UCRL-PROC-217725

Introduction to First-Principles Electronic Structure Methods: Application to Actinide Materials

John E. Klepeis

December 14, 2005

2005 MRS Fall Meeting
Boston, MA, United States
November 28, 2005 through December 2, 2005

Introduction to First-Principles Electronic Structure Methods: Application to Actinide Materials

John E. Klepeis
Lawrence Livermore National Laboratory
Livermore, CA 94551

ABSTRACT

The purpose of this paper is to provide an introduction for non-experts to first-principles electronic structure methods that are widely used in the field of condensed-matter physics, including applications to actinide materials. The methods I describe are based on density functional theory (DFT) within the local density approximation (LDA) and the generalized gradient approximation (GGA). In addition to explaining the meaning of this terminology I also describe the underlying theory itself in some detail in order to enable a better understanding of the relative strengths and weaknesses of the methods. I briefly mention some particular numerical implementations of DFT, including the linear muffin-tin orbital (LMTO), linear augmented plane wave (LAPW), and pseudopotential methods, as well as general methodologies that go beyond DFT and specifically address some of the weaknesses of the theory. The last third of the paper is devoted to a few selected applications that illustrate the ideas discussed in the first two-thirds. In particular, I conclude by addressing the current controversy regarding magnetic DFT calculations for actinide materials. Throughout this paper particular emphasis is placed on providing the appropriate background to enable the non-expert to gain a better appreciation of the application of first-principles electronic structure methods to the study of actinide and other materials.

INTRODUCTION

First-principles electronic structure methods are quantum mechanical methods for numerically solving the Schrödinger (nonrelativistic) or Dirac (relativistic) equation for systems of electrons

$$H\Psi(r) = E\Psi(r) \tag{1}$$

Equation (1) is simply the time-independent Schrödinger equation. The term "first-principles" means that there is no empirical fitting, or equivalently, no adjustable parameters. The term "*ab initio*" is sometimes used instead and means the same thing. This aspect of the methodology is important in order for it to be predictive. In other words, when applied to a new system it is not necessary to make any adjustments but rather there is an expectation that the method will be applicable. Another important quality that makes first-principles methods predictive is that they are accurate. In the present context that means they have been well tested against experiment. Naturally these methods do have limitations and I will discuss some of them in this paper. Another aspect, which could also be viewed as a limitation, is that first-principles methods are very computationally expensive. In condensed-matter physics the term first-principles usually refers to a method that is based on density functional theory (DFT). The first two-thirds of this paper will be devoted to a somewhat technical overview of DFT. I believe that it is important to understand a few technical details in order to better appreciate the application of these methods. Most of the material that I will discuss in this paper can found in much greater detail in a recent book written by Richard Martin [1].

DENSITY FUNCTIONAL THEORY (DFT)

In 1964 Hohenberg and Kohn[2] proved that the exact ground state energy for a system of interacting electrons is a universal functional of the electron density (a functional is simply a function of another function)

$$E_{\mathrm{GS}} = E[\rho(r)] \tag{2}$$

This theorem is the basis of the name density functional theory—a theory concerning a functional of the density. Note that the energy and the density are the fundamental quantities of the theory. This theorem is important for two reasons. Firstly, it says that all of the complexities of the full many-body quantum-mechanical problem are captured by a universal functional. Secondly, once we have the electron density we can then obtain all of the ground state properties of the system. However, there is also an important limitation because the theorem is only an existence proof since the universal functional in equation (2) is actually unknown.

A year later in 1965 Kohn and Sham[3] provided the means for turning DFT into a practical method. They assumed that the electron density could be written in terms of single-particle wavefunctions $\phi_k(r)$ of an auxiliary non-interacting system

$$\rho(r) = \sum_k \left|\phi_k(r)\right|^2 \tag{3}$$

The assumption in equation (3) leads to a set of one-electron Schrödinger-like equations

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + V_{\mathrm{eff}}[\rho(r)]\right]\phi_k(r) = \varepsilon_k \phi_k(r) \tag{4}$$

