PHYSICAL REVIEW B VOLUME 48, NUMBER 8 15 AUGUST 1993-II

Additional condition for transferability in pseudopotentials

Michael Teter
Corning, Inc., Corning, New York 14831
and Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, New York 14853-2501

(Received 9 November 1992)

The chemical property of hardness is extremely important in the correct prediction of electron transfer between reacting chemical systems. Hardness is defined to be one-half the second derivative of the total energy of a chemical system with respect to the number of electrons. The requirement that the total energies of atom and pseudoatom match to second order with arbitrary changes in valence-state occupancy yields major decreases in the errors made when using psudopotentials. The concept also clarifies the role of core density in pseudopotential application. A practical prescription is presented for generating such pseudopotentials and their corresponding core densities, and several *ab initio* results are compared with those of high-quality norm-conserving pseudopotentials. Significant improvements are found.

I. INTRODUCTION

The intent of this work is to show one way to significantly reduce errors inherent in the pseudopotential method. Modern pseudopotentials already are at a high state of development, but the introduction of gradient corrections[1] and other improvements in density-functional theory are showing the possibility of further order-of-magnitude reductions in the already small errors inherent in the local-density approximation (LDA).[2] As the theory itself becomes more capable, users will become less tolerant of errors introduced by pseudopotential approximations. The direction taken will be to introduce into pseudopotential construction the concept of chemical hardness in a state-dependent manner. The reason for the introduction is to ensure the correctness of the total energy to second order in valence-state occupancies. The errors in hardness in a good norm-conserving pseudopotential are small, but they are about the same size as the errors in the pseudopotential itself. The constraints imposed by this concept are moderate, but the improvements in the prediction of material properties are significant. These changes usually result in wave functions which require higher kinetic-energy plane waves than a comparable norm-conserving pseudopotential and so a plane-wave basis set will require more plane waves. This difficulty is ameliorated by two facts: (1) The added details in the wave function are real and physically significant; (2) new solution methods scale nearly linearly with the number of plane waves and so the impact on solution time is bearable.

Those wishing to simulate systems of interacting atoms have three reasonable choices for basis sets: bases which resemble atomic orbitals, plane-wave or grid-based methods, and mixtures of the two. Most computational chemists use the first set, which is usually based on Gaussian functions. Solid-state physicists have traditionally used the second because of the requirements of periodicity for crystals. When neither of the first two are appropriate by themselves, mixed basis sets are used.[3] A major advantage of atomic orbital and mixed bases is that they can represent energetic core states with only a few parameters. To represent a deep 1*s* core state using an equally spaced grid requires spatial resolution significantly better than $1/Z$, where Z is the atomic number of the nucleus. To describe the 1*s* state of silicon within a 10-bohr cube requires a minimum of 10^8 grid points, or, in momentum space, approximately 10^7 plane waves within the cutoff sphere. Even with new plane-wave solution algorithms which scale nearly linearly with the number of plane waves,[4-6] solving for the wave functions using a basis set this large is a tedious numerical task. It is, however, perfectly feasible given sufficient computer memory.

The need for this high spatial resolution has traditionally been bypassed through the use of pseudopotentials which eliminate the core states. Even so, systems which have 1*s*, 2*p*, 3*d*, 4*f*, or 5*g* valence wave functions require deep pseudopotentials because these wave functions are nodeless in the radial direction and the pseudopotential cannot soften the electron-nuclear interaction by eliminating oscillations in the wave function and thereby reducing the corresponding kinetic energy. If one wishes to generate pseudopotentials for an atom which has a nodeless valence wave function, nonlocal pseudopotentials are generally required to accurately model the atom because different angular momentum states of the same principle quantum number require very different potentials. The number of plane waves required to model the states resulting from these deep potentials is large. Since the use of regular nonlocal pseudopotentials scales as the square of the number of plane waves, deep nonlocal pseudopotentials have become the bottleneck in the application of the new solution methods. Thus whole families of important materials have been excluded from plane-wave simulation for several years. These included materials with ions containing 2*p* valence states such as oxides, 3*d* transition metals, and 4*f* rare earth metals.

The use of separable pseudopotentials as proposed by Kleinman and Bylander[7] allows norm-conserving nonlocal pseudopotentials to be used in a plane-wave formulation without destroying the nearly linear scaling of the new algorithms. Allan and Teter[8] first showed that the deep $2p$ pseudopotential in oxides could be accurately simulated using separable pseudopotentials. The use of separable nonlocal pseudopotentials coupled with the new fast solution algorithms has opened up light atom chemistry and $3d$ transition metals[9] to fast plane-wave simulation.

The simplicity of using a constantly spaced grid coupled with extremely fast solution algorithms makes the plane-wave pseudopotential (PWP) method fully competitive with normal computational chemistry methods even for isolated molecules.[10] Forces and higher derivatives are much easier to generate in the PWP method than in methods where the basis set depends upon atomic positions. The large energies associated with the core states do not confound the delicate energetic changes associated with chemical reactions among the valence electrons. The only coherent argument against their use is the error in the pseudopotential approximation itself. With modern norm-conserving pseudopotentials, that error is small, but not necessarily negligible. Thus the goal of this work is to determine and significantly reduce the dominant source of error remaining in the use of pseudopotentials.

