega) + E_{Mad}(N, 0; \Omega) + E_{rel}(N, 0; \Omega). \quad (26)$$

We can go through the same argument for the system of $N - 1$ atoms and one vacancy. The energy of the uniform electron gas for this system is $(N - 1)Z\varepsilon(NZ/\Omega)$. We can write the energy of the ideal vacancy system as

$$E^0(N - 1, 1; (N - 1)\Omega/N) = (N - 1)Z\varepsilon(NZ/\Omega) + (N - 1)e_0(NZ/\Omega)$$

$$+ E_{Mad}^0(N - 1, 1; (N - 1)\Omega/N) + E_{rel}^0(N - 1, 1; (N - 1)\Omega/N) \quad (27)$$

where E_{Mad}^0 and E_{rel}^0 are the Madelung and electronic relaxation energies for the unrelaxed ionic positions. When we take the difference of energies in equation (25), the contributions from the uniform electron gas cancel, and we get

$$E_f^0 = E_{Mad}^0(N - 1, 1; (N - 1)\Omega/N) - [(N - 1)/N]E_{Mad}(N, 0; \Omega)$$

$$+ E_{ref}^0(N - 1, 1; (N - 1)\Omega/N) - [(N - 1)/N]E_{rel}(N, 0; \Omega). \quad (28)$$

The Madelung energies are, of course, trivial to calculate. We can therefore focus on the electronic relaxation energies for the perfect and defective systems. A similar formula can be written for the ionically relaxed formation energy E_f.

It will be noted that the present calculations yield the formation energy at zero temperature (for zero pressure, this is equal to the enthalpy of formation). The quantity measured experimentally is the enthalpy of formation at a relatively high temperature. The two quantities are, of course, not identical, as has been stressed by, e.g., Catlow *et al* (1981), who discussed the thermodynamic relations needed in a calculation of high-temperature defect enthalpies. The work of Jacucci *et al* (1981) indicates that in aluminium the enthalpy of formation increases by perhaps as much as 0.1 eV between 0 and 860 K. This is not insignificant, but is not large enough to affect the arguments of the present paper.

2.4. Perturbation theory

A number of previous calculations on defects in simple metals have been based on perturbation theory, in which the energies are expanded up to some order (usually second) in the strength of the electron–core pseudopotential. The main calculations to be presented later do not, of course, rely on perturbation theory, but it will be extremely helpful to be able to compare them with perturbation results, as we shall see. It is relatively straightforward to perform perturbation calculations up to third order in the pseudopotential, and we summarise here the relations we shall need.

We expand the electronic relaxation energy E_{rel} per repeating cell in powers of the periodic potential $V(r)$ of equation (3). The expansion is most conveniently written in terms of the Fourier components, which we define by

$$\hat{V}(G) = (1/\Omega) \int dr\, e^{-iG \cdot r} V(r). \quad (29)$$

To third order in V, we write

$$E_{rel} = \tfrac{1}{2}\Omega \sum_G{}' \chi_2(G)|\hat{V}(G)|^2$$

$$+ \tfrac{1}{3}\Omega \sum_{G_1 G_2 G_3}{}' \delta_{G_1 + G_2 + G_3, 0} \chi_3(G_1, G_2, G_3)\hat{V}(G_1)\hat{V}(G_2)\hat{V}(G_3) + \ldots \quad (30)$$

where the prime on the summation indicates the exclusion of zero wavevectors. The quantity $\chi_2(G)$ is the familiar linear response function describing the density disturbance caused by a weak external potential. Since our main calculations are based on the LDA, χ_2 and χ_3 must also be treated in this way. The LDA for χ_2 is well known (Taylor 1978), and takes the form

$$\chi_2(G) = \chi_2^0(G)/[1 - \chi_2^0(G)(4\pi/G^2 + \mu_{xc}')] \tag{31}$$

where $\chi_2^0(G)$ is the response function for non-interacting electrons and μ_{xc}' means $\mathrm{d}\mu_{xc}/\mathrm{d}\rho$. A general discussion of χ_3 has been given by Lloyd and Sholl (1968) and other authors (Hammerberg and Ashcroft 1974), though we have been unable to find an explicit LDA treatment. The LDA form for χ_3 is, however, straightforward to derive, and we find

$$\chi_3(G_1, G_2, G_3) = \xi_2(G_1)\xi_2(G_2)\xi_2(G_3)\chi_3^0(G_1, G_2, G_3) + \tfrac{1}{2}\mu_{xc}''\chi_2(G_1)\chi_2(G_2)\chi_2(G_3) \tag{32}$$

where χ_3^0 is χ_3 for non-interacting electrons, $\xi_2(G)$ is defined as

$$\xi_2(G) = 1/[1 - \chi_2^0(G)(4\pi/G^2 + \mu_{xc}')] \tag{33}$$

and μ_{xc}'' means $\mathrm{d}^2\mu_{xc}/\mathrm{d}\rho^2$. An expression for $\chi_3^0(G_1, G_2, G_3)$ in closed form is implicit in the paper of Lloyd and Sholl (1968).

3. Techniques

3.1. The total energy

We need to calculate numerically the total free energy of equations (10), (11) and (12) for given amplitudes a_{qiG}, occupation numbers f_{qi} and ionic positions R_i. As usual, we impose a wavevector cut-off: the amplitude variables are the a_{qiG} for the set of G such that $\tfrac{1}{2}|G + q|^2 < E_{cut}$ for some chosen energy E_{cut}. Our scheme for calculating the various terms in the energy follows quite closely that of Car and Parrinello (1985), and we now summarise this briefly.

3.1.1. Kinetic energy. This can be evaluated directly as

$$E_K = \sum_q w_q \sum_i f_{qi} \sum_G |G + q|^2 |a_{qiG}|^2. \tag{34}$$

3.1.2. Electron–core energy. This is conveniently evaluated in terms of the Fourier transform $\hat{\rho}_G$ of the density:

$$\hat{\rho}_G = \int \mathrm{d}r\, \mathrm{e}^{-iG\cdot r} \rho(r). \tag{35}$$

From equation (2)

$$E_c = \sum_G \hat{V}_G \hat{\rho}_{-G} \tag{36}$$

where (equation (3))

$$\hat{V}_G = \Omega^{-1} \hat{v}_G S(G) \tag{37}$$

with $S(G)$ the structure factor

$$S(G) = \sum_i e^{-iG \cdot R_i}$$ (38)

and $\hat{v}_G$ the Fourier transform of the pseudopotential

$$\hat{v}_G = \int dr\, e^{-iG \cdot r} v(r).$$ (39)

The zero-wavevector term in equation (36) needs careful treatment, since $\hat{v}_G$ diverges here; this has been discused by Ihm *et al* (1979).