Equation (4) looks similar to the Schrödinger equation for a hydrogen atom and has the natural interpretation of a single electron existing in an effective density-dependent potential arising from all of the other electrons (mean field). While this interpretation can be very useful in many circumstances, there is in fact no formal justification for this view. The user should therefore exercise caution and be aware of the risk involved. In particular, the total energy E_{GS} and electron density $\rho(r)$ are the fundamental quantities, not the one-electron eigenvalues (band structure) ε_k and wavefunctions $\phi_k(r)$. As we have seen, the one-electron eigenvalues and wavefunctions are actually just mathematical constructs. It is therefore somewhat ironic that DFT-based methods are sometimes referred to as "band-structure" methods, in reference to the one-electron eigenvalues. Nonetheless, equation (4) is important to the theory itself because it is amenable to standard numerical methods. The procedure is to first make a guess for the one-electron wavefunctions in order to construct the effective potential V_{eff}, solve equation (4) for a new set of wavefunctions, and repeat this process until the input and output wavefunctions are the same (self-consistent).

Some strengths and weaknesses of density functional theory

The real strength of DFT-based methods is in calculating ground-state properties that can be obtained from the total energy. For example, by calculating the energy as a function of volume one can obtain the pressure

$$P = -\frac{\partial E}{\partial V} \tag{5}$$

Similarly, the equilibrium volume can be calculated by determining the volume at which the calculated pressure is zero (minimum in the energy)

$$V_0 = V(P = 0) \tag{6}$$

The second derivative (curvature) of the energy is related to the bulk modulus

$$B = V\frac{\partial^2 E}{\partial V^2} \tag{7}$$

By calculating the energy (and pressure) as a function of volume for a material in different crystal structures it is possible to predict pressure-induced phase transitions, as illustrated schematically in figure 1. Calculations of this type by Yin and Cohen[4] in 1982 for the case of silicon were among the first to demonstrate the power of DFT-based methods. A more recent

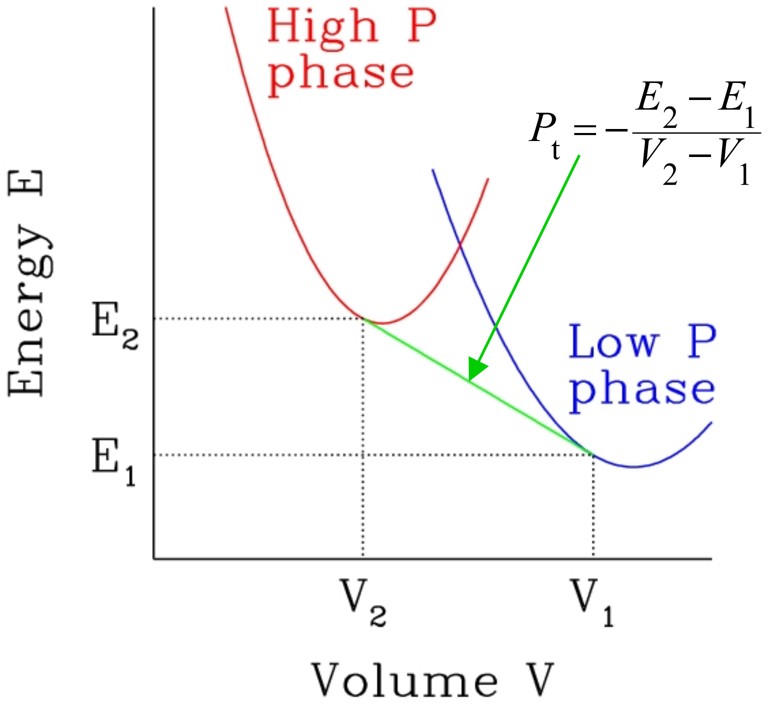

Figure 1. The transition pressure P_t between low- and high-pressure phases can be obtained by finding the common tangent (green line) between the energy as a function of volume curves for the two phases.