Modern pseudopotentials have a long history. A short review and list of references may be found in the book by Cohen and Chelikowsky.[11] For those interested in more detail, the excellent review by Pickett is recommended.[12] The breakthrough in the construction of *ab initio* pseudopotentials which were both practical and accurate came from the implementation of norm-conservation in nonlocal pseudopotentials. A norm-conserving pseudopotential conserves the normalization of the pseudo-wave-function inside the core region so that the wave function outside the core resembles as nearly as possible that of the all-electron atom. Norm-conservation was first implemented in empirical pseudopotentials by Topp and Hopfield,[13] in *ab initio* local ionic pseudopotentials by Starkloff and Joannopoulos,[14] and stressed in the creation of nonlocal pseudopotentials by Hamann, Schluter, and Chiang.[15] This work was refined by Bachelet, Hamann, and Schluter.[16]

If a spherical boundary is placed around an atom and one requires that outside that sphere the valence wave functions of the pseudoatom are exactly the same as those of the all-electron atom, then the boundary conditions are matching eigenvalues, matching wave functions, and matching derivatives on the surface of the sphere. To guarantee a solution to Schrödinger's equation outside the sphere in the all-electron potential, only the ratio of the wave-function derivative to the wave function (log derivative) is required to be correct on the surface of the sphere. For the normalized wave function to have the correct magnitude outside the sphere, however, it is required that the integral of the square of the wave function inside the sphere be the same for atom and pseudoatom. This is the norm-conservation condition. This condition guarantees that outside the spherical surface, the magnitude as well as shape of the electron density will be the same for the norm-conserving pseudoatom and the all-electron atom. This will preserve as closely as possible electrostatic interactions between valence wave functions. Norm conservation also has a deeper result. It guarantees that the log derivative of the pseudo-wave-function on the surface of the sphere will vary in the same manner as that of the all-electron atom with small variations in energy around the eigenvalue.

Transferability in a pseudopotential is the quality that the pseudoatom described by the pseudopotential accurately mimic the quantum-mechanical behavior of the all-electron atom in a wide range of environments. After the success of norm-conserving pseudopotentials in many applications, there has been a great tendency to associate transferability with the correct variation of log derivatives of the pseudo-wave-function with energy. Thus many attempts to improve norm-conserving pseudopotentials have centered around the log derivatives of the atomic and pseudoatomic wave functions and attempting to improve the energy range over which they match. The extended-norm-conserving condition of Shirley *et al.*[17] is a significant attempt to improve pseudopotentials past the norm-conserving condition. One major difficulty with using their formulation is that their analytic expression for the extended-norm-conservation condition is not integrable over a radial node in the wave function. In practice the condition must be implemented by matching the log derivatives at a sequence of energy points around the eigenvalue. The recent work of Chou[18] shows one way of performing this construction approximately. Vanderbilt[23] achieves log derivative matching over an extended energy range by a different construction technique which will be discussed later.

The normal method of pseudopotential construction is to generate a pseudo-wave-function with the desired properties and then invert Schrödinger's equation to yield the pseudopotential. If a series of wave functions are generated at different energies, inversion does not guarantee that they will yield the same potential. They usually do not. Chou forms a pseudopotential from a linear combination of the resulting pseudopotentials which gives the best match to log derivatives at the reference energies. She also directly associates transferability with the correct energy behavior of the log derivatives. A numerically tedious, but more general, approach is to bypass the construction of pseudo-wave-functions altogether and parametrize the pseudopotential itself. For any given set of parameters, the pseudopotential may be evaluated with regard to any number of variables including log derivatives of pseudoatomic wave functions over an extended energy range. A penalty function may then be evaluated which compares the resulting pseudo-wave-function properties with those of the all-electron atom, and the pseudopotential may then be optimized via a nonlinear least-squares procedure. Unfortunately, extending the energy range over which log derivatives match usually yields only a small improvement in the error associated with using a good norm-conserving pseudopotential. The reason for this seems to be that the error in the log derivatives of the pseudo-wave-function with energy is not the dominant error in pseudopotential use. A simple example is that of the oxygen atom. Given a large core radius of 2 bohr, it is possible to make a set of soft norm-conserving pseudopotentials whose log derivatives match those of the all-electron atom to better than 1% over a 1 hartree energy range. Using these pseudopotentials, the equilibrium radius of the O_2 molecule is 7.5% higher than that of the experimental value. The vibration frequency is in error by over a factor of 2. This example will be discussed further in Sec. IV.

Concentration on the log derivatives has also led to methods to generate pseudopotentials whose wave functions require a minimum number of plane waves,[19,20] known as soft pseudopotentials, because it is often possible to generate soft norm-conserving pseudopotentials with good log derivatives. The use of the adjective soft should not be confused with the concept of chemical hardness to be introduced later. A soft pseudopotential is to be desired since minimization of the number of plane waves required for convergence facilitates the solution process. The real wave function often has details which require many plane waves for their description. To soften the wave function requires that these details be smoothed. This smoothing process quite often leads to large atomic core radii since outside the core radius, the all-electron and pseudo-wave-functions are identical. One may only modify the wave function inside the core radius. If the details which are smoothed are physically significant, their effect is lost. In the absence of a coherent theory of transferability, the two techniques that the pseudopotential user has to increase transferability are to decrease the core radius for a given state or to include core states in the creation of the pseudoatom. Thus the desire for soft pseudopotentials seems to be at variance with the desire for transferability and chemical accuracy. With the new plane-wave algorithms, it is relatively painless to add plane waves, and associated costs such as computer memory are dropping rapidly. Thus the tradeoff of accuracy for softness is becoming less and less justifiable. It is not unreasonable to request that of all the pseudopotentials of equal performance, the softest one be chosen, but as mentioned above the matching of log derivatives alone is *not* a good measure of transferability.

A major contribution to transferability in pseudopotentials came from Louie, Froyen, and Cohen,[21] who realized that the deletion of the electron density of the core states removed by the pseudopotential construction gave a nonlinear error in the exchange-correlation potential in the LDA. By restoring the core density, they were able to show that atoms would respond correctly over a wider range of conditions and that many of the problems with modeling sodium were overcome. Their method is practical and almost without fail yields improvement in transferability. There are a few conceptual difficulties, however. For example, why is the core density of the all-electron atom the best one to be added to improve the behavior of the pseudo-wave-functions of the pseudoatom? If it is not the best density, how may one determine which density is best? Bylander and Kleinman[22] attacked the problem from a different point of view by showing that it was possible to generate an exchange potential with the core which corrected the problems with sodium without resorting to core density.

Vanderbilt[23] has provided a procedure for generating multireference soft separable pseudopotentials with log-derivative matching over an extended energy range. He is able to keep his wave functions soft because he does not require that they be norm-conserving in the usual sense. Instead he adds a compensating valence density to the density of the wave functions to achieve a norm-conserving condition. Vanderbilt thereby achieves softer pseudopotentials at a cost of a somewhat more complex projection operation.

