Vacancies and impurities in aluminum and magnesium

N. Chetty and M. Weinert
Department of Physics, Brookhaven National Laboratory, Upton, New York 11973-5000

T. S. Rahman
Department of Physics, Kansas State University, Manhattan, Kansas 66506

J. W. Davenport
Department of Physics, Brookhaven National Laboratory, Upton, New York 11973-5000
(Received 1 May 1995)

The vacancy formation energies and (Mg, Al, and Si) impurity heats of solution are calculated for Al and Mg using a first-principles pseudopotential approach and large supercells. While the interaction of the defects considered here are already negligible for reasonably small unit cells, adequate sampling of the Brillouin zone is found to be essential for these metallic systems, even for systems containing more than 100 atoms per unit cell; e.g., the vacancy formation energy of Al for 108 atoms per cell has the incorrect sign if only the Γ point is sampled. When the volume and structural relaxations are treated consistently, heats of formation and solution and relaxation volumes are obtained that are in good agreement with the available experimental data. Simple trends in the relaxations around the impurities in the various materials can be understood in terms of the size of the impurities compared with the host atoms. Contrary to some commonly used models, the energetics of the impurities are found to be dominated by electronic, rather than elastic, contributions. The defect-induced changes to the local electronic structure are also discussed.

I. INTRODUCTION

Impurities and defects play an important role in determining the properties of real materials. The presence of defects breaks the translational symmetry of the system. Differences in atomic size and in the number of electrons of an impurity will perturb both the electronic and structural properties. The presence of a defect may cause the formation of localized states, which in turn may drastically alter the transport properties. Likewise, the difference in size between the host and the impurity atom may cause the lattice to locally distort and induce long-range strains. These distortions may affect the mechanical properties by changing the elastic properties or may even cause a phase transition. These impurity-bulk interactions may be attractive or repulsive. Even when the interaction is repulsive, there will be an equilibrium concentration of impurities at finite temperatures. Simple defects such as vacancies will behave similarly, although defects generally will have an energy cost associated with their creation (repulsive interaction).

In the study of phase diagrams, studies of impurities provide important information about the low concentration regime. While the properties of defects and impurities are important for a fundamental understanding of the behavoir of materials, much of the basic thermodynamical information is very difficult or sometimes impossible to obtain from experiment.

The local density theory[1–3] has proved surprisingly accurate in predicting ground state properties of solids; it has become fairly routine to extract results that compare favorably with experiment. The time is now approaching when computer experiments may be considered reliable enough to be as useful as real experiments, with the added advantage that computer simulations can give new insights and enhance our understanding of real materials, often pointing to the essential physics at play. Hypothetical structures such as metastable and unstable phases are readily studied on the computer, but are not necessarily realizable in the real experimental world. Extreme conditions of pressure and temperature that may be inaccessible to present experimental means may be easily considered in a computer simulation of the material.

To treat isolated impurities and defects, one needs to consider large—in principle infinite—systems since the translational symmetry is lost. Although methods do exist that can treat the isolated impurity,[4] a common approximation is to use supercells and artificially reimpose the translational symmetry by repeating the defects. The advantage of a supercell approach is that the same methods can be used for both the ideal and defect-impurity systems. Such methods are also applicable to disordered systems[5] such as liquids[6–12] or random alloys, where the calculational cell might have no bearing on any underlying symmetry of the system and is purely an artifact of the calculation. Notwithstanding the inexorable interest in looking at larger and larger systems, we believe that there remain a number of interesting phenomena that can be studied using cells of order of a few hundred atoms, for example, the phase stability of alloys including low concentration alloys at the 1% level, dislocations and strain fields, surface reconstructions, etc.

Over the years, most first-principles calculations for large (~100 atom) systems have been in the field of semiconductors, both because of the obvious technological interest and also because such systems are relatively simpler to study compared to metals. Metals pose unique and interesting problems from a physics and computational point of view and a greater effort is required to derive accurate results.[13–18] The main point here is the often complicated nature of the Fermi surface that must be described by a detailed sampling of the Brillouin zone (BZ). The variable occupation of states adds to the number of degrees of freedom that affects the overall stability of the solutions of the Kohn-Sham (KS) equations.

In this work, we show that large metallic systems can be studied accurately. We emphasize the need for accurate sampling of the BZ, even for systems of a relatively large number of atoms. To expound on these points, we have chosen as our prototypical systems fcc Al and hcp Mg. Al, because of its useful properties of lightness, strength, and resistance to wear, is the material of choice in many technological applications. The Al-Mg alloy phase diagram includes a number of complicated phases and presents a challenge to explain theoretically.[19] In this paper, we calculate the vacancy formation energies in Al and Mg, and the heats of solution of impurities of Si and Mg in Al and of Si and Al in Mg. Considering these same-row elements in adjacent columns readily facilitates comparisons that are useful in interpreting the results. We attain converged results by considering various samplings of the BZ together with different sizes of the system cell. Where possible, we compare our results with experiment.

In Sec. II and the Appendixes, we discuss our approach for the iterative solution of the KS equations and the minimization of the total energy with respect to the atomic positions. The studies of **k**-point convergence and the results for the Al and Mg systems are given in Sec. III, and a brief summary with conclusions is given in Sec. IV.

II. METHOD

The standard self-consistent solution of the Kohn-Sham (KS) equations is a nonlinear process: A starting Hamiltonian is constructed from a guessed input charge density and the Hamiltonian (within a given basis) is diagonalized to obtain the KS eigenvalues and eigenvectors. The Fermi level is then determined, and the charge density computed. This output charge density is mixed carefully with the input density to construct a new Hamiltonian, and the process is repeated until self-consistency is achieved and the total energy is stationary.

The availability of soft pseudopotentials[20–24] has made the study of a great variety of solid state systems using plane waves feasible. The main computational hurdle that needs to be overcome in traditional exact-diagonalization techniques[25] is the so-called $\mathcal{O}(N^3)$ problem, where the computational time scales with the third power of the system size N.

The diagonalization of the Hamiltonian is generally the computationally intensive task. Conventional methods require the storage of the entire Hamiltonian. A number of years ago, Car and Parrinello[26] presented an efficient means of diagonalizing the Hamiltonian. Since then, a number of iterative schemes[27,28] have been devised that work on the same principle: A set of states is iterated upon by the Hamiltonian, usually followed by an orthogonalization procedure.[29] The main attribute of such iterative schemes is that the entire Hamiltonian need not be stored. Since one is interested only in the lowest (occupied) states of the system, it suffices to simply compute the action of the Hamiltonian on a subspace of occupied states, rather than on all the states. This procedure can be done efficiently using fast-Fourier-transform (FFT) techniques for a plane wave basis. The end result is the set of the lowest-energy eigenstates, which is all that is needed for the determination of the ground state properties of the system under consideration.

In our work, we use the plane wave pseudopotential technique[30] to solve the KS equations. Use is made of the Vosko-Wilk-Nusiar[25,31] local density parametrization of the exchange-correlation potential and energy. The ionic pseudopotentials are of the Troullier-Martins[22] form, from which the Kleinman-Bylander[32,33] fully nonlocal separable potentials are constructed. For the sampling of the BZ, we use the Monkhorst-Pack[34] special points technique; the symmetry of the lattice is used to reduce the set of **k** points to the irreducible part of the BZ.

The iterative diagonalization procedure that we employ is a variant of a preconditioned steepest descent (PSD) algorithm. The algorithm is applied to the electronic problem for a fixed set of ionic positions. There is no line minimization of the energy along the direction of steepest descent; instead, we take a step along this direction inversely proportional to the kinetic energy of the state. Details concerning the convergence of the general PSD method and its implementation in a plane wave basis are given in the Appendixes. Our complete procedure, which we now discuss, differs in a number of ways from other work.

Given a set of n_s trial wave functions at each **k** point and a potential, the states are propagated according to Eq. (B5). The number of states, n_s, is typically ~1% of the total number of plane waves and is larger than the number of occupied states. At this point, these (nonorthogonal) trial states are diagonalized with respect to the Hamiltonian H:

$$\sum_{m=1}^{n_s} \langle \psi_{\mathbf{k}n} | H | \psi_{\mathbf{k}m} \rangle \, D_{\mathbf{k}mn} = \varepsilon_{\mathbf{k}n} \sum_{m=1}^{n_s} \langle \psi_{\mathbf{k}n} | \psi_{\mathbf{k}m} \rangle \, . \quad (1)$$

The coefficients $D_{\mathbf{k}mn}$ are then used to determine the exact eigenfunctions, with eigenvalues $\varepsilon_{\mathbf{k}n}$, for the subspace spanned by the states $|\psi_{\mathbf{k}n}\rangle$:

$$|\tilde{\psi}_{\mathbf{k}n}\rangle = \sum_{m=1}^{n_s} D_{\mathbf{k}mn} |\psi_{\mathbf{k}m}\rangle \, . \quad (2)$$

The subspace diagonalization serves the purpose of orthogonalizing the states and of determining the next approximation to the KS eigenvalues. Other methods for

orthogonalizing the states besides a subspace diagonalization could be considered, especially since this step is an $\mathcal{O}(N^3)$ process. (In practice, the FFT's are still the rate limiting step for the system sizes under consideration.) However, replacing the subspace diagonalization by, for example, a Gram-Schmidt procedure would have several disadvantages.