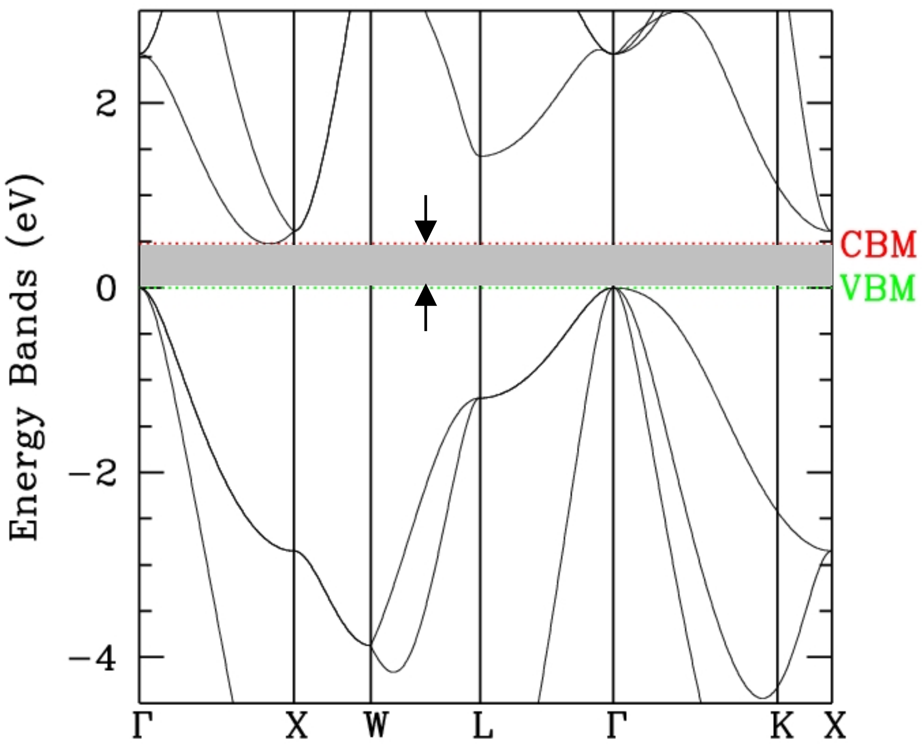

Figure 2. The calculated band structure of silicon showing a band gap of 0.5 eV. The experimental gap is 1.2 eV, more than a factor of two larger.

example is the work of Söderlind and Landa[5] for the actinide material americium. They find a phase transition from the face-centered cubic (fcc) structure to a face-centered orthorhombic (fco) structure at a pressure of 100 kbar. The corresponding experimental transition pressure is 107 kbar.

As I stated previously, it can be risky to interpret the one-electron eigenvalues ε_k as giving the band structure of the material. In particular, it is well known that the calculated band gaps in semiconductors and insulators are too small in comparison to experiment. As shown in figure 2 the calculated band gap for silicon is 0.5 eV but the experimental gap is 1.2 eV, more than a factor of two larger. One view of this result is that the method does not work very well. An alternate view more in keeping with the discussion here is that the one-electron eigenvalues are not the fundamental quantities in the theory and that care must be taken in their interpretation.

Local density approximation (LDA) and generalized gradient approximation (GGA)

Hohenberg and Kohn[2] proved that a universal functional of the density exists but since the form is unknown it is necessary to make an approximation is order to carry out practical calculations. Kohn and Sham[3] also addressed this issue. They started by separating out the known contributions to the functional from the unknown piece containing the complex many-body interactions

$$E[\rho(r)] = \text{Classical Coulomb (electrostatic) energy} + \text{Kinetic energy} + E_{\text{xc}}[\rho(r)] \qquad (8)$$

The classical coulomb energy is simply the energy of a classical distribution of charge. The unknown piece is labeled as the exchange-correlation energy E_{XC}. Kohn and Sham proposed that the contribution to E_{XC} at each point r be evaluated using E_{XC} of the uniform electron gas corresponding to the local density $\rho(r)$ and then integrated to obtain the total E_{XC}. This approximation is called the local density approximation (LDA). In practice the exchange correlation energy for the uniform electron gas is obtained from parameterizing the results of accurate quantum Monte Carlo simulations for the electron gas[6]. The term "local" in this context means that the exchange-correlation energy at a given point depends only on the density at that same point. In a "non-local" theory the exchange correlation energy at one point could depend on the density at other points as well. This idea is illustrated schematically in figure 3. In a local approximation the exchange-correlation energy at point r_1 is the same for both the "A" and "B" electron densities because they are both equal to ρ_1 at this point. However, in a non-local theory the exchange-correlation energy for A and B at r_1 could be different because the electron densities are not the same at other points.