The state of the art in highly transferable pseudopotentials is to construct them with small core radii with a norm-conserving procedure and to include core densities. If they are to be used in a separable manner then multiple reference states are advantageous not only for accuracy, but also to avoid ghost states.[24,25]

Since extending the energy range of correct log derivatives does not necessarily enhance the performance of a pseudopotential, let us reexamine what errors we may be making in the pseudopotential approximation. As isolated atoms are allowed to interact new states are produced which yield new eigenvalues, and in many cases significant transfer of electrons from one atom to another. Thus the changes in wave functions with eigenvalue and with occupancies are interrelated. The eigenvalue of a state is the derivative of the total energy with respect to the occupation of that state. By requiring the same eigenvalues, one has enforced the condition that small changes in the occupancies of atomic valence states yield the same energy changes in both the atom and pseudoatom. The norm-conservation condition enforces the condition that if the eigenvalues of atom and pseudoatom have the same changes, the valence wave functions for both the pseudoatom and the all-electron atom will change in the same manner on the surface of the sphere. There is, however, no guarantee that for a given perturbation, the change in the eigenvalues of the pseudoatom will be correct. Thus ensuring the correct eigenvalue changes under perturbation will remove an important source of error in pseudopotentials. This work will attempt to remedy the situation and guarantee that for at least one important case, namely, that of arbitrary changes in valence-state occupancy, changes in the eigenvalues of atom and pseudoatom will track exactly.

II. CONSERVATION OF DETAILED HARDNESS

If isolated systems composed of electrons and nuclei are allowed to interact, one result of the interaction is to bring the chemical potentials of the electrons of the isolated systems into agreement. The electron chemical potential μ of the system is defined as

$$\mu=\frac{\partial E}{\partial n},$$

where E is the total energy of the system and n is the number of electrons.

Despite the fact that the number of electrons in a system is an integer, throughout this work we will treat electron number and state occupancies as continuous variables. Thus the concept of differentiation with respect to occupancies has at least a mathematical meaning.

The minimization of the total energy E requires that electron transfer take place between the subsystems until the chemical potential is constant throughout the interacting systems. Given two systems, expanding the chemical potential of each linearly in the number of electrons and requiring that the two chemical potentials be equal give

$$\mu_1+\frac{\partial\mu_1}{\partial n_1}\delta n_1=\mu_2+\frac{\partial\mu_2}{\partial n_2}\delta n_2 .$$

Conservation of electrons implies that

$$\delta n_1=-\delta n_2$$

and ignoring constraints on the occupancies then gives

$$\delta n_1=\frac{\mu_2-\mu_1}{\dfrac{\partial\mu_1}{\partial n_1}+\dfrac{\partial\mu_2}{\partial n_2}} .$$

Thus the number of electrons transferred between the two systems depends upon the difference in their chemical potentials divided by the sum of the derivatives of the chemical potentials with respect to the number of electrons. One-half of this derivative is defined by Parr and Pearson[26] to be the "hardness" of the system η

$$\eta=\frac{1}{2}\frac{\partial^2E}{\partial n^2} .$$

The electron chemical potentials of the systems and the hardnesses of the systems then have equal roles in determining the amount of electron transfer between them.

There is an indeterminate number of ways to arrange the eigenvalue spectra of a pseudoatom to reproduce the derivative of the total energy with respect to electron number of an all-electron atom. If the occupancy of the highest occupied valence state is to be infinitesmally changed, then only the eigenvalues of that state in the atom and pseudoatom must match. If the concept is generalized to allow changes in any of the valence states, however, then all of the eigenvalues of corresponding valence states in the atom and pseudoatom must match. The usual norm-conserving pseudopotential construction accomplishes this matching essentially exactly. Similarly, there are again an indeterminate number of ways to make the hardnesses of the atom and pseudoatom match. If, however, arbitrary changes are again allowed in valence-state occupancies, then the entire matrix of second derivatives of the total energy with respect to the various occupancies must match in order to have the correct changes of the eigenvalues of atom and pseudoatom with such changes in occupancy.

We therefore define a hardness matrix for the atom

$$\eta_{ij}=\frac{1}{2}\frac{\partial^2E}{\partial f_i\partial f_j},$$

where f_i is the occupation of the ith state. The eigenvalue ϵ_i is

$$\epsilon_i=\frac{\partial E}{\partial f_i}$$

and combining the two expressions

$$\eta_{ij}=\frac{1}{2}\frac{\partial\epsilon_i}{\partial f_j} .$$

This expression has an explicit contribution resulting from the direct dependence of the density on the state occupancy and an implicit part which results from the relaxation of the wave functions with changes in state occupancy. While the exact hardness expression includes the effects of self-consistency, we will treat the wave functions as frozen in this work for two reasons: (1) the proper treatment of self-consistency greatly extends the necessary computation and (2) the inclusion of self-consistency makes no detectable difference in the quality of the pseudopotential.

The expression for the electron density in terms of the occupation numbers is

$$\rho(r)=\sum_i f_i\rho_i(r) ,$$

where

$$\rho_i(r)=\psi_i^*(r)\psi_i(r) .$$

The LDA expression for ϵ_i is

$$\epsilon_i=-\tfrac{1}{2}\int\psi_i^*(r)\nabla^2\psi_i(r)d^3r+\int\rho_i(r)V_{\text{ext}}(r)d^3r+\int\int\frac{\rho_i(r)\rho(r')}{|r-r'|}d^3r\,d^3r'+\int\rho_i(r)\mu_{\text{xc}}[\rho(r)]d^3r .$$

Assuming frozen wave functions, the only terms in the eigenvalue expression which have dependence on the total density $\rho(r)$ and therefore nonzero derivatives with respect to occupation number are the Hartree term and the exchange-correlation term. Using the result that

$$\frac{\partial}{\partial f_i}=\rho_i\frac{\partial}{\partial\rho} ,$$

the expression for the derivative of the ith eigenvalue with respect to the jth occupation number generates the following form for η_{ij}:

$$\eta_{ij}=\frac{1}{2}\int\int\frac{\rho_i(r)\rho_j(r')}{|r-r'|}d^3r\,d^3r'+\tfrac{1}{2}\int\rho_i(r)\rho_j(r)\frac{\partial\mu_{\text{xc}}}{\partial\rho}d^3r .$$

This expression has a very simple interpretation. The variation of f_j will change the density which will affect the eigenvalue of state i in two distinct manners: first through electrostatic coupling and second through changes in the exchange-correlation potential. Typically the electrostatic term dominates the exchange-correlation term by an order of magnitude and is always positive.