The direct evaluation of $\hat{\rho}_G$ in wavevector space

$$\hat{\rho}_G = \sum_q w_q \sum_i f_{qi} \sum_{G'} a^*_{qiG'} a_{qiG'+G}$$ (40)

would be extremely time-consuming, since the number of operations would be proportional to the square of the number of Gs. Instead, we calculate the density $\rho_l = \rho(r_l)$ on a grid of points r_l in real space and construct $\hat{\rho}_G$ by Fourier transformation:

$$\hat{\rho}_G = \omega_1 \sum_l e^{-G \cdot r_l} \rho_l$$ (41)

where ω_1 is the volume per grid point. The ρ_l are obtained from equation (13), after Fourier transformation of the amplitudes. The usual fast Fourier technique reduces the number of operations to $\nu \ln \nu$, where ν is the number of Gs. Note that equation (40) gives non-zero $\hat{\rho}_G$ for $|G|$ up to approximately twice the cut-off $|G|$ used for the amplitudes. The real-space grid must be taken fine enough to account for this. If this is done, then the Fourier-transform method exactly reproduces equation (40).

3.1.3. Hartree energy. This is likewise evaluated in terms of $\hat{\rho}_G$:

$$E_H = (2\pi/\Omega) \sum_{G \neq 0} G^{-2} |\hat{\rho}_G|^2$$ (42)

the zero-wavevector term being excluded (Ihm *et al* 1979).

3.1.4. Exchange–correlation energy. This is evaluated on the real-space grid:

$$E_{xc} = \omega_1 \sum_l \rho_l \varepsilon_{xc}(\rho_l).$$ (43)

3.1.5. Madelung energy. The Coulomb interaction energy E_i of the ions is calculated using the standard Ewald technique.

3.1.6. Ground-state energy. Our calculation will yield the minimum free energy A for a specified temperature T. The use of this finite-temperature scheme is merely a device, whose main purpose is to smooth discontinuities at the Fermi level (see § 3.4 below). Our real interest is in the ground-state energy E_0. It is easily shown that for small T the free energy deviates from E_0 by a quantity quadratic in T: $A = E_0 - \frac{1}{2}\gamma T^2$; and that the deviation of the energy E is equal and opposite: $E = E_0 + \frac{1}{2}\gamma T^2$. Our best estimate for the ground-state energy will therefore be $\frac{1}{2}(E + A)$, whose deviation from E_0 will only be $O(T^3)$. Our numerical results for the energy presented in § 4 are calculated in this way.

3.2. Energy minimisation

Our task is to determine numerically the minimum (free) energy of the perfect and defective systems, and in the latter case to minimise this energy with respect to the ionic positions. One way of doing this involves the diagonalisation of the single-particle Hamiltonian using an estimate for the density $\rho(r)$; the resulting occupied orbitals are used to construct a new estimate for the density and the process is iterated to self-consistency. If the ionic positions are to be relaxed, this whole operation must be repeated in a search for the minimum total energy. It has been stressed by Car and Parrinello (1985) that this complicated nested set of processes can be condensed into a single iterative search in which the total energy E is minimised simultaneously with respect to the occupied wavefunctions and the ionic coordinates. In the scheme they propose, both the wavefunctions and the coordinates are treated as classical dynamical variables. The minimisation is performed by a type of damped dynamics. Their scheme has been used with great success in the study of grain-boundary and surface structure in germanium (Payne *et al* 1987, Needels *et al* 1987) and in the calculation of the relaxed bulk structure of α-cristobalite (Allan and Teter 1987).

The approach we shall describe here is in the same general spirit as the Car–Parrinello method, but differs from it in important respects. The guiding idea is that we search for the overall minimum of the free energy as a function of the plane-wave amplitudes a_{qiG} of the occupied orbitals, the occupation numbers f_{qi} and the ionic coordinates R_i. At this minimum, the single-particle equations (14) are satisfied and the forces on the ions all vanish. Instead of using damped dynamics, though, we perform the search by the conjugate-gradient technique, which we expect to be considerably more efficient. A number of important problems arise in the development of this approach for a metal, as we shall discuss later in this section. For the moment, we focus on minimisation with respect to the amplitudes; the occupation numbers and the ionic positions will be considered later.

First, we recall the conjugate-gradient idea (Fletcher and Reeves 1964, Fletcher 1980). Suppose we have a space of vectors x and a real function $\varphi(x)$ whose minimum we seek iteratively. At the kth step, we find ourselves at $x^{(k)}$ and we look for the next estimate $x^{(k+1)}$ along the search direction $s^{(k)}$:

$$x^{(k+1)} = x^{(k)} + \alpha^{(k)}s^{(k)} \tag{44}$$

where $\alpha^{(k)}$ is chosen so as to minimise φ along this direction. The conjugate-gradient prescription for $s^{(k)}$ is a generalisation of the steepest-descent method. In the latter, $s^{(k)}$ would point down the gradient of φ:

$$s^{(k)} = f(x^{(k)}) \tag{45}$$

where the 'force' f is

$$f = -\nabla\varphi. \tag{46}$$

In the conjugate-gradient method, information about the forces in previous steps is used to pick a more intelligent search direction. The algorithm is

$$s^{(k)} = f^{(k)} + \beta^{(k)}s^{(k-1)} \tag{47}$$

with $\beta^{(k)}$ given by

$$\beta^{(k)} = |f^{(k)}|^2/|f^{(k-1)}|^2. \tag{48}$$

The process is initiated by taking $s^{(0)} = f^{(0)}$. The reasons for the efficacy of this strategy have been discussed by, for example, Fletcher (1980).

In our problem, the plane-wave amplitudes a_{qiG} for each occupied orbital (qi) are the components of a vector in a (complex) vector space of dimension ν_{qi}, the number of Gs. We are looking for the minimum energy as a function of the set of all occupied orbitals, so our vector space is the sum of the vector spaces for the orbitals; the overall dimension is thus $\Sigma \nu_{qi}$. The force corresponding to f (equation (46)) has components

$$F_{qiG} = -\partial E/\partial a^*_{qiG} \tag{49}$$

where for the purpose of this derivative a_{qiG} and a^*_{qiG} are treated as independent variables. The calculation of the F_{qiG} from the formulae of § 3.1 is straightforward.