The local density approximation is surprisingly simple but has also been stunningly successful. Clearly the approximation should work best in cases where the density is slowly varying, or equivalently, when the density gradients are small (the gradient is simply the three-dimensional first derivative). In some sense the local density approximation can be considered as the first constant term in a Taylor-series-like expansion. Clearly one could try to improve upon the approximation by including information not only about the density but also the density gradient

$$E_{\text{xc}} = E_{\text{xc}}[\rho(r), \nabla\rho(r)] \qquad (9)$$

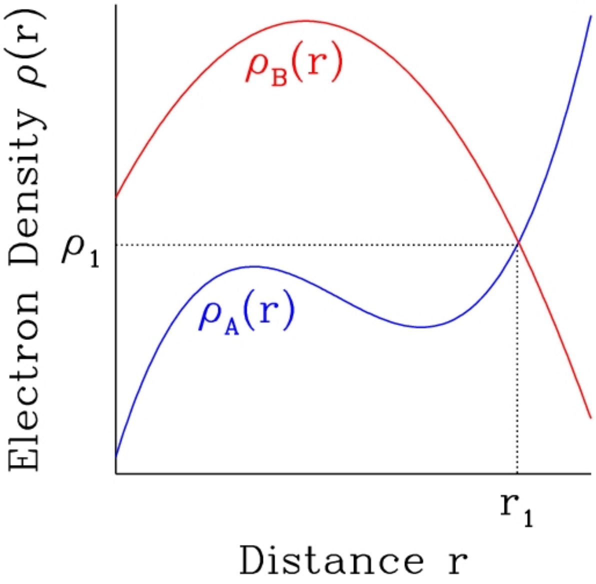

Figure 3. In a local approximation the exchange-correlation energy at point r_1 is the same for both the "A" and "B" electron densities because they are both equal to ρ_1 at this point. However, in a non-local theory the exchange-correlation energy for A and B at r_1 could be different because the electron densities are not the same at other points.

However, the naïve idea of simply using a Taylor series does not work because the density gradients are too large and the Taylor series does not converge. It is therefore necessary to be more careful in the manner of including density gradient information, leading to the term *generalized* gradient approximation (GGA). While it is possible to obtain information about the gradient dependence of the exchange-correlation energy from quantum Monte Carlo simulations and other analyses, it proved difficult to formulate a generalized gradient approximation that uniformly improved upon the results of the local density approximation. For this reason the early GGAs were not uniformly successful. However, they are more generally accepted now with the caveat that some GGAs are better suited to condensed-matter systems while others are better suited to finite-sized molecules and clusters.

Generalized gradient approximations are most appropriate in cases where there are large density gradients. One example is the case of isolated atoms. It had been know for a long time that the cohesive energies (energy to separate a system into isolated free atoms) obtained using the local density approximation were inaccurate. It was long assumed that the problem was with the free atom limit and since there are large density gradients in isolated atoms it is not surprising that GGAs improve calculated cohesive energies in comparison to experiment. Density gradients are also large in actinide materials (high Z) and therefore GGAs are commonly used when studying these materials. Another less obvious example is the case of pressure-induced phase transitions, as illustrated schematically in figure 4. If the relative energy difference between low- and high-pressure phases (see figure 1) increases with the use of a generalized