It is postulated in this work that errors in the hardness matrix of a norm-conserving pseudoatom dominate its transferability. An error in hardness gives a second-order error in the total energy with changes in occupancy and a first-order error in the variation of a valence-state eigenvalue. If the eigenvalue change is wrong, then matching log derivatives to higher order than first in energy is useless. If the energy changes match correctly to higher order, then extended norm-conservation may very well be significant.

The diagonal parts of the matrix resemble self-energies. They do not necessarily dominate the off-diagonal parts. The more nearly the densities of the individual states of the pseudoatom and atom match, the closer the electrostatic terms will become. The role of core density also becomes clear. Even if the densities of the valence states were perfect matches, $\partial\mu_{xc}/\partial\rho$ depends upon the total density including the core density. For this term to be correct, some form of the core density must be included. It is not necessary to include the exact core density. Since the pseudo-wave-functions are not the real wave functions, the exact core density would not give the exact derivatives anyway. It is merely necessary to bring the exchange-correlation part of the hardness matrix into better agreement with that of the all-electron atom than that caused by the complete neglect of the core charge. A reasonable smooth form as shown in the Appendix can, given the proper multiplier and core radius, reduce the disagreement between the all-electron case and the pseudoatom case by 2–3 orders of magnitude. The relationship between the Louie-Froyen-Cohen[21] and Bylander-Kleinman[22] treatment of sodium becomes clearer also. It is possible to partially compensate deficiencies in the exchange-correlation term by variations in the electrostatic term. The performance of norm-conserving pseudopotentials may be improved either with or without core charges if the hardness matrix of the pseudoatom can be made to more closely match that of the all-electron atom.

In the development of this work, the best results have been found by treating the electrostatic and exchange-correlation contributions to the hardness matrix separately and attempting to make each part match independently in the atom and pseudoatom without introducing compensating errors. Enforcing constraints on the electrostatic interaction between states couples the pseudopotentials and makes their generation an iterative process. After creating several pseudopotentials in this manner, it was found that a single constraint per state on its electrostatic self interaction in addition to the norm-conservation condition gave nearly as much improvement in the electrostatic part of the hardness matrix as fully coupling the states. This is reasonable. The additional constraint makes the densities of atomic and pseudoatomic states resemble one another more closely. As the self-interaction is improved, the interactions with more distant states should also improve. Thus only one additional constraint per state is necessary—the matching of the electrostatic self-energies of the atomic and pseudoatomic states. This always brings the log derivatives into better agreement as energies deviate from the eigenvalue. The converse is not necessarily true. Better log derivatives may or may not improve the hardness.

Core density only appears when unscreening the pseudopotential. A core density is added to the valence charge densities after the pseudopotential construction has been completed and the core density is varied to minimize the sum of squares of the differences of the exchange-correlation parts of the hardness matrices of the atom and pseudoatom. The sum of squares of differences can usually be improved by 2–4 orders of magnitude in this manner. The pseudopotentials are then unscreened by removing the pseudovalence density. The suggestion of Louie, Froyen, and Cohen[21] is followed in that the core density is used for exchange-correlation only and not in any of the electrostatic terms. This avoids strong electrostatic interactions which ideally should subtract, but in practice do not always exactly compensate. The form of the core density is chosen to be smooth in the sense of having an exact Fourier cutoff as well as decaying rapidly in real space. This is important for not creating errors in forces related to errors of integration of the exchange-correlation functions.

III. CONSTRUCTION OF HARDNESS CONSERVING PSEUDOPOTENTIALS

The scalar relativistic all-electron atom is solved on a radial grid in a manner very similar to that described by Hamann.[27] There are a few differences. First, because of the manner in which solutions are to be used on the grid, a much finer grid is used. It was found that the exponential grid usually used placed far too many points near the nucleus. After a great deal of numerical experimentation, the following grid was determined to be the best among those tested in the sense that it seemed to give the best overall accuracy for the least number of points:

$$r(n)=100\left[\frac{n}{n_{\max}}+\frac{1}{100}\right]^5-\frac{1}{10^8}$$

where n varies from 0 to $n_{\max}$. This gives a grid which grows nearly exponentially near the nucleus, but the growth factor dies down later and therefore the grid distributes a larger percentage of the points in the valence region of the atom. In this work $n_{\max}$ was chosen to be 2000. This large number is necessary because the method used to determine the optimum pseudopotential was nonlinear least-squares minimization of a penalty function. It was required that the wave functions be determined to at least nine significant figures so that the effects of very small changes in the coefficients describing the pseudopotentials could be tracked. Accuracy was checked by comparing numerical results with known analytic results such as hydrogenic solutions with high atomic number and harmonic-oscillator solutions. To avoid relativistic singularities near the nucleus, the nuclear potential was taken to be

$$V_n(r)=\frac{-Z}{(r^2+r_n^2)^{1/2}},$$

where r_n was chosen to be the approximate size of the nucleus, 10^{-5} bohr. The use of a finite nucleus also allows the radial wave function to start as r^l at the origin. Here l is the angular momentum quantum number. The atomic solutions are iterated until the sum of squares of the changes in the eigenvalues are smaller than 10^{-20}.