If we denote the amplitudes at the kth iteration by $a^{(k)}_{qiG}$ and the associated forces by $F^{(k)}_{qiG}$, then the iterative scheme is

$$a^{(k+1)}_{qiG} = a^{(k)}_{qiG} + \alpha^{(k)} S^{(k)}_{qiG} \tag{50}$$

where the search direction is

$$S^{(k)}_{qiG} = F^{(k)}_{qiG} + \beta^{(k)} S^{(k-1)}_{qiG} \tag{51}$$

with

$$\beta^{(k)} = \sum_{qiG} |F^{(k)}_{qiG}|^2 \Big/ \sum_{qiG} |F^{(k-1)}_{qiG}|^2. \tag{52}$$

3.3. Orthonormality

The scheme we have just outlined is of course incomplete in one essential respect: we have not accounted for the orthonormality of the single-particle orbitals. For each wavevector q, all the eigenfunctions $\psi_{qi}(r)$ must be orthonormal:

$$\int_\Omega \mathrm{d}r\, \psi^*_{qi}(r)\psi_{qj}(r) = \delta_{ij} \tag{53}$$

where the integral goes over a single supercell. With the amplitudes defined as in equation (22), this requires

$$\sum_G a^*_{qiG} a_{qjG} = \delta_{ij} \tag{54}$$

for every q. The search process should not therefore be free to range over the entire vector space, but should be constrained to the manifold of a_{qiG} satisfying equation (54). The same condition applies in the Car–Parrinello method, and various ways have been used to enforce it (Car and Parrinello 1985, Payne *et al* 1988). The procedure we have adopted consists of two steps. First, instead of taking the search direction given by equation (51), we take a modified search direction such that if the $a^{(k)}_{qiG}$ satisfy the orthonormality condition then so do the $a^{(k+1)}_{qiG}$ to linear order in $\alpha^{(k)}$. Secondly, once $\alpha^{(k)}$ has been chosen, we apply a correction to the $a^{(k+1)}_{qiG}$ so that they exactly satisfy the condition.

Consider first the search direction. Suppose we have a set of amplitudes a_{qiG} which are exactly orthonormal. Now consider vectors t_{qiG} such that for A_{qiG} defined by

$$A_{qiG} = a_{qiG} + \alpha t_{qiG} \tag{55}$$

we have

$$\sum_G A^*_{qiG} A_{qjG} = \delta_{ij} + \mathrm{O}(\alpha^2). \tag{56}$$

The set of all such t_{qiG} forms a subspace which we call the tangent subspace (since the manifold of A_{qiG} is tangential to the constraint manifold). We now require that the search direction lie in the tangent subspace. The simplest way of ensuring this is to take the projection of $S_{qiG}^{(k)}$ onto this subspace. It is readily shown that this projection $\bar{S}_{qiG}^{(k)}$ is

$$\bar{S}_{qiG}^{(k)} = S_{qiG}^{(k)} - \tfrac{1}{2} \sum_{jG'} [S_{qiG'}^{(k)} a_{qjG'}^{(k)*} + a_{qiG'}^{(k)} S_{qjG'}^{(k)*}] a_{qjG}^{(k)}. \tag{57}$$

If we now use $\bar{S}_{qiG}^{(k)}$ instead of $S_{qiG}^{(k)}$ in equation (50), then we ensure that the $a_{qiG}^{(k+1)}$ satisfy equation (54) to order $(\alpha^{(k)})^2$. But we want them to satisfy it *exactly*. In order to ensure this, we apply a correction $\delta a_{qiG}^{(k+1)}$:

$$a_{qiG}^{(k+1)} = a_{qiG}^{(k)} + \alpha^{(k)} \bar{S}_{qiG}^{(k)} + \delta a_{qiG}^{(k+1)}. \tag{58}$$

Of course, there is considerable freedom in the choice of the $\delta a_{qiG}^{(k+1)}$. To fix them uniquely, we require that the correction be as small as possible in the sense that for each q,

$$\sum_{iG} |\delta a_{qiG}^{(k+1)}|^2$$

has the smallest possible value. This specifies the correction uniquely. It is straight-forward to show that if we write

$$A_{qiG}^{(k+1)} = a_{qiG}^{(k)} + \alpha^{(k)} \bar{S}_{qiG}^{(k)} \tag{59}$$

then

$$a_{qiG}^{(k+1)} = \sum_{j} M_{qij} A_{qjG}^{(k+1)} \tag{60}$$

where $\mathbf{M}_q$ is a Hermitian matrix equal to $\mathbf{L}_q^{-1/2}$, the elements of the matrix $\mathbf{L}_q$ being

$$L_{qij} = \sum_{G} A_{qiG}^{(k+1)*} A_{qjG}^{(k+1)}. \tag{61}$$

3.4. Variable occupation numbers

In the ground state, the occupation numbers f_{qi} are equal to 1 for $\varepsilon_{qi} < \mu$ and 0 for $\varepsilon_{qi} > \mu$, where ε_{qi} are the Kohn–Sham eigenvalues at wavevector q and μ is the chemical potential (Fermi energy). In the case of a metal, this discontinuity of f_{qi} as a function of energy is extremely troublesome, for reasons connected with the Brillouin zone sampling (Fu and Ho 1983, Needs *et al* 1986). A particularly useful way of seeing this is in terms of perturbation theory (§ 2.4). If our calculation scheme is to be reliable, it must agree with low-order perturbation theory when the pseudopotential is sufficiently weak. Now if we were to expand the ground-state energy produced by our calculation scheme in powers of the pseudopotential, the associated response functions would implicitly be calculated by Brillouin zone sampling. But in perturbation theory we do not have to rely on Brillouin zone sampling, since we know the response functions analytically. The implication is that the sampling we adopt should yield response functions that agree closely with the analytical values. The sampling approximation to $\chi_2^0(G)$ is

$$\chi_2^0(G) \simeq \frac{2}{\Omega} \sum_{q} w_q \sum_{G'} \frac{f_{q+G'}^0 - f_{q+G+G'}^0}{\varepsilon_{q+G'}^0 - \varepsilon_{q+G+G'}^0} \tag{62}$$

where ε_k^0 are the non-interacting single-particle energies and f_k^0 are the associated occupation numbers. Now the point is that Brillouin zone sampling is effective with a small number of q-vectors only if the function being sampled is smooth in reciprocal space. But because of the occupation numbers, the function sampled in equation (62), far from being smooth, actually becomes discontinuous at zero temperature.

There is another related reason why we get trouble. Since we cannot know the self-consistent eigenvalues in advance, we do not know how many occupied states there will be at each q. As the iteration progresses to self-consistency, eigenvalues at different q will generally cross each other and the Fermi energy, and this would require a discontinuous change of occupation numbers. Such discontinuities would presumably play havoc with the minimisation scheme.

The solution we have adopted to these problems is to allow the f_{qi} to vary continuously in the range $(0, 1)$. This has the effect both of smoothing the sampled function and of eliminating discontinuities due to level crossing. A convenient way of formulating this idea is to consider the calculation formally at finite temperature, and this is the reason for introducing the finite-temperature generalisation of density-functional theory outlined in §2.1.