First, since we are interested in metals, we must allow for the possibility that some states may be either partially or completely unoccupied. Because n_s is greater than the number of occupied states (and hence the occupation numbers are not all equal), a unitary transformation of the subspace states will not leave the density, or any other quantity that can be written as a trace, invariant; the (fractional) occupation numbers of metallic systems are associated with eigenstates, *not* linear combinations of eigenstates.

A second, but more subtle, consequence of our implementation of the subspace diagonalization is that we do not have the instabilities during the iterative diagonalization that have been observed elsewhere.[35,36] The instabilities can be related to the fact that the propagation of the trial states does not maintain orthogonality of the states and that the subspace spanned by the trial states changes. These points are crucial and point out a difficulty in using updating schemes more sophisticated than the PSD method. Methods such as conjugate gradients effectively use information from a number of previous iterations; unfortunately, the space spanned by the trial states changes between steps, both because of the orthogonalization (done by the Gram-Schmidt procedure, for example) and the equation of motion. The result is that the information about previous steps contained in the updating scheme is given with respect to a rotated (and different) space. If the rotation of the space is large, then this previous information can force the states in a completely wrong direction. The PSD method, on the other hand, uses information only from the present iteration, i.e., local information, and will not suffer from this problem. Moreover, since the subspace of the previous step was diagonalized—thereby determining the minimum of that subspace—the search need only occur in the orthogonal subspace. Since no information about the orthogonal subspace exists from the previous iterations, there does not appear to be any advantage to methods that make explicit use of previous steps. Note that this combination of the PSD method and subspace diagonalization does not suffer from the "long valley" problem of standard steepest descent methods and that the convergence and stability of the method are quite reasonable. Whether a method better than the PSD method can be found for optimally determining the new subspace is an open question, but any such method must take into account the rotation of the subspace resulting from orthogonalization.

Once the subspace diagonalizations and rotations for each $\mathbf{k}$ point have been carried out, these states are fractionally occupied according to a Fermi distribution

$$f_{\mathbf{k}n} = \left[\exp\left\{\frac{(\varepsilon_{\mathbf{k}n} - \varepsilon_f)}{k_B T}\right\} + 1\right]^{-1}, \tag{3}$$

with a thermal broadening $k_B T$ and Fermi level ε_f which is adjusted to accommodate the fixed number of electrons. The additional degrees of freedom that arise in the determination of the occupation numbers $f_{\mathbf{k}n}$ require that an entropylike term[14] be included to maintain the overall variational nature of the total energy expression:

$$\begin{aligned} E_{\text{tot}} = &\sum_{\mathbf{k}n} \varepsilon_{\mathbf{k}n} f_{\mathbf{k}n} - \frac{1}{2}\int d\mathbf{r}\, d\mathbf{r}'\, \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + \gamma_{\text{Ewald}} \\ &+ \int d\mathbf{r}\, \{\epsilon_{\text{xc}}[\rho(\mathbf{r})] - \mu_{\text{xc}}[\rho(\mathbf{r})]\} + \alpha Z \\ &+ k_B T \sum_{\mathbf{k}n} [f_{\mathbf{k}n} \ln f_{\mathbf{k}n} + (1-f_{\mathbf{k}n}) \ln(1-f_{\mathbf{k}n})], \end{aligned} \tag{4}$$

where the charge density is given by

$$\rho(\mathbf{r}) = \sum_{\mathbf{k},n} |\psi_{\mathbf{k}n}(\mathbf{r})|^2 f_{\mathbf{k}n}, \tag{5}$$

μ_{xc} and ϵ_{xc} denote the exchange-correlation potential and energy, respectively, and γ_{Ewald} and αZ are the Ewald constant and the $\mathbf{g} = 0$ term of the local pseudopotential interaction energy, respectively.[25] Although minimizing this functional form of the total energy appears to be equivalent to minimizing the free energy of the system[37] at finite temperature, the "entropy" term[14] is a consequence of the fractional occupation and need not be associated with a physical temperature.

We use the rotated states as the input for the next iterative step [cf. Eq. (B5)] and then use a linear mixing of the charge density to update H. For large systems, the smaller size $\mathbf{g}$ vectors tend to amplify errors in the charge density because of the long-range nature of the Coulomb potential, the so-called "charge sloshing" problem. The use of Eq. (B4) effectively limits the time step from increasing indefinitely for small $\mathbf{g}$ vectors since the preconditioner remains bounded by $1/S$. This ensures that the low-energy components of the wave function are mixed in carefully, and are not allowed to dominate the self-consistency process (see the Appendixes).

Once the electronic wave functions and the charge density for a given set of atomic positions are converged, the Hellmann-Feynman forces[38] are calculated.[14] To generate the new set of atomic positions, we use a modified[39,40] Broyden-Fletcher-Goldfarb-Shano (BFGS) quasi-Newton update scheme where the Oren-Spedicato scaling is used in the second step to make the process self-scaling. In order to make the whole process automatic, we use a simple criterion in the first step that determines how far to move along the direction of the force: We assume that the forces on the atoms are comparable to the forces an atom sees during a phonon vibration. The distance to move an atom can then be related to the force, the mass (proportional to the atomic number), and a characteristic phonon frequency, for which we use the Debye temperature. With this method, we are able to rapidly minimize the total energy with respect to atomic positions. Even for large low-symmetry systems, only 3-5 atomic steps are needed to have all forces less than 10^{-4}–10^{-5} hartrees/a.u. and the total energy converged

to within μhartrees for a cell with N_a atoms.

Finally, we give some of the computational details of the calculations discussed in Sec. III. The wave functions are expanded in plane waves with energies up to 12 Ry, although some tests were done for cutoffs of 16 and 20 Ry. Since the energy differences were converged to a few meV per cell, we will only report the 12 Ry calculations. The temperature broadening for the Fermi distribution function is $k_BT = 0.001$ in hartrees, corresponding to approximately room temperature. We include a relatively large number of states above the Fermi level, including approximately 25% more states than we would occupy had the system been a band-gap material. The shift parameter defined in Eq. (B5) for the preconditioned matrix of the iterative diagonalization scheme is set to $S = 0.5$.

III. RESULTS

As discussed in the Introduction, vacancies and impurities are important from both a scientific and technological point of view. Diffusion in solids is often mediated by vacancies; thus, the mobility of vacancies and the interaction between vacancies and other vacancies or impurities in the crystal affect the physical and transport properties of the material. The formation energies of defects and heats of solution of impurities are important thermodynamical quantities that are related to and control the physical properties.

The vacancy formation energy ΔH_v (at T=0 K) is defined as the energy required to create a vacancy in the bulk. For a supercell of N atoms of metal A, ΔH_v is given by

$$\begin{aligned}\Delta H_v &= E_{\rm vac}[N-1] - (N-1)\,\mu_A \\ &= E_{\rm vac}[N-1] - \frac{N-1}{N}\,E_A[N]\,, \end{aligned} \tag{6}$$

where $E_{\rm vac}[N-1]$ is the total energy of the supercell containing $N-1$ atoms and the single vacancy, and the chemical potential μ_A is the energy per atom in the bulk obtained from the total energy $E_A[N]$ of a bulk supercell containing N atoms of element A. Atomic relaxations have the effect of reducing ΔH_v. The heat of solution ΔH of an impurity of atom B in metal A will have a form similar to Eq. (6):

$$\Delta H = E_{\rm impurity}[N-1] - (N-1)\,\mu_A - \mu_B\,, \tag{7}$$

where μ_A and μ_B are the corresponding chemical potentials (total energies per atom) of materials A and B, respectively, in their ground state (diamond structure Si, hcp Mg, and fcc Al). While μ_A is calculated using parameters equivalent to those used for the impurity system, μ_B is obtained from separate bulk calculations which are converged with respect to **k**-point sampling. Note that ΔH is independent of constant shifts in the atomic energy; thus the calculated heats of solution are independent of, for example, the typical local density overestimate of cohesive energies or the choice of zero of the pseudopotential.

A. Size effects in supercells

The vacancy formation energy of Al has been studied by a number of different authors.[41–51] The results from a number of different self-consistent local density approximation (LDA) calculations vary from around[45,46,48,51] 0.55 eV, to[50] 0.71 eV, to about[47,49] 0.85 eV, with some older values[44] as high as 1.5 eV; for comparison, the experimental value[52] is 0.67±0.03 eV. In this section, we use the vacancy and Si and Mg impurities in Al as test cases for convergence studies of the size effects; the physics of the results will be discussed later.

There are two types of size effects that will be considered: (1) unit cell size (N_a atoms per cell) and (2) **k**-point sampling (N_k^0 points in the Brillouin zone). If one is interested in determining the properties of the isolated defect, then N_a should be increased until the results are independent of N_a. For a given N_a, the electronic structure may in turn depend on the choice of N_k^0. To get a converged result, both N_a and N_k^0 must be considered. While at first glance N_a and N_k^0 appear to be independent, they are, in fact, closely related: Consider a crystal with N unit cells of N_a atoms, i.e., a crystal with $N_{\rm tot} = NN_a$ atoms. Then Born–von Karman periodic boundary conditions generate exactly N points in the BZ, although this structure of the BZ is often ignored. If the logic is now inverted, the choice of **k**-point sampling and supercell size together determines the effective total number of atoms in the crystal:

$$N_{\rm tot} = N_a N_k^0\,. \tag{8}$$

In order to make a reasonable approximation to the bulk, $N_{\rm tot}$ needs to be large enough. Obviously, in practical calculations there is a trade-off between supercell size and **k**-point sampling. The simplest example of this behavior is that for ideal systems there are various combinations of N_a and N_k^0 that yield the same $N_{\rm tot}$. These different choices will give *identical* results *if* the different sets of **k** points are equivalent (the well-known "band-folding" effect).