The states for which pseudopotentials are to be generated are chosen along with their core radii, and the pseudopotentials are parametrized in the following manner:

$$V_p(r)=\sum_{i=1}^{n} a_i g_i(r)+h(r)V_{scf}(r) \; .$$

The $g_i(r)$ are independent functions which have a value of one at the origin, have $i-1$ nodes, and then go to zero at r_c with zero slope. The function which multiplies the all-electron potential $h(r)$ starts at zero at the origin with zero derivatives through fourth order and goes to one at r_c with zero slope. These functions will be described in the Appendix. V_{scf} is the total self-consistent potential of the all-electron atom. The core radius in this construction is therefore of the Kerker[28] type and not of the Hamann-Schluter-Chiang type. After r_c in a Kerker construction, the wave function is exactly the all-electron wave function and the potential is the all-electron potential. In a Hammann-Schluter-Chiang construction, the pseudo- and all-electron wave-functions and potentials gradually blend into one another after r_c. Reasonable values of r_c in this work range from a minimum value slightly inside the last radial maximum of $r\psi_i(r)$ to a maximum defined by

$$r_{\max}=\left[\int_0^{\infty}\psi_i(r)^2 4\pi r^4 dr\right]^{1/2} ,$$

where $\psi_i(r)$ has been normalized such that

$$\int_0^{\infty}\psi_i(r)^2 4\pi r^2 dr=1 \; .$$

The squares of the differences between certain features of the all-electron wave function and those of the pseudo-wave-function are collected into a penalty function and minimized. These features include ratios of function values and derivatives at r_c not only for the eigenvalue, but for ten other points covering a range of one-half a hartree on either side of the eigenvalue. Log derivatives as such are not used because a zero in the wave function will generate a singularity in the log derivative. Instead, the following transformed log derivatives (TLD's) are used:

$$\frac{\psi(r)\dfrac{\partial\psi(r)}{\partial r}}{\dfrac{\partial\psi(r)^2}{\partial r}+\psi(r)^2} \; .$$

It is clear that if the numerator and denominator are both divided by $\psi(r)^2$ the above expression depends only on the log derivative of $\psi(r)$. The advantage of the TLD's in avoiding singularities is obvious since $\psi(r)$ and its derivative cannot be zero at any place other than the origin without both being zero everywhere as long as the potential is restricted from going to positive infinity.

There exists a small inconsistency in a pseudopotential construction in which the nonrelativistic Schrödinger equation is inverted after the pseudo-wave-function which satisfies the appropriate conditions has been generated. If the atom has been solved using scalar relativity, the wave function outside the core radius is a solution of the scalar relativistic equation and not Schrödinger's equation, creating relative errors of the range $10^{-4}-10^{-6}$. The inversion of Schrödinger's equation is necessary since the scalar relativistic wave equation may not be inverted easily. Given the pseudopotential rather than the pseudo-wave-function, one may then use the scalar relativistic wave equation in solving for the pseudo-wave-function avoiding this minor inconsistency. Unfortunately the use of the scalar relativistic wave equation slightly decouples the log derivatives from the norm-conserving condition. Thus the generalized norm-conserving condition of Hamann[27] for energies other than the eigenvalue is used in addition to the TLD's and the squares of deviations at the 11 energy points are added to the penalty function. The square of the difference between the electrostatic self energies of the atomic state and pseudoatomic state is also added to the penalty function.

If a cutoff energy is specified, the spherical Bessel transform of the pseudo-wave-function is constructed and the square of the fraction of the kinetic energy which lies above the cutoff energy is added to the penalty function as specified by Rappe *et al.*[19] This construction must be used delicately since attempts to grossly soften the pseudopotential destroy its quality. On the other hand, it is occasionally necessary because the optimization routine will sometimes deepen the pseudopotential by a factor of 2 or more to gain less than a part per thousand in the penalty function.

Normally, if the pseudopotential were to be used in a separable construction, such detailed attention to the performance of the pseudopotential would be wasted, since only the norm-conserving characteristic is guaranteed to survive being case into separable form. If, however, a pseudo-wave-function from one of the other energies is added to the reference set after orthogonalization in a Blochl[24] construction, then the details of the pseudopotential behavior are conserved over a much broader energy range. The use of a second reference state projects out most of the ghost states which are possible in the Bylander-Kelinman construction and also allows the maximum angular momentum pseudopotential to be used as the local pseudopotential. The cost paid for a second reference state is an increase in projection time, but since the second reference state is suitably orthogonal to the first, one may disable its use to see if the difference is significant. One may also make a gain by being permitted to use the highest angular momentum pseudopotential as local and avoid the most complex projection.

A surprisingly good second reference wave function may be generated for any state by simply multiplying the initial reference state by r^2 and orthogonalizing to the first using Blochl's weighted orthogonalization.[24] The reason that this works is that if the pseudopotential is an even power series in r, then the radial wave function for any angular momentum l is of the form r^l times an even power series in r. The first term is not energy dependent and thus the leading term in the energy derivative of the wave function is r^{l+2}, so that r^2 times the reference wave function mimics the energy derivative of the wave function for small r and makes a very good second reference wave function when suitably orthogonalized. The use of an exact wave function at a different energy is to be preferred, but this approximate procedure works quite well if such a wave function is not easily available. It is also a trivial procedure for those possessing a separable pseudopotential with a single reference wave function who want to see the improvement inherent in adding a second.

Finally, if a ghost state appears despite the double reference state, it is because one of the nonlocal potentials has dipped below the local potential. In this case, the square of a final residual may be added to the penalty function: the integral of the local potential minus the nonlocal one over those regions where the difference is positive. This final residual is then used along with all the others in an iterative manner to adjust the various pseudopotentials until the ghost state disappears.