It is straightforward to rewrite this finite-temperature scheme in terms of the plane-wave amplitudes and including Brillouin zone summation. In the practical minimisation, the amplitudes and the f_{qi} are, of course, varied simultaneously. We have found it inconvenient and unnecesary to treat the f_{qi} by conjugate gradients, and a much simpler method suffices. At a given iteration k, we have values of the amplitudes $a_{qiG}^{(k)}$ and the occupation numbers $f_{qi}^{(k)}$ for the ith orbital at wavevector q. These are used to calculate the total energy, and also the single-particle expectation values η_{qi} defined as in equation (16). These η_{qi} when substituted in equation (18) would give new occupation numbers, which we write as $f_{qi}^{(k)} + \Delta f_{qi}^{(k)}$. The search for the next values of the occupation numbers is made according to

$$f_{qi}^{(k+1)} = f_{qi}^{(k)} + \gamma^{(k)} \Delta f_{qi}^{(k)}. \tag{63}$$

It can be shown that this is guaranteed to reduce the free energy for some positive $\gamma^{(k)}$.

It will be seen that we have to pay a price for this finite-temperature device. We must now work with all orbitals for which f_{qi} is appreciable. However, since we require a good approximation to the zero-temperature situation, we must take $k_B T \ll \mu$, so that the number of additional orbitals is not very great. It should be added that other methods for smoothing the Fermi discontinuity have also been described in the literature (Fu and Ho 1983, Needs *et al* 1986). The relationship between our proposed method and these other methods requires further study.

3.5. Difficulties of ill-conditioning

There are, however, other difficulties to be faced in the finite-temperature scheme. These arise because the equations to be solved are very ill-conditioned. There are two reasons for this. The first is that the free energy depends only weakly on the amplitudes a_{qiG} of orbitals having small occupation numbers. The other has to do with the fact that the true minimum is obtained only for exact eigenfunctions of the single-particle Hamiltonian. If the temperature were exactly zero, the free energy would be invariant under a unitary transformation of the orbitals. At low temperature, it is almost invariant, so that the free energy is insensitive to such transformations.

The first problem is easy to deal with: we have only to work with suitably scaled amplitudes. We find that if we go to variables $a_{qiG}/\sqrt{f_{qi}}$ then this source of trouble disappears.

The second problem is more vicious, and its solution requires a modification in strategy. Since unitary transformations play a central role here, it is natural to separate displacements of the amplitudes into two parts. Consider such a displacement:

$$a_{qiG} \rightarrow a_{qiG} + \delta a_{qiG} \tag{64}$$

for given q. This might represent a unitary transformation

$$a_{qiG} + \delta a_{qiG} = \sum_j U_{ij} a_{qjG} \tag{65}$$

where U_{ij} is unitary. The necessary and sufficient condition for this, given that orthonormality is preserved, is that the δa_{qiG} be linear combinations of the a_{qiG}. Then let us represent a general displacement as

$$\delta a_{qiG} = \delta_{\parallel} a_{qiG} + \delta_{\perp} a_{qiG} \tag{66}$$

where $\delta_{\parallel} a_{qiG}$ is a linear combination of the a_{qiG} and $\delta_{\perp} a_{qiG}$ is orthogonal to all the a_{qiG}:

$$\sum_G a_{qiG}^* \delta_{\perp} a_{qjG} = 0 \tag{67}$$

for all i, j. Our strategy will then be to apply conjugate gradients only to the orthogonal displacements $\delta_{\perp} a_{qiG}$, and to use another procedure, now to be described, for the $\delta_{\parallel} a_{qiG}$; this will be devised so as to ensure that the single-particle Hamiltonian rapidly becomes diagonal in the space of the a_{qiG}.

In order to explain this, suppose that we have some amplitudes and occupation numbers, and hence some potential $V_{\mathrm{eff}}(r)$, and let us define

$$h_{qij} = \int_{\Omega} \mathrm{d}r\, \psi_{qi}^*(r) [-\tfrac{1}{2}\nabla^2 + V_{\mathrm{eff}}(r)] \psi_{qj}(r). \tag{68}$$

Perhaps the obvious way of forcing the orbitals to be eigenfunctions would simply be to diagonalise this matrix. But this would be unsatisfactory. It would not, of course, give us the self-consistent eigenfunctions, and there is no guarantee that it would even reduce the energy. What we prefer to do is to displace the wavefunctions so that they become more nearly eigenfunctions, while ensuring that the energy is reduced. Let us write the parallel displacements as

$$\delta_{\parallel} \psi_{qi}(r) = \alpha \sum_j W_{ij} \psi_{qj}(r). \tag{69}$$

In order for orthonormality to be preserved to linear order in α, W_{ij} must be anti-Hermitian. We also want the W_{ij} to be such that the energy decreases as α increases from zero. The condition can be shown to be

$$\sum_{ij} h_{qij} W_{ij} (f_i - f_j) < 0. \tag{70}$$

Now let us write W_{ij} as

$$W_{ij} = \zeta_{ij} h_{qji}/(h_{qii} - h_{qjj}) = \zeta_{ij} h_{qji}/(\eta_{qi} - \eta_{qj}) \tag{71}$$

where ζ_{ij} is symmetric—this will make W_{ij} anti-Hermitian. The motivation for this is that if we take all $\zeta_{ij} = 1$, then we just get the first-order perturbation formula for the eigenfunctions of h_{qij}. Now consider condition (70). As we go to self-consistency, the

occupation numbers will go to their thermal values, equation (18), so that $f_i > f_j$ if $\eta_i < \eta_j$. In this case, equation (70) is automatically satisfied for $\zeta_{ij} = 1$. On the way to self-consistency, it may happen for some (i, j) that $f_i - f_j$ and $\eta_i - \eta_j$ have the same sign; if this happens, we put $\zeta_{ij} = 0$ for such (i, j), to ensure that condition (70) remains satisfied. We also modify ζ_{ij} if $|h_{qij}|/(\eta_{qi} - \eta_{qj})$ is too large, since perturbation theory is valid only for small values of this ratio. We choose some maximum value (normally equal to 0.2) for $|W_{ij}|$ and reduce ζ_{ij} from unity, if necessary, so that this maximum is not exceeded.

3.6. Relaxation of the ions

The relaxation of the ions to their equilibrium positions is actually one of the lesser problems in the present work, which is why we have left it till last. Car and Parrinello (1985) proposed a 'simulated annealing' procedure for simultaneously relaxing the ions and the electrons, and stressed the virtues of this approach when there are many competing minima. In the present problem (and many others), we should not expect to find more than a single minimum, so that a much simpler method suffices. Initially, we attempted to treat the ions and the wavefunctions on the same footing in a combined conjugate-gradient scheme. This can be made to work, but turns out to be very slow. The method we have adopted makes use of the fact that it is very much easier to relax the ions to their zero-force positions than to relax the electrons to self-consistency.