In Table I, we give the calculated formation energy

TABLE I. The ideal (unrelaxed, a=7.5056a_B) vacancy formation energy of Al (ΔH_v, in eV) computed for different size unit cell and various samplings of the Brilloiun zone. $N_{\mathbf{k}}$ is the number of (special) **k** points sampled and $N_{\rm tot}$ is the effective number of sites in the crystal.

Atoms	$N_{\mathbf{k}}$	$N_{\rm tot}$	ΔH_v
32	1	32	−0.14
32	4	2048	0.80
32	10	6912	0.68
32	20	16384	0.83
32	35	32000	0.82
108	1	108	−0.53
108	4	6912	0.68
108	10	16384	0.82
256	1	256	1.20

of an ideal vacancy in Al (at fixed lattice constant) for different values of N_a and N_k, where N_k is the number of symmetry inequivalent points in the first BZ. Also given are the corresponding values of $N_{\rm tot}$. For the cases N_k=1, the **k**-point sampling is limited to the standardly used Γ point. For equal values of $N_{\rm tot}$, but different values of N_a, the energies per atom for bulk Al are identical. First note that for N_k=1 all the values of ΔH_v—even for 256 atoms per cell—are significantly in error, with the 32- and 108-atom cells predicting the wrong sign. The Γ-point samplings, with their small values of $N_{\rm tot}$, would seem to suggest that there are long-range vacancy-vacancy interactions in Al and that significantly larger cell sizes are necessary.

Such a conclusion, however, is incorrect. First consider the convergence of the 32-atom cell with N_k. The results are converged to ~0.01 eV for 20 special points. Comparisons of equivalent **k**-point sets ($N_{\rm tot}$) for the 32- and 108-atom cells, show that the vacancy-vacancy interaction differs by less than ~0.01 eV between these choices of N_a. Thus the vacancy-vacancy interactions appear to be almost negligible already for the 32-atom cell, but only if the BZ is adequately sampled. From the results in Table I, it is clear that in order to get comparable accuracy using Γ-point-only sampling would require unit cells on the order of at least 10^3–10^4 atoms. Since the computational cost is linear in **k** points, increasing N_k rather than N_a is clearly a more efficient way to increase $N_{\rm tot}$; this trivial way to get effectively $\mathcal{O}(N)$ scaling is only appropriate when N_a is large enough to include the important long-range interactions.

The reason that **k**-point sampling beyond the Γ point is necessary is simply that the electronic structure for small effective crystal sizes does not resemble bulk Al. In Fig. 1, we show the total density of states for the 108-atom cell of pure Al, sampled either at only the Γ point or at four special **k** points. While the four **k**-point sampling generally already has the correct shape of the Al bulk density of states, the Γ-point sampling does not resemble a nearly free electron density of states. These results, and the need for adequately sampling the BZ, should not be particularly surprising since the itinerant electrons in metallic systems are quite sensitive to the boundary conditions imposed by the finite crystal size. An important conclusion that can be drawn is that simply using the largest supercells possible is not only an inefficient use of resources, but also may lead to incorrect results; bigger is not necessarily better.

Similar convergence studies were also carried out for Mg and Si impurities (cf. Table II). The impurity-impurity interactions are again basically negligible for the 32-atom cells. As before, **k**-point sampling was found to be important: For the Si impurity, the errors resulting from sampling the Γ point only were smaller than in the case of the vacancy, but still unacceptably large (~0.7 eV for the 32-atom cell, ~0.2 eV for the 108-atom cell). Moreover, the results in Tables I and II show that the convergence with **k** points (for fixed N_a) is different for each system. For example, the Mg results are converged to ~0.01 eV with respect to cell size (for comparable **k**-point samplings), but show poorer convergence with respect to **k**-point sampling than do the Si results. These results demonstrate that some convergence studies are necessary for any new system.

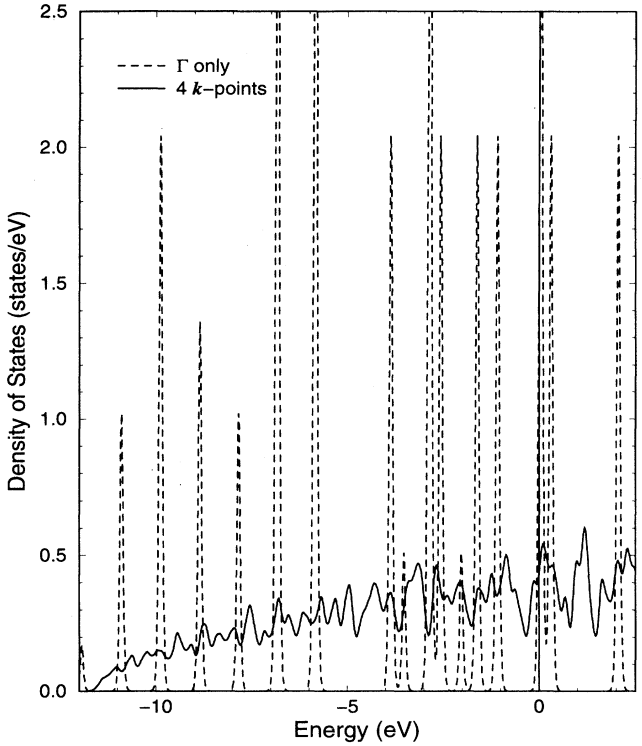

FIG. 1. Comparison of the electronic density of states of the 108-atom Al cell using Γ-point (solid line) and 4 special **k**-point (dashed line) sampling. [Each eigenvalue is Fermi broadened, $\sim\mathrm{sech}^2(\varepsilon-\varepsilon_i)/(2k_BT)$, with k_BT=0.001 hartree.]

In addition to the ideal (unrelaxed) impurities, the relaxed heats of solution are also given. The relaxation energy, however, is quite insensitive to the **k**-point sampling, even though the relaxed atomic positions for different choices of N_a and N_k varied slightly. Likewise, and perhaps somewhat surprisingly, the relaxation energy was also rather insensitive to the cell size as can be seen by comparing the 32- and 108-atom results in Table II; even for the vacancy, with its significantly larger relaxation energy, the corresponding difference is less than 0.01 eV.

TABLE II. The heats of solution (in eV) of Si and Mg impurities in Al for various unit cell sizes and number of special **k** points ($N_{\mathbf{k}}$) at fixed lattice constant (a=7.5056a_B).

	32 atoms		108 atoms
$N_{\mathbf{k}}$	10	35	4
Si impurity			
unrelaxed	0.42	0.41	0.41
relaxed	0.41	0.40	0.40
Mg impurity			
unrelaxed	0.03	0.07	0.02
relaxed	0.01	0.05	0.00

B. Al vacancy

The presence of impurities and vacancies causes a number of changes in the properies of a material; one of the most obvious is a measurable change of the lattice constant with concentration of defects. This macroscopic fractional change in volume $\Delta V/V_0$, together with the concentration of defects c, can be used to define an effective size of an impurity (formation) volume V^F or equivalently a relaxation volume $\Delta V^{\rm rel}$,

$$\frac{\Delta V^{\rm rel}}{\Omega_0} = \frac{V^F}{\Omega_0} - 1 = \frac{1}{c}\frac{\Delta V}{V_0}, \tag{9}$$

where Ω_0 is the volume per atom of a host atom. For cubic systems, this can be written in terms of the lattice constants of the ideal (a_0) and defect (a) systems as

$$\frac{\Delta V^{\rm rel}}{\Omega_0} = \frac{1}{c}\left[\left(\frac{a}{a_0}\right)^3 - 1\right]. \tag{10}$$

Taking the volumes into account correctly can have important quantitative effects on the calculated values. A rough estimate of the error in total energy resulting from the use of a fixed lattice constant for both the bulk and the impurity systems for an N atom cell is

$$\Delta E \approx N\frac{1}{2}B\Omega_0\left(\frac{\Delta V^{\rm rel}}{\Omega_0}\right)^2 c^2, \tag{11}$$

where B is the bulk modulus; ΔE is simply the energy (within the harmonic approximation) required to change the volume of each atom by an amount $c\Delta V^{\rm rel}$. Since $c\sim N^{-1}$, the error vanishes as $c\to 0$ and $N\to\infty$, but for finite values there will be a contribution. For the Al vacancy, the difference between a 32- and 108-atom cell is $\sim$0.01 eV, consistent with the results in Table I.

The fully relaxed heat of formation and relaxation volume for the Al vacancy are given in Table III. The error bars given are based on estimates of the variations with cell sizes and **k**-point sampling. Note that small changes in the lattice constants can have large effects on the calculated values of $\Delta V^{\rm rel}$ and that the problem becomes more severe for larger cells (smaller concentrations). The energies given in Table III are based on **k**-point samplings of up to $N_{\rm tot}$= 87 808 atoms, corresponding to 2744 **k** points in the irreducible wedge for fcc Al; these larger samplings were not done for the unrelaxed structures and thus were not given in Tables I and II. Interestingly, the **k**-point convergence for the impurity systems appears to be faster than the convergence for bulk Al, so that most of the error bar is associated with bulk Al. (Note that the changes in energy for different **k**-point samplings are small on the scale of what one normally considers: To get an absolute convergence of 0.01 eV for a 108-atom bulk cell with respect to **k**-point sampling requires that the bulk total energy be converged to 3 μhartrees/atom.)