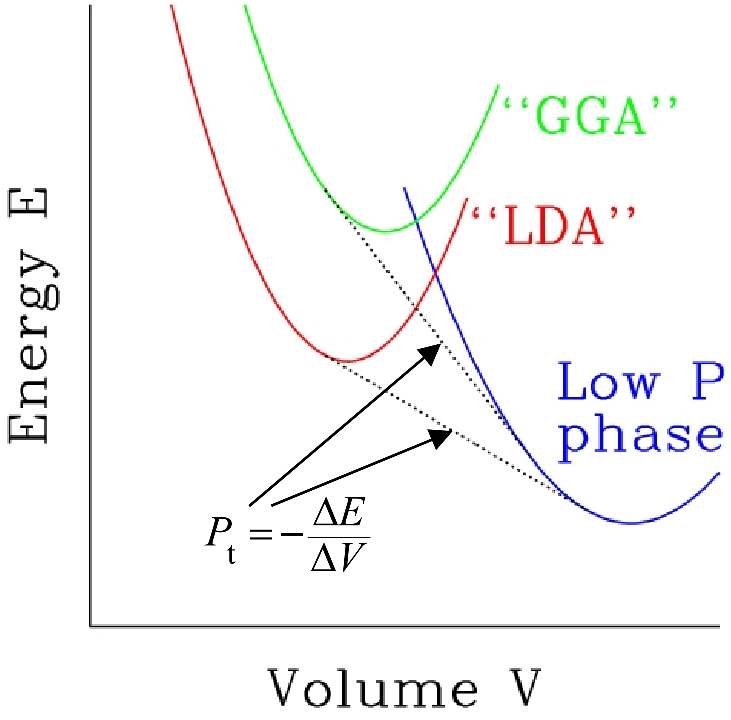

Figure 4. If the relative energy difference between low- and high-pressure phases (see figure 1) increases with the use of a generalized gradient approximation then the calculated transition pressure can increase relative to the pressure calculated using the local density approximation.

gradient approximation then the calculated transition pressure can increase relative to the pressure calculated using the local density approximation. In addition, calculations using GGAs generally result in larger equilibrium volumes and smaller bulk moduli in comparison to calculations based solely on the local density approximation. GGAs certainly do not solve all of the failings of the local density approximation. For example, GGAs do not fix the semiconductor gap problem because this is not a density gradient issue.

Pseudopotentials, LMTO, LAPW, and all that

Pseudopotential, linear muffin-tin orbital (LMTO), linear augmented plane wave (LAPW), and many other methods are simply different numerical procedures for solving the one-electron Schrödinger-like equations given above in equation (4). The most important point to note is that all of these methods are used to solve the same density functional theory equations and therefore they can all be equally accurate *if properly applied*. It is certainly true that different groups often obtain differing results but it is very frequently the case that the differences are numerical in nature. A primary difference among the various methods is the manner in which the core electrons are treated. In pseudopotential-based methods the core wavefunctions are assumed to be unchanged in the solid (or liquid) and therefore they are removed from the problem and replaced by an effective ion core or "pseudopotential." Conversely, in all-electron methods such as LMTO and LAPW both the core and valence electrons are treated self-consistently. All-electron methods are sometimes prefixed with the phrase full-potential (FP). This phrase refers to the fact that in certain methods the crystal potential is assumed to be spherically symmetric in the regions of space near the atoms but that a full-potential method does not make this approximation and should therefore be more accurate. At an even finer level of detail the difference between methods such as LMTO and LAPW is in the particular mathematical functions that are used to describe the one-electron wavefunctions. There can be various reasons for choosing a particular method over others. Pseudopotential methods scale better to large system sizes and are most often used in parallel computing applications. All-electron methods are less frequently used in general but are usually preferred when the core electrons are important. Examples include the presence of shallow core states (e.g. actinides) and in high-pressure studies since at higher pressures the core states shift upwards in energy toward the valence.

Going beyond density functional theory

Density functional theory has well-known limitations including the treatment of excited electron states (e.g. semiconductor band gaps, excitons—bound electron-hole pairs) and strong electron-electron correlation (e.g. localized f-electrons in actinide materials). A description of these methods is beyond the scope of the present paper but I will briefly mention a few important examples. The GW method[7] is used to calculate quasiparticle energies (excited states) and has been very successful in calculating accurate band gaps for semiconductors and insulators. Methods based on the two-particle Bethe-Salpeter equation[8] include the effect of the electron-hole interaction and are able to accurately treat excitons. Dynamical mean field theory (DMFT)[9] has the potential to combine the first-principles aspect of density functional theory with a more accurate treatment of strongly correlated electrons. It has been applied to f-electron materials including the volume collapse transition in cerium[10] and the phonon spectrum of δ-phase plutonium[11].