In the expression for the total energy, the ionic coordinates appear explicitly only in E_c and E_i. For a given $\rho(r)$, the calculation of the forces on the ions is orders of magnitude more rapid than the electronic part of the calculation. For fixed a_{qiG} and f_{qi}, the relaxation of the ions to the positions for which the 'forces' $-\partial E/\partial R_i$ vanish to high accuracy can be accomplished in a very few (usually three or four) iterative steps. Our procedure thus involves the relaxation of the ions to their exact zero-force positions at each iterative step for the amplitudes and occupation numbers. The iterative relaxation of the latter variables is performed exactly as described before, the electronic forces F_{qiG} being evaluated for the exact zero-force positions of the ions.

It will be noted that this method is exactly the opposite of the more obvious method in which the electrons are relaxed to self-consistency before the ionic forces are calculated. This latter method would, of course, take far longer, because it is the iteration to self-consistency which is the time-consuming process. It should also be stressed that the ionic forces $-\partial E/\partial R_i$ are not the Hellmann–Feynman forces that would be found for the ground state with the given R_i. This does not matter, though: all that is required is that our overall procedure should yield the minimum of the free energy with respect to the amplitudes, occupation numbers and positions.

3.7. Pseudopotential and exchange–correlation function

The pseudopotential for aluminium used in our numerical calculations is a local form due to Goodwin (1987). Its Fourier transform (equation (39)) is

$$\hat{v}_k = -(4\pi/k^2)[(Z - AR)\cos(kR) + (A/k)\sin(kR)]\exp[-(k/k_c)^N] \tag{72}$$

the values of the parameters being $Z = 3$, $A = 0.11065$, $R = 1.15$, $k_c = 3.5$, $N = 6$ in atomic units (energies in hartrees). This pseudopotential is constructed so as to reproduce the experimental value for the equilibrium lattice parameter, and it has

been shown to give satisfactory values for the bulk modulus and the phonon frequencies (Goodwin 1987).

Our calculations use the exchange–correlation energy $\varepsilon_{xc}(\rho)$ due to Ceperley and Alder (1980), as parametrised by Perdew and Zunger (1981).

3.8. Brillouin zone sampling

The Brillouin zone sampling is performed using the special-points scheme of Chadi and Cohen (1973). With our finite-temperature device, it is sufficiently accurate to use two special points in the irreducible 1/48th of the Brillouin zone, as we shall show.

4. Numerical results

Our concern here is not only with the absolute numerical results but also with showing that the various technical problems are under adequate control. Apart from the convergence to self-consistency, which poses no particular problem, there are four technical matters that demand attention: (i) the energy cut-off; (ii) the number of sampling vectors; (iii) the 'temperature'; and (iv) the size of the supercell. All four can be studied by suitable comparisons with perturbation results. There are three convenient sizes of supercell, which contain 8, 16 and 27 lattice sites, and correspond respectively to FCC, BCC and FCC periodic geometry. We have made calculations on the perfect lattice, the ideal vacancy and the relaxed vacancy, and we discuss the results for these systems in turn. In all the results to be reported, electronic self-consistency has been achieved to better than 10^{-6} Hartree for the total energy of the system.

4.1. The perfect lattice

Extensive trial calculations have led us to conclude that an energy cut-off of 6.4 Hartree, a temperature of one-tenth the free-electron Fermi energy, and two Chadi–Cohen sampling points give adequate precision, and all results are obtained with these parameters, unless otherwise stated. The quoted energy cut-off is the one we use for the equilibrium value of the lattice parameter a_0; for other values of a_0, the cut-off is scaled so that the number of plane waves for a given supercell is always the same.

Calculations on the 16-site perfect lattice for a series of lattice parameters yield an equilibrium a_0 of 7.598 au. This is very close to the value expected for the pseudopotential we are using (Goodwin 1987); the experimental value is $a_0 = 7.64$ au (Simmons and Wang 1971).

We now compare with perturbation theory by making calculations in which the pseudopotential is scaled by a factor C between 0 and 1. If second-order perturbation theory was exact, the electronic relaxation energy E_{rel} would be proportional to C^2, so it is convenient to study the quantity $E_{rel}/N_{site}C^2$, where N_{site} is the number of sites in the supercell. Results for the three supercells are displayed in figure 1. By fitting to a fifth-degree polynomial, we extract the two leading coefficients in the expansion

$$E_{rel}/N_{site}C^2 = b_0 + b_1C + \ldots \tag{73}$$

corresponding to second- and third-order perturbation theory. The exact values of these coefficients can be calculated from equation (30)—here the wavevector sums
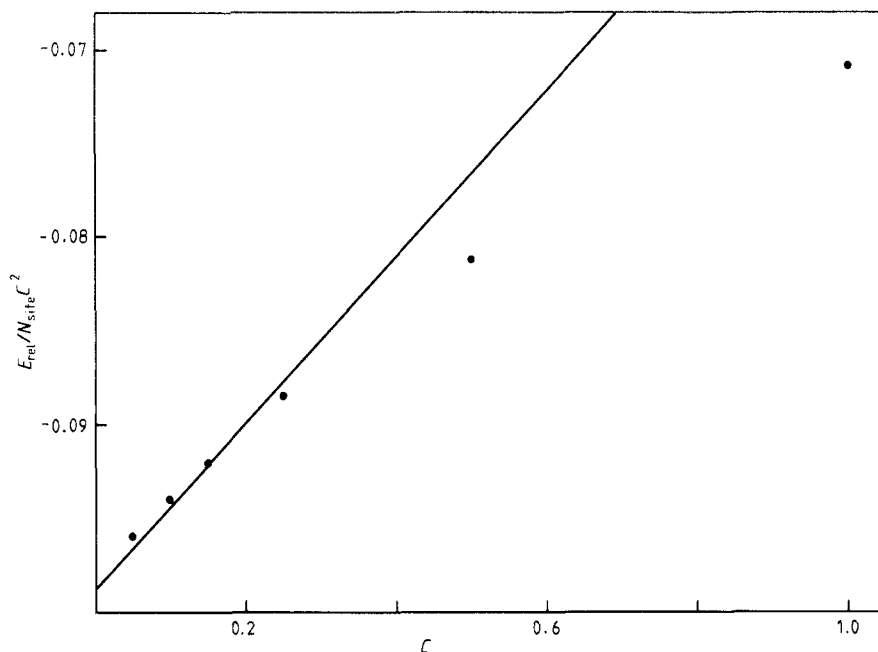
Figure 1. Electronic relaxation energy E_{rel} for the perfect lattice for different values of the scaling factor C of the ionic pseudopotential. Dots show values obtained from the full calculation; results for the three sizes of supercell are indistinguishable on the scale of the plot. The straight line shows the result of perturbation theory up to third order.

Table 1. Expansion coefficients b_0 and b_1 for the electronic relaxation energy of the perfect lattice: comparison of values extracted from the full calculation and exact values from perturbation theory. Values are in hartrees.