Our calculated heat of formation of a vacancy in Al is 0.66±0.03 eV, in good agreement with experiments and also reasonably consistent with previous calculations. This value differs significantly from the values given in the previous section because of volume effects (the results in Tables I and II used a bulk lattice constant 0.4% larger than the calculated one) and relaxation effects. For a given lattice constant, the relaxation energy of the atoms is $\sim$0.08 eV. As expected, the relaxation of the first shell of atoms around the impurity is inward by 1.9% of the nearest-neighbor distance. The second and third shells of atoms also relax inward by 1.0% and 0.7%, respectively. The relaxation of these further shells appears at first glance to converge rather slowly, but the motion of these shells can be understood quite simply if one considers the *relative* displacements of the shells: The distance between the first and second (second and third) shells is decreased by 0.2% (0.1%) compared to the ideal structure, values that are an order of magnitude smaller than the changes relative to the vacancy. Thus the relaxation of the more distant shells of atoms is dominated by a desire to maintain the local bond lengths and coordination of the atoms.

The hard-core repulsion between the first shell of atoms limits the relaxation around the vacancy to $\sim$2% and leaves a significant void. The calculated relaxation volume of -0.33 (in units of Ω_0) for the vacancy is in good agreement with (at least some of) the experimental results and the previous value of -0.28 calculated by DeVita and Gillan.[48] As expected, our value of $\Delta V^{\rm rel}$ falls between the limits of no relaxation ($\Delta V^{\rm rel}$=0) and complete collapse of the volume ($\Delta V^{\rm rel}$=-1). An important consequence of a value for $\Delta V^{\rm rel}$ of this order is that elastic energies [cf. Eq. (11)] are too small to describe the vacancy formation energy alone; i.e., electronic effects are necessary.

TABLE III. The fully relaxed formation enthalpies and heats of solution (ΔH^F) and the relaxation volumes ($\Delta V^{\rm rel}$) for the vacancy and Mg and Si impurities in Al; all energies in eV and relaxation volumes in units of the bulk atomic volume of Al. The error bars on the calculated values are similar for all systems considered.

		Calculated	Experiment
ΔH^F	vacancy	0.66±0.03	0.67±0.03[a]
			0.64±0.04[b]
			0.60−0.77[c]
	Mg	0.07	0.20[d]
			0.08[e]
			0.06[f]
	Si	0.37	0.51[d]
$\Delta V^{\rm rel}$	vacancy	−0.33±0.05	−0.38[g]
			−0.05[c]
	Mg	0.30	0.34[d]
	Si	−0.18	−0.13[d]

[a] "Recommended value" of Ref. 52.
[b] Reference 68.
[c] See references in Ref. 52.
[d] Reference 53.
[e] Reference 64.
[f] Reference 65.
[g] Reference 69.

C. Impurities of Si and Mg in Al

Both Si and Mg are important constituents in Al alloys. As such, the properties of dilute amounts of these impurities are of interest. In Table III the heats of solutions and relaxation volumes are given for these systems. In both cases the heats are positive, i.e., repulsive. The positive heat for Mg is somewhat surprising since Al and Mg do form ordered, albeit complicated, alloys. The sign, however, is consistent with various experimentally derived values. The values from Ref. 53 are derived from the (high-)temperature dependence of the concentration of impurities. The extrapolation of these data is such that we expect these values to be upper bounds to the T=0 K heats, but of the correct sign. The other available data for ΔH^F for Mg in Al bracket our calculated value.

The relaxation volume for Mg is comparable to that of the vacancy, but of opposite sign (Mg is larger than Al), while the relaxation energy is rather small compared to the vacancy. Likewise, the atomic relaxations are smaller than in the case of the vacancy, with the first shell moving outward 0.9%, the second inward 0.7%, and the third shell moving outward by less than 0.1%; just as for the vacancy, the relative differences between shells are approximately an order of magnitude smaller. The calculated relaxation volume of 0.30 is comparable to, but slightly smaller than, the 40% difference one would expect from pure atomic sizes; such deviations arise from the fact that the atoms are not simply hard spheres, but that the electrons associated with each atom deform.

The relaxations for Si are even smaller than for Mg, with $\Delta V^{\rm rel}$=−0.18. Since Si is smaller than Al, the first shell of atoms relaxes inward by ∼0.5% and the relaxation energy is ∼0.01 eV. The calculated heat of 0.37 eV is smaller, as expected, than that deduced from the temperature dependence of the Si concentration.

Overall, the calculated results are in good agreement with experiment, where available. Taken as a set, the calculated results give us confidence that we are able to describe these types of systems correctly.

Beyond simply obtaining numbers, we can also obtain new insights from the calculations. Based on our results, we argue that the magnitude of the Si heat of solution can be understood by considering electronic contributions to the energy, rather than strain-elastic contributions. In the Al lattice, the Si atom sees an fcc environment. fcc Si (calculated at the Al lattice constant) looks similar to Al; i.e., a rigid band model for the density of states describes fcc Si quite well (Fig. 2). Hence, a Si impurity in Al will behave much as a nearly free-electron impurity (see also Sec. III G), although the Si ground state is the cubic diamond structure. The energy difference between the cubic diamond and fcc structures for Si is ∼0.5 eV/atom, which is comparable to the heat of solution. Thus the major contribution to the heat of solution is electronic in nature: Si prefers the tetrahedral coordination of the diamond structure rather than the fcc environment found in Al. These results point out that simple strain arguments[53] based on size alone are not able to provide a consistent description of the energetics of impurities and defects, although such models are commonly used.

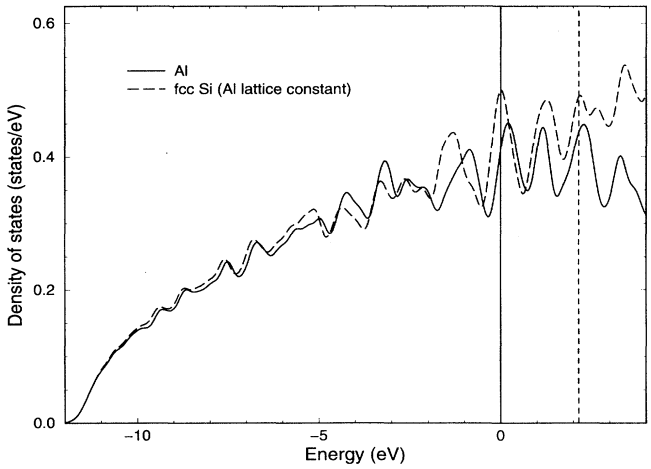

FIG. 2. The density of states for the fcc Al (solid line) and fcc Si (at the Al lattice constant) (dashed line) cell. The bottom of the bands are aligned; Fermi levels for each system are given by the vertical lines.

D. Bulk properties of Mg

Although Al has been studied quite often, relatively few calculations for Mg exist. Mg crystallizes in the hcp structure and undergoes a pressure-induced martensitic transition to the bcc phase at around 50 GPa.[54] In this section, we summarize the bulk properties of hcp, fcc, and bcc Mg that we need to describe the vacancy and impurities results. We use 90, 100, and 110 special **k** points in the irreducible part of the zone for the hcp, bcc, and fcc unit cells, respectively. In Table IV, we tabulate the results for the lattice constant, bulk modulus, and cohesive energy. The results are in reasonable accord with available experimental and other theoretical results.[54–56]

We tested the convergence of the bulk properties of the hcp structure as a function of cutoff energy (12 Ry and 20 Ry) and the number of **k** points (90 and 168 special points) and found the structural properties tabulated in Table IV to be well converged.

TABLE IV. The in-plane lattice constant a and c/a ratio for the hcp phase, and the nearest-neighbor distance $a_{\rm NN}$ for the hcp, bcc, and fcc phases. (All lengths are in atomic units.) The bulk moduli (B, in Mbar) and cohesive energies (E_c, in eV/atom) for Mg in the hcp, bcc, and fcc phases are also given. The experimental values (extrapolated to T=0 K) are given in parentheses (Ref. 56).

	hcp	bcc	fcc
a	5.88 (6.02)		
c/a	1.62 (1.623)		
$a_{\rm NN}$	5.85	5.70	5.88
B	0.40 (0.35)	0.39	0.39
E_c	1.80 (1.51)	1.77	1.79

E. Vacancy formation energy of Mg

The vacancy formation energy of Mg in the hcp structure was determined at a fixed lattice constant; based on the results for the Al vacancy, we expect these calculated heats of formation to differ from the fully volume-relaxed heats by on the order of 0.01–0.03 eV. We first considered a 36-atom cell with 10 ($N_{\rm tot}$=3600) and 36 ($N_{\rm tot}$=23328) special **k** points for the bulk and the vacancy problem, computing both the bulk and perturbed systems in the same structure and with the same sampling of the BZ to minimize errors. To investigate the finite size effects, we then created the vacancy in a 96-atom cell with a sampling of 6 special points ($N_{\rm tot}$=9216) in the BZ.