APPLICATIONS OF DENSITY FUNCTIONAL THEORY

There are far too many applications of density functional theory to even attempt to list them all here. Instead I will briefly describe a few examples to illustrate the wide variety of problems that can be treated. Applications directly related to the total energy are typically more reliable. I have already discussed the equation-of-state (pressure as a function of volume) and pressure-induced phase transitions. Other examples include calculations of the elastic constants and phonon spectrum, which are also obtained from the total energy. I will briefly describe quantum-mechanical forces on the atoms and molecular dynamics simulations below. Although I have cautioned that interpretation of the one-electron eigenvalues can be risky I do not want to give the impression that they are of no use. On the contrary, there are many applications of density functional theory that focus on the eigenvalues (band structure or electronic density-of-states). Many experimental probes, including electron energy-loss spectroscopy (EELS), x-ray absorption spectroscopy (XAS), photoemission spectroscopy (PES), inverse photoemission spectroscopy (IPES), and optical spectroscopy, all provide varying degrees of information regarding the band structure. In some cases an accurate theoretical treatment requires methods that go beyond density functional theory. In addition to bulk materials it is also possible to apply density functional theory to systems in lower dimensions. I will describe the application to surfaces, impurities, and defects below. Another area of broad application is magnetic systems and I will conclude this paper by discussing spin-polarized density functional theory and the current controversy regarding magnetic calculations for actinide materials.

Molecular dynamics and thermal properties

The quantum-mechanical forces on the atoms can be obtained directly from the total energy

$$M_i \ddot{R}_i = F_i = -\frac{\partial E}{\partial R_i} \qquad (F = Ma) \tag{10}$$

Equation (10) corresponds to Newton's second law of motion and can be used directly as the basis for molecular dynamics simulations. These simulations correspond to a classical treatment of the atoms but a fully quantum-mechanical treatment of the electrons. This methodology is called quantum molecular dynamics (QMD) and it allows the study of thermal properties of condensed matter systems, including melt curves, temperature-dependent phase diagrams, the atomic structure of liquids, and many others. QMD simulations are very computationally intensive, especially for high-Z materials such as actinides. Recently QMD simulations have been carried out for liquid uranium[12]. The forces in equation (10) can also be used to determine the optimal atomic positions for complex crystalline phases. The procedure is to let the atoms move in the direction of the forces until the residual forces are essentially zero. The resulting atomic structure is labeled as "relaxed" and corresponds to at least a local minimum in the total energy and is therefore predicted to be at least metastable.

In contrast to QMD simulations, classical molecular dynamics (MD) simulations replace equation (10) with a force law that is obtained empirically or semi-empirically. The result is that there is no quantum mechanics because the electrons have been removed from the problem. In essence the empirical force law corresponds to integrating out the electronic degrees of freedom, leaving only effective forces between the atoms. Various names are associated with classical MD simulations, including modified embedded atom method (MEAM), model generalized pseudopotential theory (MGPT), Lennard-Jones, etc. All of these names correspond to a particular choice for the empirical force law. The same thermal properties can be studied as in the case of QMD simulations. However, classical MD simulations are much more computationally efficient and therefore it is possible to study much larger systems and longer time scales than is possible with QMD simulations.

Surfaces, impurities, and defects

It is frequently advantageous to be able to use the same three-dimensional crystal code to also study lower dimensional systems such as surfaces (two dimensions) and defects (zero dimensions). The procedure is to create artificial three-dimensional periodicity by constructing a "supercell" that preserves periodic boundary conditions in all three dimensions. This procedure is illustrated schematically in figure 5 for a two-dimensional surface problem. Two-dimensional sheets (or "slabs") of bulk material are separated by two-dimensional vacuum regions. Normal periodic boundary conditions are preserved in the direction parallel to the surface but the slabs and vacuum regions are also repeated periodically in the direction perpendicular to the surface. It is necessary to test the convergence of whatever quantity is being calculated with respect to the perpendicular thickness of both the slab and vacuum regions. A one-dimensional line or wire could be handled in a similar manner.