Site no.	b_0		b_1	
	Full calc.	Pert.	Full calc.	Pert.
8	−0.09834	−0.09887	0.04385	0.04464
16	−0.09824	−0.09887	0.04395	0.04464
27	−0.09812	−0.09887	0.04371	0.04464

are cut off at a value corresponding to an energy of 8.5 Hartree, which is enough to render the residual error completely negligible. Values of the expansion coefficients are compared in table 1. The very close agreement confirms the adequacy of our chosen parameters. Note that the small differences between the results for different supercells are solely due to the Brillouin zone sampling and give a measure of the error due to sampling.

4.2. The ideal vacancy

We have done calculations on the vacancy system with supercells containing 8, 16 and 27 sites (7, 15 and 26 atoms). The lattice parameter a_0 is taken in each case so that
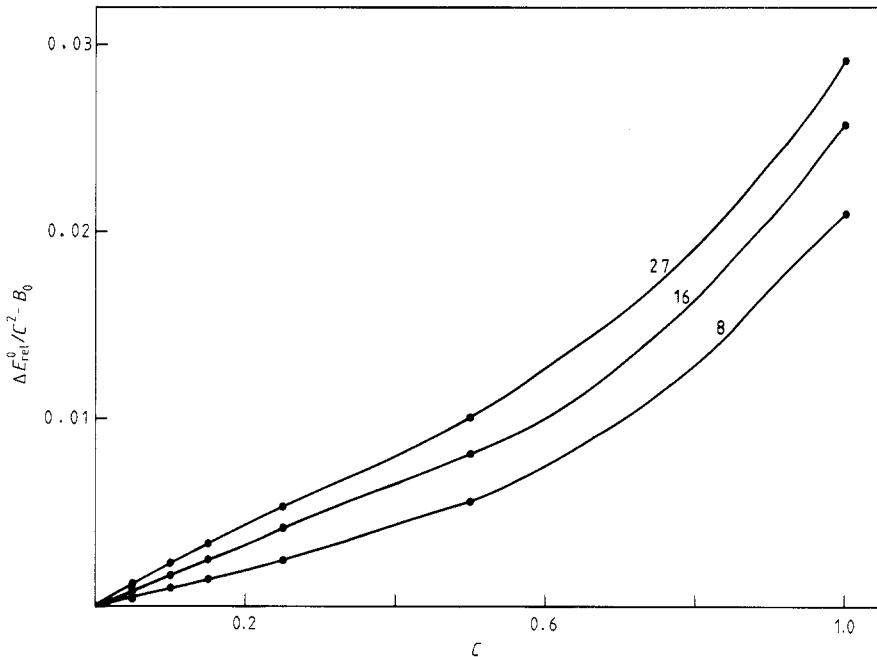
Figure 2. Electronic relaxation contribution ΔE_{rel}^0 to the ideal vacancy formation energy for different values of the scaling factor C of the ionic pseudopotential. The three sets of results are for supercells containing 8, 16 and 27 lattice sites.

the number of atoms per unit volume is the same as in the perfect crystal (§ 2.3), and the energy cut-off is adjusted accordingly.

We first study the agreement with perturbation theory, by performing the calculations with different scale factors C for the pseudopotential. As we have shown in § 2.3, the key quantity for the vacancy formation energy is the difference of electronic relaxation energy for the defective and perfect systems, which is given for the ideal vacancy by

$$\Delta E_{\text{rel}}^0 = E_{\text{rel}}^0(N - 1, 1; (N - 1)\Omega/N) - [(N - 1)/N]E_{\text{rel}}(N, 0; \Omega). \tag{74}$$

In figure 2 we display our results for $\Delta E_{\text{rel}}^0/C^2$ for the three sizes of supercell. By fitting as before to a fifth-degree polynomial, we extract the two leading coefficients in the expansion

$$\Delta E_{\text{rel}}^0/C^2 = B_0 + B_1 C + \ldots \tag{75}$$

corresponding to second- and third-order perturbation theory. The exact values of these coefficients are then calculated by applying equation (30) to the defective system with the same values of lattice parameter as in the full calculation and a cut-off energy of 8.5 Hartree. The values of the coefficients are compared in table 2. The close agreement with the exact results confirms the adequacy of the energy cut-off, the temperature and the number of sampling vectors used in the full calculations.

We now use the results to calculate the energy of formation of the ideal vacancy. The most obvious way to do this is simply to use equation (28) as it stands. The ionic Madelung energies and hence the difference

$$\Delta E_{\text{Mad}}^0 = E_{\text{Mad}}^0(N - 1, 1; (N - 1)\Omega/N) - [(N - 1)/N]E_{\text{Mad}}(N, 0; \Omega) \tag{76}$$

Table 2. Expansion coefficients B_0 and B_1 for the electronic relaxation energy of vacancy formation: comparison of values extracted from the full calculation and exact values from perturbation theory. Values in hartrees.

Site no.	B_0		B_1	
	Full calc.	Pert.	Full calc.	Pert.
8	−0.5478	−0.5494	0.0090	0.0071
16	−0.7873	−0.7896	0.0166	0.0198
27	−0.9446	−0.9433	0.0220	0.0231

Table 3. Electronic relaxation and Madelung terms, ΔE_{rel}^0 and ΔE_{Mad}^0, in the formation energy E_f^0 of the ideal vacancy calculated for three sizes of repeating cell. Energies in hartrees (eV in parentheses).

Site no.	ΔE_{rel}^0	ΔE_{Mad}^0	E_f^0
8	−0.5268	0.5527	0.0259 (0.71)
16	−0.7616	0.7852	0.0236 (0.64)
27	−0.9154	0.9366	0.0212 (0.58)

Table 4. The part of the electronic relaxation energy for vacancy formation $\delta\Delta E_{rel}^0$ not accounted for by second- and third-order perturbation theory, and the resulting improved estimate for the vacancy formation energy. Energies in hartrees (eV in parentheses).

Site no.	$\delta\Delta E_{rel}^0$	E_f^0
8	0.0120	0.0224 (0.61)
16	0.0091	0.0246 (0.67)
27	0.0072	0.0236 (0.64)

are straightforward to calculate to high precision using the usual Ewald procedure. Our results for ΔE_{rel}^0, ΔE_{Mad}^0 and E_f^0 for the three supercells are listed in table 3.