Our results for the ideal and fully relaxed vacancies are tabulated in Table V. We note that the result of 0.81 eV for the relaxed vacancy formation energy for the smaller unit cell with the higher sampling of the BZ compares closely with the result of 0.83 eV for the larger system cell. These results indicate that one can get a reasonably good estimate of the vacancy formation energy by studying a relatively small atom cell accurately; this is consistent with our earlier observation for the case of the Al vacancy.

The difference between the results for the smaller and larger unit cell might point to the relative attractive nature of the vacancy-vacancy interaction in Mg: The distance between the vacancies in the 36-atom cell is 17.6 Bohr, whereas that in the 96-atom cell is 23.5 Bohr; the reduced heat of formation at the shorter distance may indicate that the vacancies have a tendency to attract each other at zero temperature at this distance. It is possible that at a certain critical density of vacancies, clustering will be favored over single isolated vacancies since it will require less energy to create more internal surface in expanding a void in magnesium than to create a single isolated vacancy, which would require breaking the maximum number of 12 nearest-neighbor bonds.

The relaxation energies are small, reducing the heat of formation of the vacancy by ~1.5%. In Fig. 3 we show the vector displacement of the atoms around the fully relaxed vacancy for the 96-atom cell. The atoms displace in an obvious way: The three nearest atoms in the layer above and the layer below the vacancy move essentially radially toward the vacancy a fractional distance of d_1=1.1% (measured relative to the nearest-neighbor distance). The six nearest atoms in the hexagonal plane containing the vacancy are displaced in the plane a fractional distance of d_2=0.7% toward the hollow site with an angular displacement of θ~12° to the radial direction. The atoms in the second nearest-neighbor shell are displaced by d_3=0.2% and move (approximately) radially toward the vacancy; the atoms in the third layer immediately above and below the vacancy move by d_4=0.6% also radially toward the vacancy. The displacement of all other atoms is less than 0.2% of the nearest-neighbor distance.

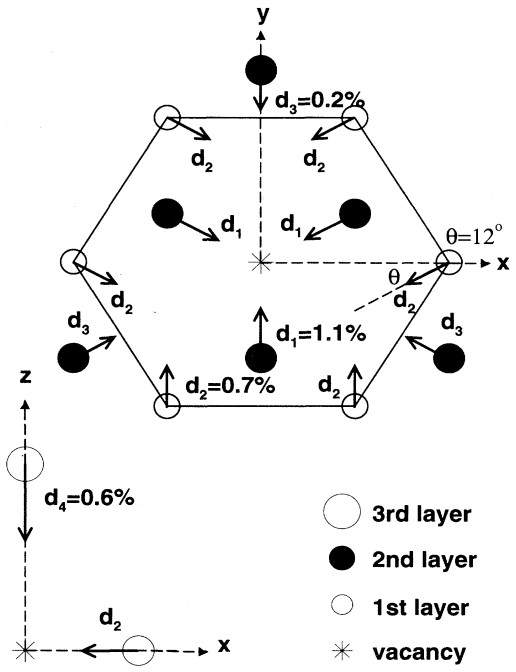

FIG. 3. The vector displacement of the atoms around the fully relaxed Mg vacancy for the 96-atom cell.

TABLE V. Vacancy formation energy (in eV) for hcp Mg for various unit cell sizes and number of special **k** points ($N_{\mathbf{k}}$).

	36 atoms		96 atoms	Expt.
$N_{\mathbf{k}}$	10	36	6	
unrelaxed	0.93	0.83	0.84	
relaxed	0.91	0.81	0.83	0.81[a]
				0.58[b]
				0.79[c]
				0.90[d]

[a]Reference 60.
[b]Reference 58.
[c]Reference 59.
[d]Reference 63.

With the loss of coordination caused by the creation of the vacancy, there is a clear tendency for the atoms in the immediate vicinity of the vacancy to move radially toward each other to gain electronic charge density.[57] As in Al, the atoms farther away from the first neighbor shell surrounding the vacancy relax in a way to maintain optimal bulk coordination. For example, the average change in the nearest-neighbor distance of an atom in the second shell surrounding the vacancy is 0.6% of the nearest-neighbor distance, whereas the minimum displacement of an atom in this shell from its ideal hcp site is 0.7%. There is a partial dimerization of the atoms in the hexagonal plane surrounding the vacancy: The atoms relax toward the void which is the hollow site, thus compensating for the loss of coordination.

There are three types of experimental methods that have been used to measure the vacancy formation energy of Mg and one of these results is in disagreement with the

other two. A dilatometric-diffractometric measurement[58] gave a value of 0.58±0.01 eV, whereas resistivity-after-quench experiments[59] yielded a result of 0.79±0.03 eV, with slightly larger values reported by others depending on the conditions of the quench: 0.81 eV from electrical resistivity measurements[60] at thermal equilibrium or 0.83–0.89 eV after quenching.[61,62] The final class of experiment used positron annihilation techniques[63] and deduced a value of 0.9±0.1 eV.

Our theoretical results clearly are in close agreement with the last two classes of experimental results and are consistent with the claim that the vacancy formation energy of Mg is greater than that of Al.[59] While the investigation of the first group appears to be thorough, including a critical analysis of various possible sources of error, we believe that these results are probably less reliable because the dilatometric and the diffractometric measurements were actually performed on different samples with possible different dislocation and impurity distributions.

F. Si and Al impurities in Mg

The heats of solution of Si and Al impurities in hcp Mg were also computed. Once again we considered a 36-atom cell with 10 and 36 special **k** points, and then a 96-atom cell with 6 special points. For the smaller system size, the impurity problem may be considered equivalent to a low concentration alloy at the ~3% level and, for the larger cell, an alloy at the ~1% level.

In Table VI, we list the heat of solution for the Si impurity and the Al impurity in Mg. For the constitutent phases, we choose the cubic diamond and fcc structures for Si and Al, respectively. Surprisingly, the relaxations are more substantial in the impurity cases than in the correponding vacancy problem: For example, in the case of the Al impurity in Mg, the three nearest atoms above and below the plane containing the impurity are displaced d_1=1.4% radially toward the impurity, and the six nearest atoms in the hexagonal plane containing the impurity are also displaced d_2=1.4%, with an angular displacement of θ~15° toward the hollow site. The displacement of all other atoms is less than 0.4%, and the relaxation energy is ~40 meV, which is about 4 times that for the vacancy. This accounts for the fact that there is a larger discrepancy in the result for the impurity problems than the vacancy problems when comparing the large unit cell calculation with the smaller cell, 36 **k**-point result: The extended nature of the relaxations in the impurity cases, especially the Si impurity, calls for an even more detailed sampling of the Brillouin zone than what we have been able to afford.

TABLE VI. Heats of solution (in eV) for Si and Al impurities in hcp Mg for various unit cell sizes and number of special **k** points ($N_{\mathbf{k}}$).

	36 atoms		96 atoms
$N_{\mathbf{k}}$	10	36	6
Si impurity			
unrelaxed	0.32	0.35	0.29
relaxed	0.27	0.28	0.19
Al impurity			
unrelaxed	0.08	0.11	0.09
relaxed	0.05	0.08	0.05

The metallic radii of Si (1.32 bohr) and of Al (1.43 bohr) are smaller than that of Mg (1.60 bohr), and the larger size and electronic mismatch are between Si and Mg rather than between Al and Mg. Our results reflect this difference in the larger heat of solution and relaxation energies of the Si impurity system compared with the Al case. There are, unfortunately, less experimental data for impurities in Mg than in Al with which to compare. Two different indirect determinations of the heat of solution of Al in Mg differ even in sign: 0.07 eV (Ref. 64) and −0.11 eV (Ref. 65). Our results clearly favor the positive heat and also agree in magnitude.

G. Impurity-induced density of states

The discussion up to now has been related to the structural properties and energetics. Additional information, however, can be obtained by looking at the electronic properties such as the local density of states directly. In Fig. 4(a), the local densities of states (LDOS) at the defect site for the Al calculations are given, along with the

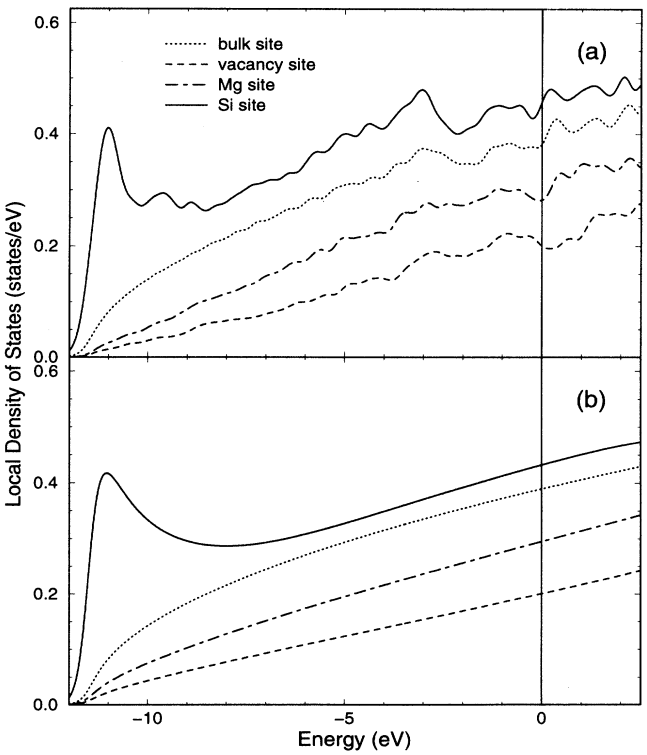

FIG. 4. (a) Calculated and (b) Clogston-Wolff model local densities of states for bulk Al and the Mg, Si, and vacancy sites in Al.