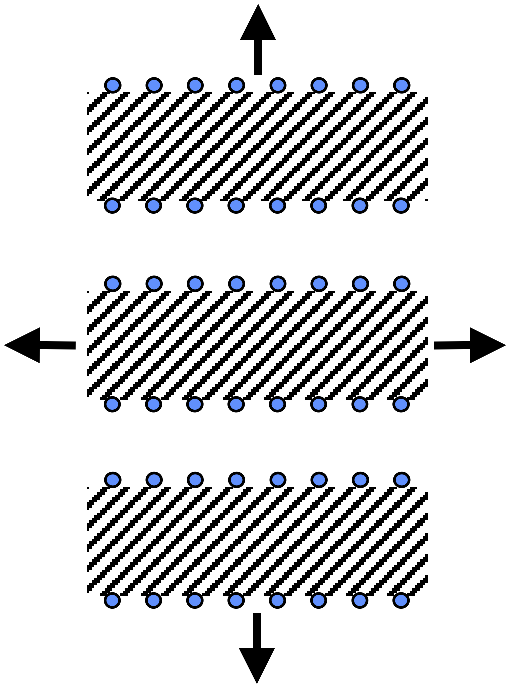

Figure 5. Schematic illustration of a surface supercell, or slab geometry, in which regions (or slabs) of two-dimensional bulk material are separated by vacuum regions. Note that each slab has two surfaces that may or may not be equivalent by symmetry and periodic boundary conditions exist for all three dimensions.

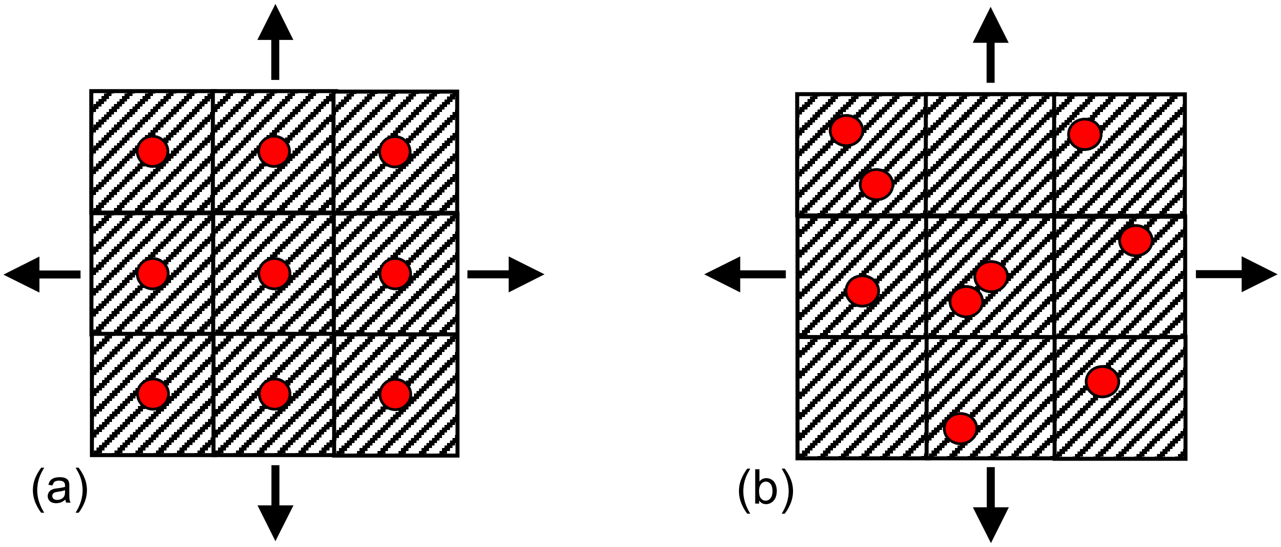

Figure 6. Schematic illustration of the supercell method for treating localized defects or impurities. Panel (a) shows an ordered array of defects while (b) shows how using a bigger supercell enables the treatment of disorder by allowing a distribution of defect sites while still preserving the overall density.

The analogous method for treating a localized defect or isolated impurity is to start with a large box containing a single defect and then periodically repeating this supercell in all three dimensions, as illustrated in figure 6a. Once again it is necessary to carry out convergence tests to insure that artificial defect-defect interactions are sufficiently small. Note that the net result is an ordered array of defects. The procedure for treating disorder is to use a bigger box (or supercell) so that a distribution of defect sites can be included, as shown in figure 6b. This solution for handling disorder is numerical and can quickly become very computationally expensive. In contrast the coherent potential approximation (CPA)[13] provides a mathematical solution to the problem of disorder. CPA-based calculations enable an accurate treatment of truly random disorder and they are much less computationally intensive than using the supercell method. However, the CPA approach is typically limited to substitutional impurities. The cerium-thorium alloy system provides a recent example in which the accurate treatment of substitutional disorder provided by the CPA yielded an improved agreement with measurements of a pressure-induced phase transition[14].