To perform the nonlinear optimization, the Levenberg-Marquart algorithm is used.[29] To initialize the optimization, only $g_1(r)$ is used, and a search is performed covering the range of -100 to 100 for its coefficient. The minimum value of the penalty function discovered in the search determines the starting value. In general, optimization is performed sequentially using first one parameter and then this result is used as the starting case for two parameters and so on until the required accuracy in the penalty function is achieved. In general fewer parameters give softer potentials and so an accurate pseudopotential using only a small number (2–4) is desirable. The optimization is computationally intensive and the creation of a pseudopotential may take an hour of supercomputer time. Adding to the difficulty of creating the pseudopotentials is the large number of choices to be made among core radii, cutoff energies the number of residuals to use, number of parameters, and the like. The only justification for such a cumbersome procedure must lie in its results. It is important to note that the only reason for the nonlinear optimization procedure lies in the desire to make a single pseudopotential have extended-norm-conserving properties in its log derivative as well as making it hardness conserving. If changes in eigenvalues with respect to occupation numbers are now essentially correct, it is only reasonable to require that changes in the wave function with eigenvalue also be as error free as possible. If a user only wants the improvement from hardness conservation, the electrostatic self-energy condition may be simply added to a regular construction procedure. This will improve the resulting pseudopotentials and also improve the energy dependence of the log derivatives.

After the pseudopotentials and pseudo-wave-functions have been created, a two-parameter core density is created, with the form

$$\rho_{\text{core}}=Cf(r/r_{\text{core}}) \ .$$

The function f is also described in the Appendix. Both the coefficient and the core radius are adjusted so that the sum of squares of the differences between the exchange-correlation parts of the hardness matrices of the atom and pseudoatom are minimized. This is also an easy feature to add to any normal pseudopotential program. It is generally worthwhile to perform this optimization whether or not the electrostatic part of the hardness matrix is conserved.

IV. RESULTS

When possible, comparisons are made between the predictions of various pseudopotentials and all-electron calculations. These calculations include spherically symmetric silicon using the atomic program and a three-dimensional series of all-electron plane-wave calculations for carbon in the diamond structure. These all-electron plane-wave calculations are presented for carbon and mark the first time, to our knowledge, that all-electron plane-wave calculations for species other than hydrogen have been performed. The reason is clear: 100 000–200 000 plane waves per band are required within the cutoff sphere. The kinetic-energy cutoff required for converging such a simulation is approximately $9Z^2$ hartrees, where Z is the atomic number. Nevertheless, the convergence time for a typical all-electron simulation of diamond in a two-atom cell with ten special k points is merely a few hours on a work station. At present, computer memory restrictions identify nitrogen as the largest atom which may be reasonably simulated in this manner. These simulations are important in that they can approach the LDA limit without using experiment as the judge of the quality of a pseudopotential. Agreement with experiment should be used to judge the basic density-functional theory, the pseudopotential should be judged by its agreement with all-electron results. Nevertheless, in the case of the oxygen diatomic molecule and the relaxed structure of α-quartz, experiment is quoted for two reasons. First, we do not have sufficient computer memory to perform the all-electron calculation, and second, agreement is embarrassingly close.

A. Silicon eigenvalues

Silicon was chosen as the test case to determine if the non-self-consistent hardness constraint truly makes self-consistent eigenvalues track with occupation number and the number of valence electrons was varied from zero to eight. The spherically symmetric atomic program was used. This test monitored the self-consistent eigenvalues of the $3s$, $3p$, and $3d$ states as silicon was taken from the plus four ion in the neon configuration to the minus four ion in the argon configuration. A harmonic-oscillator potential was added to the silicon ion potential in order to stabilize the negative ions. The form of the potential was

$$V(r)=r^2/100$$

and is added to all of the test cases.

Table I shows the self-consistent eigenvalues of the $3s$, $3p$, and $3d$ states as the number of valence electrons varies. The various pseudopotentials were a norm-conserving pseudopotential, the same norm-conserving pseudopotential with core density added by the prescription in Sec. III, and finally an extended-norm and hardness-conserving (ENHC) pseudopotential with core density. The addition of the core density makes a major

TABLE I. Self-consistent eigenvalue differences in hartrees from all-electron calculations for various pseudopotentials and occupations.

State occupation	All-electron eigenvalue	Norm conserving	Norm conserving with core	Hardness conserving with core
$3s^0$	−1.790 24	+0.195 40	+0.021 48	−0.015 86
$3p^0$	−1.440 41	+0.175 29	+0.032 71	−0.011 43
$3d^0$	−0.992 51	+0.137 05	+0.011 96	−0.005 09
$3s^1$	−1.416 50	+0.029 86	+0.016 05	+0.000 09
$3p^0$	−1.089 99	+0.029 50	+0.019 21	+0.000 87
$3d^0$	−0.663 83	+0.014 29	+0.005 89	+0.000 01
$3s^2$	−1.001 06	+0.010 82	+0.008 40	+0.001 56
$3p^0$	−0.695 80	+0.011 42	+0.009 71	+0.001 79
$3d^0$	−0.302 12	+0.003 78	+0.002 75	+0.000 60
$3s^2$	−0.629 77	+0.002 49	+0.002 50	+0.000 75
$3p^1$	−0.345 29	+0.003 26	+0.003 22	+0.000 78
$3d^0$	+0.009 15	+0.000 56	+0.000 56	+0.000 27
$3s^2$	−0.291 16	−0.000 82	−0.000 21	+0.000 06
$3p^2$	−0.270 28	+0.000 52	+0.002 86	+0.000 65
$3d^0$	+0.283 40	+0.000 18	+0.000 17	+0.000 12
$3s^2$	+0.011 17	−0.001 55	−0.000 98	−0.000 24
$3p^3$	+0.256 04	−0.000 71	−0.000 59	−0.000 20
$3d^0$	+0.520 74	+0.000 55	+0.000 40	+0.000 13
$3s^2$	+0.276 87	−0.001 26	−0.000 85	−0.000 26
$3p^4$	+0.504 42	−0.000 54	−0.000 56	−0.000 19
$3d^0$	+0.726 08	+0.000 81	+0.000 63	+0.000 17
$3s^2$	+0.508 86	−0.000 77	−0.000 47	−0.000 20
$3p^5$	+0.721 70	−0.000 20	−0.000 28	−0.000 10
$3d^0$	+0.906 61	+0.000 87	+0.000 72	+0.000 20
$3s^2$	+0.712 55	−0.000 36	−0.000 13	−0.000 12
$3p^6$	+0.913 54	+0.000 04	−0.000 02	−0.000 02
$3d^0$	+1.069 23	+0.000 80	+0.000 71	+0.000 19