LDOS at a bulk site for reference. As expected, the density of states below the Fermi level is increased for Si and decreased for Mg and the vacancy. The vacancy site still has significant charge (~1.2 electrons), providing further evidence of the delocalized nature of the electrons in Al. The Si LDOS shown are similar to those calculated[66] previously using a Green's function technique.

In order to understand the changes in the LDOS, we consider a simple model. From the calculations of fcc Si and Mg at fixed lattice constant (cf. Fig. 2), we know that a rigid band model describes the density of states of fcc Si (and Mg) reasonably well, although the LDOS of states shown in Fig. 4(a) of course do not resemble the bulk Al LDOS. To describe the effect of the defect, we consider a Clogston-Wolff[67] model. In this model, the one-particle Green's function for the perturbed system is related to the unperturbed system by the Dyson equation

$$G = G_0 + G_0 V_I G\,, \tag{12}$$

where V_I is the impurity strength and is assumed spherically symmetric. The Green's function (for cubic systems) can be decomposed into ℓ for $\ell \leq 2$. Since the imaginary part of the Green's function is simply proportional to the LDOS $n(\varepsilon)$,

$$G(\varepsilon) = H(\varepsilon) - i\pi n(\varepsilon)\,, \tag{13}$$

and the real part of the Green's function $H(\varepsilon)$ is related by a Kramers-Kronig relationship (Hilbert transform) to $n(\varepsilon)$,

$$H(\varepsilon) = P\int_{-\infty}^{+\infty} d\omega\, \frac{n(\omega)}{\varepsilon - \omega}\,, \tag{14}$$

the perturbed density of states can be obtained from the unperturbed LDOS: The perturbed LDOS, for each ℓ separately, is given by

$$n(\varepsilon) = \frac{n_0(\varepsilon)}{(1 - H_0 V)^2 + (\pi V n_0)^2}\,. \tag{15}$$

The strength of the impurity potential for each ℓ depends on the total number of electrons in that band and is related to V_I by $V = V_I/(2\ell + 1)$.

To apply this model to the impurities in Al, we consider a free-electron density of states scaled to have the same occupied bandwidth and number of electrons as Al. This LDOS is further broken up into ℓ contributions based on a simple plane wave decomposition. The impurity strength parameter V_I is determined from the shift in Fermi level of the free-electron DOS required to obtain the correct number of electrons (2 for Mg, 4 for Si, 1.25 for the vacancy). The resulting LDOS for this model are given in Fig. 4(b). Overall the model reproduces the LDOS very well, including the magnitude. This agreement is all the more remarkable considering that there are no arbitrary adjustable parameters. The difference in shape between the LDOS of Si and Mg (and the vacancy) results from the difference in sign of the impurity potential; for Si it is attractive (necessary to bind an extra electron), while for Mg and the vacancy, it is repulsive. The peak at the bottom of the band in the case of Si is mainly s like in character and demonstrates that even in a free-electron metal, electrons of different ℓ are affected differently. For the repulsive impurity potentials, there is a suppression of the density of states that changes not only the overall magnitude of the LDOS, but also the shape near the bottom of the band.

These modifications of the electronic structure are ultimately what give rise to the positive heats of solution for the these impurities. Although we have not carried out the analogous calculations for impurities in Mg, similar results are expected, except that now Al, Si, and the vacancy all will have repulsive interactions. The results of this section demonstrate that the significant changes in the electronic structure can be understood within a simple model, but a model that goes beyond the simple elastic (size effect) models often used to describe the impurity heats of solution.[53]

IV. SUMMARY AND CONCLUSIONS

We have presented our computational scheme for solving the self-consistent Kohn-Sham equations using the plane wave pseudopotential technique. The algorithm is based on a preconditioned steepest descent method and a subspace diagonalization. This last step is important beyond simply orthogonalizing the states and obtaining the occupation numbers; in our experience, it also provides a method that avoids the computational instabilities that have been observed by others. By using a properly variational total energy expression, including variable occupation numbers, consistent total energies and forces are obtained for large cells. Combined with an efficient quasi-Newton relaxation scheme, the forces are used to fully relax the internal structural parameters in a small number of atomic steps.

For the metallic systems studied here, adequate sampling of the Brillouin zone was found to be crucial. For example, the vacancy formation energy of Al calculated with just Γ-point sampling for a 108-atom cell produces a result of the wrong sign. On the other hand, with careful **k**-point sampling, effective defect-defect interactions appear to be rather short ranged and the heats of solution and formation are already quite well described by unit cells of ~30 atoms. If one were to use Γ-point sampling of the zone only, at least an order of magnitude more atoms per unit cell would be required to get comparable accuracy.

The calculated properties of the vacancies and simple impurities are in generally good agreement with the available experimentally derived results; where conflicting experimental values exist, the calculations distinguish between the different values. In all cases considered, the heats of solution of impurities and vacancy formation are positive, including the low concentration limits of the Al-Mg system, a system which is known to form complex ordered alloys. The relaxations about the impurities and vacancies show trends that follow rather simple arguments based on size effects of the various defects. Although size effects are important, the dominant con-

tribution to the heats appears to be electronic in nature, rather than elastic. While this separation into electronic and elastic contributions is somewhat arbitrary since the elastic properties are determined by the electrons, explanations based on differences in bonding have the virtue of providing simple local pictures and explaining the small relaxation energies found.

The results and convergence studies presented here provide benchmarks and tests of the abilitiy of first-principles electronic structure theory to accurately describe the defects and impurities of these metallic systems. The results are encouraging and suggest that careful calculations can be used not only to model the properties of defects, impurities, and alloys at the 1% level for real materials, but also to obtain physical insight on a microscopic level.

ACKNOWLEDGMENTS

We thank C. Leung, G. Castilla, F. Manzo, R. B. Marr, J. Pasciak, and R. F. Peierls for useful discussions and B. Liu for pointing out Refs. 64 and 65. N.C. would like to thank E. McFadden for help with the computational resources. This work was supported by the Division of Materials Sciences, U.S. Department of Energy, under Contract No. DE-AC02-76CH00016, by the U.S. National Science Foundation under Grant No. DMR-91-20440, by the High Performance Computing and Communications Program of the U.S. Department of Energy through the Partnership in Computational Sciences, and by a grant of computer time at the National Energy Research Supercomputer Center, Livermore, CA.

APPENDIX A: CONVERGENCE OF PRECONDITIONED STEEPEST DESCENT METHOD

We are interested in computing the eigenstates of the Hamiltonian H,

$$H|\xi_i\rangle = \varepsilon_i|\xi_i\rangle\,. \tag{A1}$$

Our algorithm for iteratively solving this set of equations closely resembles the first time derivative or steepest descent method,[16,27]

$$\frac{\partial}{\partial t}|\psi_i\rangle = (\lambda_i - H)|\psi_i\rangle, \tag{A2}$$

where the stationary solutions of this equation are the eigenstates of H. The Rayleigh quotient

$$\lambda_i = \frac{\langle\psi_i|H|\psi_i\rangle}{\langle\psi_i|\psi_i\rangle} \tag{A3}$$

is the expectation value of the energy in the state $|\psi_i\rangle$. It is important to emphasize that the "time" t in these equations is not a proper time—this is *not* the time-dependent Schrödinger equation—but simply an auxiliary parameter introduced to label the iterative steps of the algorithm. There is no simple analytical solution to Eq. (A2), but we can determine the convergence properties of this scheme by expanding the state $|\psi_i\rangle$ in terms of the eigenstates of H,

$$|\psi_i\rangle = \sum_\alpha a_\alpha|\xi_\alpha\rangle\,. \tag{A4}$$

Substituting into Eq. (A2) gives the time evolution of the coefficients

$$\frac{\partial}{\partial t}a_\alpha = a_\alpha(\lambda_i - \varepsilon_\alpha)\,. \tag{A5}$$

It follows quite simply that

$$\frac{|a_\alpha^t|}{|a_\beta^t|} = \frac{|a_\alpha^0|}{|a_\beta^0|}\exp[(\varepsilon_\beta - \varepsilon_\alpha)t]\,, \tag{A6}$$

where the superscripts t and 0 refer to the corresponding iteration steps. Thus the coefficient of the lowest-energy state will be the most rapidly growing term. Keeping in mind that the state has to remain normalized, the action of the equation of motion is to simply project out the lowest-energy eigenstate present in $|\psi_i\rangle$. If the initial state has a nonzero overlap with the lowest-energy state in the entire system, then this equation will simply project out that state.

One could use the above procedure to compute the required eigenstates iteratively: First one would compute the lowest eigenstate; then the next to lowest eigenstate would be computed by generating a starting iterate which is orthogonal to the lowest, and so forth. This procedure is not a desirable way to solve the problem.

Instead, we would like to propagate all of the states at the same time rather than project out the states one at a time. This is because in the electronic structure problem, the Hamiltonian is a complicated function of the occupied states and has to be updated during the evolution of the states. In addition, simultaneous eigenstate updates are natural for implementation on large scale parallel computing environments. However, simultaneous iteration gives rise to problems of maintaining orthogonality between the trial states which is necessary to prevent the states from all descending to the ground state. The equation of motion [Eq. (A2)] by itself does not guarantee the orthogonality of the states for continously varying time.