The precision of these results can be improved by a very simple device. We have shown that the expansion coefficients B_0 and B_1 are in close agreement with their exact values. Nevertheless, there are errors, which are not negligible on the scale of the formation energy. We can use our knowledge of the exact values to eliminate this part of the error. Let us denote by $\delta\Delta E_{rel}^0$ the part of ΔE_{rel}^0 not accounted for by second- and third-order perturbation theory:

$$\delta\Delta E_{rel}^0 = \Delta E_{rel}^0 - B_0 - B_1. \tag{77}$$

Taking ΔE_{rel}^0, B_0 and B_1 all from our full calculations, we obtain the values of $\delta\Delta E_{rel}^0$ given in table 4. Since we know that these calculations give B_0 and B_1 correct to about 10^{-3} Hartree, we expect that the value of $\delta\Delta E_{rel}^0$ will also be correct to this kind of accuracy. We now calculate ΔE_{rel}^0 by writing

$$\Delta E_{rel}^0 = B_0^{exact} + B_1^{exact} + \delta\Delta E_{rel}^0 \tag{78}$$

where B_0^{exact} and B_1^{exact} are the exact values calculated from the perturbation

Table 5. Expansion coefficients B_0^{exact} and B_1^{exact} and Madelung term ΔE_{Mad}^0 for the vacancy formation energy $(E_f^0)^{\text{pert}}$ calculated by third-order perturbation theory for a range of cell sizes; supercell symmetry is noted in parentheses in the first column.

Site no.	B_0^{exact}	B_1^{exact}	ΔE_{Mad}^0	$(E_f^0)^{\text{pert}}$
8 (FCC)	-0.5494	0.0071	0.5527	0.0104
16 (BCC)	-0.7896	0.0198	0.7852	0.0155
27 (FCC)	-0.9433	0.0231	0.9366	0.0164
64 (FCC)	-1.1518	0.0238	1.1434	0.0154
125 (FCC)	-1.2805	0.0250	1.2716	0.0161
128 (BCC)	-1.2846	0.0244	1.2756	0.0155
216 (FCC)	-1.3678	0.0253	1.3585	0.0160

expression. This procedure should give a result for ΔE_{rel}^0 and hence E_f^0 to a precision of better than 10^{-3} Hartree (0.03 eV). Results for E_f^0 obtained by this procedure are given in table 4. It must be stressed that this precision concerns only the effects of energy cut-off, temperature and Brillouin zone sampling. Since both the full calculations and the perturbation calculations are performed in supercell geometry, there will remain errors due to the interaction between the vacancies. However, the perturbation scheme also allows us to assess this source of error, as we now show.

4.3. Supercell effects

The full calculations become very time-consuming for supercells containing more than 27 sites. But it is clear from the results we have just presented that the major part of the electronic relaxation energy is accounted for by second- and third-order perturbation theory. The remainder $\delta \Delta E_{\text{rel}}^0$ is still important on the scale of the vacancy energy, but is considerably smaller than the other terms. In particular, its variation with supercell size should be smaller than that of those other terms. This means that we can use perturbation theory to gauge the supercell corrections. The perturbation calculations are, of course, far more rapid than the full calculations, so that we can go to much bigger supercells. We give in table 5 the values of B_0^{exact} and B_1^{exact}, together with the perturbation value of the formation energy $(E_f^0)^{\text{pert}} = \Delta E_{\text{Mad}}^0 + B_0^{\text{exact}} + B_1^{\text{exact}}$ for a wide range of supercell sizes. This shows that the variation due to supercell effects as we go beyond a supercell of 27 sites is no more than 10^{-3} Hartree. The indication is that the value for E_f^0 for the 27-site supercell in table 4 represents the energy of formation of an isolated ideal vacancy to a precision of about 10^{-3} Hartree.

4.4. The ionic relaxation energy

We have used the procedure described in § 3.6 to relax the electrons and the ionic positions simultaneously. For the eight-site system, the translational symmetry prevents relaxation, so the calculations have been done only for the 16- and 27-site systems. The difference between the fully relaxed energy of the vacancy system and

Table 6. Ionic relaxation energy ΔE_f and final results for the relaxed vacancy formation energy E_f. Energies in hartrees (eV in parentheses).

Site no.	ΔE_f	E_f
16	−0.0036	0.0210 (0.57)
27	−0.0031	0.0205 (0.56)

its ground-state energy when the ions are fixed on their regular lattice sites gives the ionic relaxation energy ΔE_f. Results for this energy, together with the relaxed formation energy $E_f = E_f^0 + \Delta E_f$ are given in table 6. For a large system, we might expect the periodicity to reduce the magnitude of ΔE_f, since it has the effect of a constraint. It is perhaps a little surprising, then, that the relaxation energy is slightly smaller for the larger system; presumably this is a consequence of interactions between the defects. However, the difference of ionic relaxation energy between the two sizes of system is small, and it seems unlikely that this energy would be significantly different for larger repeating cells. Our result of 0.56 eV is quite close to the experimental value of 0.66 eV (Triftshäuser 1975, Berger *et al* 1978, Fluss *et al* 1978).

5. Discussion

The results we have presented indicate that it is now possible to calculate defect energies in the simple metals with adequate precision. Several possible reasons can be suggested for our remaining small discrepancy with experiment for the vacancy energy in aluminium: (i) our calculations refer to zero temperature, but the experiments are performed at high temperature; (ii) the pseudopotential we have used is not good enough; and (iii) the LDA is not good enough. In fact, there is some indication from the work of Jacucci *et al* (1981) that there is a significant temperature dependence of the enthalpy of formation in aluminium. According to their results, an increase of ~0.1 eV might be expected on going from zero temperature to 860 K. This would be enough to account for our discrepancy. But it also seems likely that, to the accuracy we are now discussing, a local representation of the pseudopotential would not be reliable. We have based our calculations on a local pseudopotential because the local approximation appears to be reasonably good for aluminium, but we would not expect it to give completely quantitative results. The extension of our scheme to treat non-local pseudopotentials will therefore be essential. A number of proposals for including non-locality in Car–Parrinello-like schemes have been made (Car and Parrinello 1987, Car *et al* 1987, Allan and Teter 1987), and the required extension seems unlikely to pose serious problems, at least for the simple metals. We are currently studying this question. Once this has been done, we hope the way will be open to the *ab initio* study of a wide range of defect properties (formation, migration and binding energies of vacancies, interstitials, impurities, etc) in the simple metals. We expect that such calculations will significantly improve our understanding of, for example, mass transport in these metals.

One of the themes of this work has been the relation between the full *ab initio*

calculation and low-order perturbation theory. The jellium calculations of Evans and Finnis (1976) suggested that non-linear terms in the electronic relaxation energy would give a contribution of ~1 eV. Our results fully confirm this: for the ideal vacancy (§ 4.2), the difference between our calculated formation energy and the second-order perturbation value is 0.8 eV. A large part of this (0.6 eV) is accounted for by the third-order term. This encourages the thought that perhaps a simple but accurate representation of the total energy as a function of ionic configuration may be found for metals like aluminium.