improvement as Louie, Froyen, and Cohen discovered. It is important to note that the core density construction used in this example is one constructed to optimize the exchange-correlation part of the hardness matrix, and is not the core charge construction of Louie, Froyen, and Cohen. Finally the remaining error is essentially removed with the hardness-conservation condition. All of the pseudopotentials had 1.75 bohrs core radii. The hardness-conserving pseudopotential had the cutoff condition added to the list of residuals so that 99.9% of the kinetic energy was below 25 hartrees. It is interesting to note that for the all of the pseudopotentials, the error is largest for the fewest valence electrons. The error essentially disappears by the time that the ion reaches the argon configuration. This is reasonable. In the limit of zero valence density, the core density is providing the entire exchange-correlation potential, and any errors in the core density will be magnified. By the time that there are eight valence electrons, the presence or absence of the core makes little difference.

B. Diatomic oxygen molecule

Two test cases are reported. Norm-conserving, and extended-norm and hardness conserving pseudopotentials are constructed with core radii of 2 bohrs for both the 2*s* and 2*p* states. The large core radius was chosen so that the norm-conserving pseudo-wave-functions required a cutoff energy of only 20 hartrees. Nevertheless, the log derivatives of the pseudo-wave functions tracked those of the all-electron atom to better than 1% over an energy which ranged from half a hartree below the eigenvalue to half a hartree above it. The 2*p* extended-norm and hardness conserving pseudopotential had 99.9% of the kinetic energy of its pseudo-wave-function below 40 hartree. A cutoff energy of 50 hartree was used throughout to ensure convergence. Two oxygen atoms were placed in a $10\times10\times15$ bohr^3 box and their energy calculated as a function of separation. A spin-polarized calculation was done with seven valence electrons in one spin state and five valence electrons in the other to ensure the correct ground-state spin configuration. The equilibrium radii and curvature are given in Table II for the two cases as well as for experiment.[31] The norm-conserving case had an equilibrium radius 7.5% too high and a curvature 60% too low. The ENHC case reproduced the experimental equilibrium radius to better than 0.5% and the curvature to within 3%.

C. Carbon under very high pressure

Total energies were calculated for three test cases. A norm-conserving pseudopotential at a cutoff of 50 hartrees and a core radius of 1.25 bohrs, an ENHC pseudopotential with a cutoff of 50 hartrees and a core radius of 1.25 bohrs, and an all-electron calculation performed

TABLE II. Equilibrium radius r_0 and $\partial^2 V(r)/\partial r^2|_{r_0}$ for the $^3\Sigma_g^-$ oxygen molecule for norm-conserving, extended norm, and hardness-conserving pseudopotentials, and experiment (Ref. 31).

	Norm-conserving	ENHC	Expt.
r_0 (bohr)	2.453	2.275	2.282
$\partial^2 V(r)/\partial r^2\|_{r_0}$ (hartree/bohr^2)	0.307	0.778	0.756

with a local ENHC pseudopotential for the 1*s* state of carbon with a 300-hartrees cutoff and a 0.4-bohrs core radius. A two-atom unit cell of diamond was simulated using ten special points under isostatic compression. Both pseudopotential calculations were taken to complete energy convergence. The lattice constant was varied from 5 to 8 bohrs. The results are given in Table III. For both pseudopotentials a constant was added to the total energies to bring their averages into agreement with the all-electron calculation. The total energies of the ENHC pseudopotential have a standard deviation of less than 2 mhartree when compared with those of the all-electron calculation. This is to be compared with a standard deviation of 49 mhartree for the norm-conserving pseudopotential. There is a reduction in the error of comparison with the all-electron case of a factor of more than 25.

Other physical properties are compared with their all-electron values. These properties include equilibrium lattice constant, bulk modulus, and pressure derivative of the bulk modulus. The all-electron values are reproduced an order of magnitude more accurately by the ENHC pseudopotential. The norm-conserving results may be improved by lowering the core radius, but the 2*s* radial node puts a limit on how far this improvement may be pushed. At the same core radius, the ENHC result will always be an order of magnitude better.

D. Quartz structure

The silicon and oxygen ENHC pseudopotentials defined earlier were used to refine the structure of α-quartz. Table IV shows the results of this work, an earlier work using norm-conserving pseudopotentials, and experiment.[30] The all-electron calculation for this structure could not be performed due to the extremely high cutoff energy required for silicon. The earlier calculation has errors in the lattice constants and coordinates of the order of 2%.[9] The results from the ENHC pseudopotentials show agreement at the 0.1% level.

V. SUMMARY

The requirement that the total energies of atom and pseudoatoms track to second order with respect to arbi-

TABLE III. Energies for diamond under isostatic compression for norm-conserving pseudopotential, all-electron, and extended norm- and hardness-conserving pseudopotential. A constant has been subtracted from both pseudopotential calculations to make their average energies the same as the all-electron case. Also included are the standard deviations of the differences from the all-electron case, the equilibrium lattice constant a_0, the bulk modulus B, and the pressure derivative of the bulk modulus.

Lattice constant (bohr)	Norm-conserving −63.112 325 (hartree)	All-electron (hartree)	ENHC −62.361 247 (hartree)
5.0	−73.559 096	−73.679 091	−73.677 824
5.8	−74.206 141	−74.242 149	−74.240 666
6.0	−74.281 435	−74.301 995	−74.300 981
6.2	−74.334 449	−74.342 446	−74.340 841
6.4	−74.360 408	−74.364 967	−74.364 244
6.6	−74.389 790	−74.374 818	−74.374 361
6.8	−74.398 456	−74.373 822	−74.373 871
7.0	−74.397 801	−74.364 842	−74.365 143
7.2	−74.389 796	−74.348 265	−74.350 026
8.0	−74.313 485	−74.247 460	−74.251 898
σ (hartree)	0.049 2		0.001 8
a_0 (bohr)	6.882	6.678	6.689
B (hartree/bohr^3)	0.013 62	0.016 28	0.015 71
$\partial B/\partial P$	3.728	3.599	3.648

TABLE IV. Structure of α-quartz for norm-conserving, hardness and extended norm-conserving pseudopotentials, and experiment.