To deal with these problems, a finite difference approximation to Eq. (A2) for small times Δt is used:

$$|\psi_i^{t+1}\rangle = |\psi_i^t\rangle + \Delta t\,(\lambda_i^t - H)|\psi_i^t\rangle\,, \tag{A7}$$

where t now labels time in units of Δt. The states are propagated for time Δt by this equation, and are then orthogonalized by some procedure. These steps are repeated until the states become stationary. This is the "steepest descent" (SD) method since the change in the state is orthogonal to the state itself, and in the direction of steepest descent of the expectation value.

In practice, H is updated during each iterative time step while the atomic positions remain fixed. It is now left to determine the optimal size of the time step Δt for

which the equations can be integrated stably. Obviously, the larger the time step, the quicker the convergence, but the greater the deviation from orthonormality. Also, for large system sizes, the time step should be sufficiently small to damp out charge oscillations due to the long-range nature of the Coulomb interaction.

For small times Δt, Eq. (A5) becomes

$$a_\alpha^{t+1} = a_\alpha^t[1 + \Delta t\,(\lambda_i^t - \varepsilon_\alpha)]\,, \tag{A8}$$

which shows that the time evolution of each coefficient a_α is governed by a characteristic frequency $\omega_\alpha = (\lambda_i - \varepsilon_\alpha)$. Since $\varepsilon_{\max} \geq \lambda_i^t \geq \varepsilon_0$, choosing

$$\Delta t \leq \frac{2}{(\varepsilon_{\max} - \varepsilon_0)} \tag{A9}$$

guarantees convergence to the lowest-energy state with a convergence rate of $1 + \Delta t\,(\varepsilon_0 - \varepsilon_1)$. Here, the subscripts 0, 1, and "max" correspond to the lowest, next-to-lowest, and maximum eigenvalues, respectively.

Thus the time step in the iterative solution of the KS equations is set by the inverse of the width of the band spectrum of the Hamiltonian. This result places a limit on the type of the system that one can realistically study using this method. For example, for a transition metal system which needs a high cutoff for the plane wave expansion of the wave functions, the corresponding large bandwidth results in a time step that might make it impractical to study large systems comprising such atoms.

Now consider a preconditioned steepest descent (PSD) method. If we replace Δt by $\Delta t\,(H - \varepsilon_0^*)^{-1}$ in Eq. (A7), we get the iteration

$$|\psi_i^{t+1}\rangle = |\psi_i^t\rangle + \Delta t\,(H - \varepsilon_0^*)^{-1}(\lambda_i^t - H)|\psi_i^t\rangle\,, \tag{A10}$$

where $\varepsilon_0^* \leq \varepsilon_0$ ensures that the preconditioner is positive definite. This leads to the following equation for the time evolution of the eigenstate coefficients:

$$a_\alpha^{t+1} = a_\alpha^t \left[1 + \Delta t\,\frac{(\lambda_i^t - \varepsilon_\alpha)}{(\varepsilon_\alpha - \varepsilon_0^*)}\right]. \tag{A11}$$

This modified algorithm will converge to the lowest-energy state if

$$\Delta t \leq 2. \tag{A12}$$

Thus the inverse of the Hamiltonian, shifted appropriately to ensure a positive definite operator, scales the high-frequency eigencomponents so that it is possible to use a large time step and—in contrast to the SD method—does not depend on the width of the spectrum of H. The choice of $\Delta t = 1$ corresponds to the well-known inverse power method.

The problem with Eq. (A10) is that it is expensive to compute the action of the inverse Hamiltonian. Thus we are led to choose an approximate form for this operation,

$$|\psi_i^{t+1}\rangle = |\psi_i^t\rangle + B\,(\lambda_i^t - H)|\psi_i^t\rangle\,. \tag{A13}$$

Here B is a symmetric and positive definite operator which should approximate the inverse of H. We have absorbed the time step Δt into the definition of B, and we will scale B such that the spectral radius γ of $I - BH$ is less than 1, where I denotes the identity operator. The state $|\psi_i\rangle$ converges to the lowest-energy state $|\xi_0\rangle$ at an asymptotic rate of

$$\frac{\varepsilon_0}{\varepsilon_1} + \gamma\left(1 - \frac{\varepsilon_0}{\varepsilon_1}\right). \tag{A14}$$

A rapid rate of convergence is achieved if γ is bounded away from 1.

APPENDIX B: IMPLEMENTATION OF PRECONDITIONED STEEPEST DESCENT METHOD

To solve the electronic structure problem, we first expand the wave functions in a basis of plane waves,

$$\langle \mathbf{r}|\psi_{\mathbf{k}n}\rangle = \psi_{\mathbf{k}n}(\mathbf{r}) = \sum_{\mathbf{g}} \exp[i(\mathbf{k}+\mathbf{g})\cdot\mathbf{r}]\, C_{\mathbf{k}+\mathbf{g},n}\,, \tag{B1}$$

where $\mathbf{k}$ is the wave vector in the BZ, n is the band index, and $\mathbf{g}$ is a reciprocal lattice vector. The Hamiltonian matrix within this basis is given by ($\hbar$=m=e=1)

$$H_{\mathbf{k}+\mathbf{g},\mathbf{k}+\mathbf{g}'} = \frac{1}{2}(\mathbf{k}+\mathbf{g})^2\delta_{\mathbf{g}\mathbf{g}'} + V_{\rm loc}(\mathbf{g}-\mathbf{g}') + V_{\rm NL}(\mathbf{k}+\mathbf{g},\mathbf{k}+\mathbf{g}')\,, \tag{B2}$$

where $V_{\rm loc}$ comprises the local pseudopotential, the Hartree potential, and the exchange and correlation potential. $V_{\rm NL}$ is the nonlocal Kleinman-Bylander separable pseudopotential.

Equation (A7) now becomes ($\mathbf{G} \equiv \mathbf{k} + \mathbf{g}$)

$$C_{\mathbf{G},n}^{t+1} = C_{\mathbf{G},n}^t + \Delta t \sum_{\mathbf{g}'} (\lambda_{\mathbf{k}n}\delta_{\mathbf{g}\mathbf{g}'} - H_{\mathbf{G},\mathbf{G}'})C_{\mathbf{G}',n}^t$$

$$= C_{\mathbf{G},n}^t - \Delta t\{\tfrac{1}{2}|\mathbf{G}|^2 + V_{\rm loc}(\mathbf{0}) + V_{\rm NL}(\mathbf{G},\mathbf{G}) - \lambda_{\mathbf{k}n}\}C_{\mathbf{G},n}^t - \Delta t \sum_{\mathbf{g}'\neq\mathbf{g}} \{V_{\rm loc}(\mathbf{g}-\mathbf{g}') + V_{\rm NL}(\mathbf{G},\mathbf{G}')\}C_{\mathbf{G},n}^t$$

$$\equiv (1 - \Delta t\,\omega_{\mathbf{G},n})C_{\mathbf{G},n}^t - \Delta t\, W_{\mathbf{G},n}\,. \tag{B3}$$

Keeping in mind that the above equation only makes sense for small times Δt, we assume that the time rate of change of $C_{\mathbf{G},n}$ is largely due only to the coefficient itself, namely, the first term in Eq. (B3), and that we can ignore the time dependence of the second term $W_{\mathbf{G},n}$. Then each Fourier component of the wave function has its own characteristic frequency $\omega_{\mathbf{G},n}$. Since the kinetic energy plays an increasingly more dominant role in the time evolution of the high-energy components, and since it is precisely these components that are responsible for the reduced time step in the SD method, we use a preconditioning matrix B that satisfies the constraint of being positive definite of the form

$$B_{\mathbf{g}\mathbf{g}'} = \{\tfrac{1}{2}|\mathbf{G}|^2 + S\}^{-1}\delta_{\mathbf{g}\mathbf{g}'}, \tag{B4}$$

where S is a constant shift and may be chosen appropriately. This choice of a point-Jacobi preconditioner correctly scales the high-kinetic-energy components, and allows in effect for a variable time step $\Delta t \sim (\omega_{\mathbf{G},n})^{-1}$ for each Fourier component of the wave function.

Equation (A13) now reads

$$C^{t+1}_{\mathbf{G},n} = C^{t}_{\mathbf{G},n} + \frac{\sum_{\mathbf{g}'}(\lambda_{\mathbf{k}n}\delta_{\mathbf{g}\mathbf{g}'} - H_{\mathbf{G},\mathbf{G}'})C^{t}_{\mathbf{G}',n}}{\frac{1}{2}|\mathbf{G}|^2 + S}. \tag{B5}$$

The result is a quicker convergence for the lower-energy components of the wave function since the maximal time step is used here, and a slower convergence for the higher-energy components. The latter terms are less important since, if the cutoff energy is high enough, these terms are necessarily small. This preconditioned matrix has the effect of making those states that are dominated by the kinetic energy to be approximately degenerate which has the tendency of making the error in the wave function proportional to the steepest descent vector.[28]

A better understanding of the properties of the equation of motion can be obtained from a real space representation. Substituting Eq. (B5) into (B1) gives

$$\psi^{t+1}_{\mathbf{k}n}(\mathbf{r}) = \psi^{t}_{\mathbf{k}n}(\mathbf{r}) + \int d\mathbf{r}'\, B(\mathbf{r}-\mathbf{r}')\,[\lambda_{\mathbf{k}n} - H]\psi^{t}_{\mathbf{k}n}(\mathbf{r}'), \tag{B6}$$

where

$$B(\mathbf{r}) = \sum_{\mathbf{R}} \frac{\exp[-i\mathbf{k}\cdot(\mathbf{r}-\mathbf{R})]}{2\pi}\,\frac{\exp[-\sqrt{2S}\,|\mathbf{r}-\mathbf{R}|\,]}{|\mathbf{r}-\mathbf{R}|}, \tag{B7}$$

and $\mathbf{R}$ is a lattice vector. Here it is seen explicitly that contributions to the wave function in the far-field limit are suppressed because of the presence of the Yukawa-like term in the update. This reduces the transfer of charge over long distances which can be an artifact of the numerical solution and is especially problematic in systems of large physical dimensions—the phenomenon of "charge sloshing."