Perturbation theory also plays a central part in the present work in providing a powerful check on the technical approximations used in the full calculations. This is a very important matter, given the high precision that must be achieved to obtain useful results for the vacancy energy. By combining our full calculations with perturbation theory in the appropriate way, we have been able to reduce the purely technical errors to ~0.03 eV, which is more than adequate for most practical purposes. One very important aspect of this concerns the effect of interactions between the defects in supercell geometry. It was suggested by Chakraborty et al (1981) and by Chakraborty and Siegel (1982) that such effects might be responsible for the large (~1 eV) difference between their calculated vacancy energy in aluminium and the experimental value. As we have seen, if we work with a 27-site supercell, the correction due to such effects is unlikely to be more than ~0.03 eV.

Lastly, we comment on the energy minimisation method itself. In developing this method, we were guided by the desire to minimise simultaneously with respect to the wavefunctions and the ionic positions, following Car and Parrinello. This approach encounters difficulties in a metal, because of Fermi-surface effects. Our proposed solution to this problem, which involves treating the system at finite temperature, is related to other methods that are in use for smoothing the discontinuity in occupation number at the Fermi energy (Fu and Ho 1983, Needs et al 1986). Although this device turns out to be efficient and accurate, it does make the method rather complicated, and it seems likely that our method could be simplified and improved. As an indication of the magnitude of the computing task, we note that a full relaxation calculation on the 27-site system presently requires about 2 h of CPU time on the Cray 2; calculations on the smaller cells are, of course, more rapid. We believe that there is considerable scope for improving the efficiency of the calculations, and that one can expect to reduce the computation time by a factor of 2 or more. This will be a task for the future.

References

Allan D C and Teter M P 1987 *Phys. Rev. Lett.* **59** 1136
Baraff G A and Schlüter M 1979 *Phys. Rev. B* **19** 4965
Baraff G A, Schlüter M and Allan G 1983 *Phys. Rev. Lett.* **50** 739
Berger A S, Ockers S T and Siegel R W 1978 *J. Nucl. Mater.* **69 & 70** 734
Callaway J and March N H 1984 *Solid State Phys.* **38** 136
Car R, Kelly P J, Oshiyama A and Pantelides S T 1984 *Phys. Rev. Lett.* **52** 1814
Car R and Parrinello M 1985 *Phys. Rev. Lett.* **55** 2471
—— 1987 *Proc. 18th Int. Conf. on the Physics of Semiconductors (Stockholm 1986)* ed. O Engström
    (Singapore: World Scientific) vol 2, p 1165
Car R, Parrinello M and Andreoni W 1987 *Proc. 1st NEC Symposium on Fundamental Approaches to New
    Material Phases (Tokyo) 1986* ed. S Sugano, Y Nishina and S Ohnishi (New York: Springer)
Catlow C R A, Corish J, Jacobs P W M and Lidiard A B 1981 *J. Phys. C: Solid State Phys.* **14** L121
Ceperley D M and Alder B J 1980 *Phys. Rev. Lett.* **45** 566
Chadi D J and Cohen M L 1973 *Phys. Rev. B* **8** 5747
Chakraborty B and Siegel R W 1982 *Point Defects and Defect Interactions in Metals* ed. J Takamura, M
    Doyama and M Kiritani (New York: North-Holland) p 93
Chakraborty B, Siegel R W and Pickett W E 1981 *Phys. Rev. B* **24** 5445
Chang R and Falicov L M 1971 *J. Phys. Chem. Solids* **32** 465
Cohen M L 1986 *Int. J. Quantum Chem.* **29** 843
Evans R and Finnis M W 1976 *J. Phys. F: Met. Phys.* **6** 483
Fletcher R 1980 *Practical Methods of Optimization* vol 1 (New York: Wiley) p 63
Fletcher R and Reeves C M 1964 *Comput. J.* **7** 149
Fluss M J, Smedskjaer L C, Chason M K, Legnini D G and Siegel R W 1978 *Phys. Rev. B* **17** 3444
Fu C-L and Ho K-M 1983 *Phys. Rev. B* **28** 5480
Goodwin L S 1987 *PhD Thesis* Cavendish Laboratory, University of Cambridge
Hammerberg J and Ashcroft N W 1974 *Phys. Rev. B* **9** 409
Harrison W A 1966 *Pseudopotentials in the Theory of Metals* (New York: Benjamin)
Heine V 1970 *Solid State Phys.* **24** 1 (New York: Academic)
—— 1984 *NATO ASI on Electronic Structure of Complex Systems (Ghent) 1982* ed. P Phariseau and W T
    Temmerman (New York: Plenum) p 1
Heine V and Weaire D 1970 *Solid State Phys.* **24** 250
Hohenberg P and Kohn W 1964 *Phys. Rev.* **136** B864
Ihm J 1988 *Rep. Prog. Phys.* **51** 105
Ihm J, Zunger A and Cohen M L 1979 *J. Phys. C: Solid State Phys.* **12** 4409
Jacucci G, Taylor R, Tenenbaum A and Doan N V 1981 *J. Phys. F: Met. Phys.* **11** 793
Kohn W and Sham L J 1965 *Phys. Rev.* **140** A1133
Levy M 1982 *Phys. Rev. A* **26** 1200
Lloyd P and Sholl C A 1968 *J. Phys. C: Solid State Phys.* **1** 1620
Louie S G, Schlüter M, Chelikovsky J R and Cohen M L 1976 *Phys. Rev. B* **13** 1654
Mermin N D 1965 *Phys. Rev.* **137** A1441
Needels M, Payne M C and Joannopoulos J D 1987 *Phys. Rev. Lett.* **58** 1765
Needs R J, Martin R M and Nielsen O H 1986 *Phys. Rev. B* **33** 3778
Payne M C, Bristowe P D and Joannopoulos J D 1987 *Phys. Rev. Lett.* **58** 1348
Payne M C, Needels M and Joannopoulos J D 1988 *Phys. Rev. B* **37** 8138
Perdew J and Zunger A 1981 *Phys. Rev. B* **23** 5048
Phariseau P and Temmerman W T (ed.) 1984 *NATO ASI on Electronic Structure of Complex Systems
    (Ghent) 1982* (New York: Plenum)
Popovic Z D, Carbotte J P and Piercy G R 1974 *J. Phys. F: Met. Phys.* **4** 351
Scheffler M 1987 *Physica* **146B** 176
Simmons G and Wang H 1971 *Single Crystal Elastic Constants and Calculated Aggregate Properties: a
    Handbook* (Cambridge, MA: MIT Press)
So C B and Woo C H 1981 *J. Phys. F: Met. Phys.* **11** 325
Srivastava G P and Weaire D 1987 *Adv. Phys.* **36** 463
Taylor R 1978 *J. Phys. F: Met. Phys.* **8** 1699
Triftshäuser W 1975 *Phys. Rev. B* **12** 4634