Magnetic density functional calculations for plutonium: fact or fiction

Magnetism can be treated using spin-polarized density functional theory in which the total density is divided into spin up and spin down densities

$$\rho_{\text{tot}}(r) = \rho_{+}(r) + \rho_{-}(r) \tag{11}$$

The local density approximation—LDA is then replaced with the local spin-density approximation—LSDA but otherwise the magnetic calculations are carried out in the same manner as nonmagnetic calculations. Söderlind and Sadigh[15] have used this approach to calculate the total energy and equilibrium volumes for the six known polymorphs of plutonium at ambient pressure. The resulting theoretical equilibrium volumes are in very good agreement with experiment and the corresponding relative energy differences are consistent with the complex experimental phase diagram. This level of agreement is truly remarkable, especially since it was thought until recently that density functional theory was not even capable of accurately treating all six phases, a number of which are believed to exhibit strong electron-electron correlation. However, the prediction within the calculations of large magnetic moments is somewhat controversial since experiments appear to suggest that there is no moment.

The question I will address here is whether or not the presence of a magnetic moment in the calculations should invalidate the results. As I have said a number of times, the total energy is the fundamental quantity in density functional theory—not the single-particle eigenvalues or wavefunctions. In particular, the total energy is variational in the density, which means that the energy is somewhat insensitive to errors in the density. The primary effect of introducing spin polarization (magnetism) in this case is to provide an extra variational degree of freedom that leads to a dramatic improvement in the calculations. It turns out that this result is not accidental because spin polarization produces a splitting of the f-band into occupied and unoccupied pieces, which is precisely the essential physics contained in more accurate correlated electron theories such as DMFT. While the exact nature of this splitting may be incorrect within spin-polarized density functional theory it appears that the variational principle nonetheless leads to robust total energies. It therefore seems to me to be a mistake to focus too much attention on the presence or absence of a magnetic moment. In the end it is more constructive to ask what we can learn from the density functional calculations. In particular, it appears that strong electron-electron correlations are present in the α-phase in addition to the higher temperature phases. Ultimately these calculations represent a challenge to correlated electron theories such as DMFT that are not yet able to provide calculations at the same level of accuracy for all of the six phases.

ACKNOWLEDGMENTS

This work was performed under the auspices of the U.S. Department of Energy by University of California, Lawrence Livermore National Laboratory under Contract W–7405–ENG–48.

REFERENCES

1. R. M. Martin, *Electronic structure: Basic theory and practical methods*, (Cambridge University Press, Cambridge, 2004).
2. P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).
3. W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).
4. M. T. Yin and M. L. Cohen, Phys. Rev. **B 26**, 5668 (1982).
5. P. Söderlind and A. Landa, Phys. Rev. **B 72**, 024109 (2005).
6. D. M. Ceperley and B. J. Alder, Phys. Rev. Lett. **45**, 566 (1980).
7. W. G. Aulber, L. Jonsson, and J. W. Wilkins, Solid State Physics **54**, 1 (2000).
8. L. X. Benedict and E. L. Shirley, Phys. Rev. **B 59**, 5441 (1999).
9. G. Kotliar and D. Vollhardt, Physics Today **57(3)**, 53 (2004); A. Georges, G. Kotliar, W. Krauth, and M. J. Rozenberg, Rev. Mod. Phys. **68**, 13 (1996).
10. K. Held, A. K. McMahan, and R. T. Scalettar, Phys. Rev. Lett. **87**, 276404 (2001).

11. X. Dai, S. Y. Savrasov, G. Kotliar, A. Migliori, H. Ledbetter, and E. Abrahams, Science **300**, 953 (2003).
12. R. Q. Hood and L. H. Yang, unpublished.
13. D. D. Johnson, D. M. Nicholson, F. J. Pinski, B. L. Gyorffy, and G. M. Stocks, Phys. Rev. **B 41**, 9701 (1990).
14. A. Landa, P. Söderlind, A. Ruban, L. Vitos, and L. Pourovskii, Phys. Rev. **B 70**, 224210 (2004).
15. P. Söderlind and B. Sadigh, Phys. Rev. Lett. **92**, 185702 (2004).