[1] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).
[2] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).
[3] E.g., see *Theory of the Inhomogeneous Electron Gas*, edited by S. Lundqvist and N. M. March (Plenum, New York, 1983).
[4] R. Podloucky, R. Zeller, and P. H. Dederichs, Phys. Rev. B **22**, 5777 (1980).
[5] G. Galli and A. Pasquarello, *Computer Simulation in Chemical Physics*, edited by M. P. Allen and D. J. Tildesley (Kluwer Academic Publishers, Dordrecht, 1993).
[6] G. Kresse and J. Hafner, Phys. Rev. B **47**, 558 (1993).
[7] G.-X. Qian, M. Weinert, G. W. Fernando, and J. W. Davenport, Phys. Rev. Lett. **64**, 1146 (1990).
[8] X.-P. Li, P. B. Allen, R. Car, M. Parrinello, and J. Q. Broughton, Phys. Rev. B **41**, 3260 (1990).
[9] J. L. Martins and R. Wentzcovitch, Solid State Commun. **78**, 831 (1991).
[10] G. Galli, R. M. Martin, R. Car, and M. Parrinello, Phys. Rev. Lett. **62**, 555 (1989).
[11] I. Stich, R. Car, and M. Parrinello, Phys. Rev. Lett. **63**, 2240 (1989).
[12] D. M. Bylander and L. Kleinman, Phys. Rev. B **45**, 9663 (1992).
[13] G. W. Fernando, G. X. Qian, M. Weinert, and J. W. Davenport, Phys. Rev. B **40**, 7985 (1989).
[14] M. Weinert and J. W. Davenport, Phys. Rev. B **45**, 13 709 (1992).
[15] M. P. Grumbach, D. Hohl, R. M. Martin, and R. Car, J. Phys. Condens. Matter **6**, 1999 (1994).
[16] C. Woodward, B. I. Min, R. Benedek, and J. Garner, Phys. Rev. B **39**, 4853 (1989).
[17] M. R. Pederson and K. A. Jackson, Phys. Rev. B **43**, 7312 (1991).
[18] Y. Yamamoto and T. Fujiwara, Phys. Rev. B **46**, 13 596 (1992).
[19] S. Narasimhan and J. W. Davenport, Phys. Rev. B **51**, 659 (1995).
[20] D. R. Hamann, M. Schlüter, and C. Chiang, Phys. Rev. Lett. **43**, 1494 (1974); G. B. Bachelet, D. R. Hamann, and M. Schlüter, Phys. Rev. B **26**, 4199 (1982).
[21] D. Vanderbilt, Phys. Rev. B **41**, 7892 (1990).
[22] N. Troullier and J. L. Martins, Phys. Rev. B **43**, 1993 (1991).
[23] A. Rappe, K. M. Rabe, E. Kaxiras, and J. D. Joannopoulos, Phys. Rev. B **41**, 1227 (1990).
[24] E. L. Shirley, D. C. Allan, R. M. Martin, and J. D. Joannopoulos, Phys. Rev. B **40**, 3652 (1989).
[25] W. E. Pickett, Comput. Phys. Rep. **9**, 115 (1989).
[26] R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).
[27] J. M. Soler and A. R. Williams, Bull. Am. Phys. Soc. **32**, 562 (1987).
[28] M. P. Teter, M. C. Payne, and D. C. Allan, Phys. Rev. B **40**, 12 255 (1989).
[29] J. Broughton and F. Khan, Phys. Rev. B **40**, 12 098 (1989).
[30] J. Ihm, A. Zunger, and M. L. Cohen, J. Phys. C **12**, 4409 (1979).
[31] S. H. Vosko, L. Wilk, and M. Nusair, Can. J. Phys. **58**, 1200 (1980).
[32] L. Kleinman and D. M. Bylander, Phys. Rev. Lett. **48**, 1425 (1982).

[33] X. Gonze, P. Kaeckell, and M. Scheffler, Phys. Rev. B **41**, 12 264 (1990).
[34] H. J. Monkhorst and J. D. Pack, Phys. Rev. B **13**, 5188 (1976).
[35] M. Needels, A. M. Rappe, P. D. Bristowe, and J. D. Joannopoulos, Phys. Rev. B **46**, 9768 (1992).
[36] A. F. Wright and S. R. Atlas, Phys. Rev. B **50**, 15 248 (1994).
[37] N. D. Mermin, Phys. Rev. **137**, A1441 (1965).
[38] R. P. Feynman, Phys. Rev. **56**, 340 (1939); H. Hellmann, *Einfuhrung in die Quantumchemie* (Deuticke, Leipzig, 1937).
[39] D. F. Shano, Math. Oper. Res. **3**, 244 (1978).
[40] S. S. Oren and E. Spedicato, Math. Programming **10**, 70 (1976).
[41] P. Minchin, A. Meyer, and W. H. Young, J. Phys. F **4**, 2117 (1974).
[42] M. W. Finnis and M. Sachdev, J. Phys. F **6**, 965 (1976).
[43] M. J. Stott, J. Nucl. Mater. **69-70**, 157 (1978).
[44] B. Chakraborty and R. W. Siegel, Phys. Rev. B **27**, 4535 (1983).
[45] M. J. Gillan, J. Phys. Condens. Matter **1**, 689 (1989).
[46] R. W. Jansen and B. M. Klein, J. Phys. Condens. Matter **1**, 8359 (1989).
[47] M. W. Finnis, J. Phys. Condens. Matter **2**, 331 (1990).
[48] A. De Vita and M. J. Gillan, J. Phys. Condens. Matter **3**, 6225 (1991).
[49] M. J. Mehl and B. M. Klein, Physica B **172**, 211 (1991).
[50] R. Benedek, L. H. Yang, C. Woodward, and B. I. Min, Phys. Rev. B **45**, 2607 (1992).
[51] H. M. Polatoglou, M. Methfessel, and M. Scheffler, Phys. Rev. B **48**, 1877 (1993).
[52] P. Ehrhart, P. Jung, H. Schulta, and H. Ullmaier, in *Atomic Defects in Metals,* edited by H. Ullmaier, Landolt-Börnstein, New Series, Group III, Vol. 25 (Springer-Verlag, Berlin, 1990).
[53] J. E. Hatch, *Aluminum: Properties and Physical Metallurgy* (American Society for Metals, Metals Park, OH, 1984).
[54] R. M. Wentzcovitch and M. L. Cohen, Phys. Rev. B **37**, 5571 (1988); J. D. Altott, P. B. Allen, R. M. Wentzovitch and J. A. Moriarty, *ibid.* **48**, 13 253 (1993).
[55] M. Y. Chou and M. L. Cohen, Solid State Commun. **57**, 785 (1986).
[56] C. Kittel, *Introduction to Solid State Physics,* 5th ed. (Wiley, New York, 1976).
[57] K. W. Jacobsen, J. K. Nørskov, and M. J. Puska, Phys. Rev. B **35**, 7423 (1987).
[58] C. Janot, D. Malléjac, and B. Georg, Phys. Rev. B **2**, 3088 (1970).
[59] P. Tzanetakis, J. Hillairet, and G. Revel, Phys. Status Solidi B **75**, 433 (1976).
[60] C. Mairy, J. Hillairet, and D. Schumacher, Acta. Metall. **15**, 1258 (1967).
[61] J. Hillairet, C. Mairy, J. Espinasse, and V. Levy, Acta Metall. **18**, 1285 (1970).
[62] C. J. Beevers, Acta Metall. **11**, 1029 (1963).
[63] A. Vehanen and K. Rytsölä, in *Proceedings of the International School of Physics*, edited by W. Brandt (North-Holland, Utrecht, 1981).
[64] B. Predel and K. Hulse, Z. Metallkd. **69**, 661 (1978).
[65] D. Luedecke and K. Hack, Z. Metallkd. **77**, 145 (1986); J. Brown and J. Pratt, Met. Trans. **1**, 2743 (1970).
[66] J. Deutz, P. H. Dederichs, and R. Zeller, J. Phys. F **11**, 1787 (1981).
[67] A. M. Clogston, Phys. Rev. **125**, 439 (1962); P. A. Wolff, *ibid.* **124**, 1030 (1961).
[68] K. G. Lynn and P. J Schultz, Appl. Phys. A **37**, 31 (1985).
[69] R. M. Emrick and P. B. McArdle, Phys. Rev. **188**, 1156 (1964).