Parameter	Norm-conserving	HENC	Expt.
a (bohr)	9.363	9.285	9.290
c (bohr)	10.304	10.209	10.215
Si_x	0.4638	0.4688	0.4697
O_x	0.4081	0.4140	0.4135
O_y	0.2758	0.2699	0.2669
O_z	0.2782	0.2839	0.2858

trary variations in valence-state occupancies when added to the usual norm-conserving conditions has led to the creation of a new class of pseudopotentials which have uniformly shown order-of-magnitude reduction in the errors associated with their use when compared with either all-electron or experimental results. These pseudopotentials conserve chemical hardness and have extended-norm-conserving characteristics. Analysis of the hardness matrix shows the reason for the importance of core densities in pseudopotential use and acts as a guide to the construction of reasonable core densities. When used in a separable construction these pseudopotentials are generated with dual reference states to not only stabilize them against ghosts, but also to ensure that the properties of these pseudopotentials track over much broader conditions than single reference ones. The condition that electrostatic self-energies of atomic and pseudoatomic states match is one that may be added to any pseudopotential construction program and together with the use of core densities should generate pseudopotentials of sufficient accuracy as to be practically indistinguishable from the corresponding all-electron calculation.

ACKNOWLEDGMENTS

The author acknowledges helpful conversations with Douglas Allan, Xavier Gonze, David Vanderbilt, Lin Wang Wang, and Andrew Rappe. He would like to thank Neil Ashcroft for the motivation to attempt all-electron plane-wave calculations, and Corning, Inc., the Cornell Laboratory of Atomic and Solid State Physics, the Cornell Theory Center, the Cornell Materials Science Center, and I.B.M. for resources made available during this work.

APPENDIX

The functional form used to describe the pseudopotential is

$$V_l(r)=\sum_{m=1}^{n} a_{lm} g_m(r)+h(r)V_{scf}(r) \ ,$$

with unknown parameters a_{lm}. Using the scaling

$$x=r/r_c \ ,$$

then in the range $0\leq x\leq 1$, the functions $g_m(x)$ have the form

$$g_m(x)=\frac{\sin(m\pi x)}{m\pi x}\sum_{j=0}^{6} b_{mj}x^{2j} \ ,$$

where

$$b_{m0}=1.000\,000\,000\,000\,000 \ ,$$

$$b_{m1}=1.644\,934\,006\,684\,822m^2 \ ,$$

$$b_{m2}=-4.715\,200\,508\,447\,231-6.409\,990\,562\,438\,041m^2 \ ,$$

$$b_{m3}=8.258\,744\,178\,343\,138+10.457\,122\,270\,276\,11m^2 \ ,$$

$$b_{m4}=-6.458\,685\,729\,920\,13-8.072\,930\,869\,018\,365m^2 \ ,$$

and

$$b_{m5}=1.915\,142\,903\,096\,105+2.380\,865\,094\,332\,069m^2 \ .$$

The m dependence of the b's has been chosen to smooth the functions. These functions have been built to incorporate the Troullier-Martins[20] condition that the coefficient of r^2 at the origin is zero. A great deal of numerical experience has shown that, while the softest pseudopotentials do not match this condition, the pseudopotentials with this condition and only one or two g terms are never very far from the optimum in terms of cutoff energy:

$$h(x)=1-(1-x^5)^2.$$

For $x>1$ the g's are all zero and $h(x)=1$.

As noted in the text, the form of the core density is

$$\rho_c(r)=\rho_{c0}f(x)^2 \ .$$

The form for $f(x)$ is

$$f(x)=\frac{\sin(2\pi x)}{2\pi x(1-x^2)(1-4x^2)} \ .$$

The Fourier transform of f is

$$f(G)=\frac{r_c^2[2\sin(\pi r_c G)+\sin(2\pi r_c G)]}{6G}$$

for G between 0 and $1/r_c$ and $f(G)=0$ after $1/r_c$. This form has been chosen to have an exact cutoff in Fourier space as well as die off quickly in real space. Practically, $f(x)^2=0$ after $x=3$.

---

[1] A. D. Becke, J. Chem. Phys. **96**, 2155 (1992).

[2] S. Lunqvist and N. H. March, *Theory of the Inhomogeneous Electron Gas* (Plenum, New York, 1983).

[3] S. G. Louie, K.-M. Ho, and M. L. Cohen, Phys. Rev. B **19**, 46 (1977).

[4] R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).

[5] M. C. Payne, J. D. Joannopoulos, D. C. Allan, M. P. Teter, and D. H. Vanderbilt, Phys. Rev. Lett. **56**, 2656 (1986).

[6] M. P. Teter, M. C. Payne, and D. C. Allan, Phys. Rev. B **40**, 12 255 (1989).

[7] L. Kleinman and D. M. Bylander, Phys. Rev. Lett. **48**, 1425 (1982).

[8] D. C. Allan and M. P. Teter, Phys. Rev. Lett. **59**, 1136 (1987).

[9] D. C. Allan and M. P. Teter, J. Am. Ceram. Soc. **73**, 3247 (1990).

[10] A. M. Rappe and J. D. Joannopoulos, and P. A. Bash, J. Am. Chem. Soc. **114**, 6466 (1992).

[11] M. L. Cohen and J. R. Chelikowsky, *Electronic Structure and Optical Properties of Semiconductors,* 2nd ed. (Springer-Verlag, Berlin, 1988), p. 16.

[12] W. E. Pickett, Comput. Phys. Rep. **9**, 115 (1989).

[13] W. C. Topp and J. J. Hopfield, Phys. Rev. B **7**, 1295 (1974).

[14] T. Starkloff and J. D. Joannopoulos, Phys. Rev. B **16**, 5212 (1977).

[15] D. R. Hamann, M. Schluter, and C. Chiang, Phys. Rev. Lett. **43**, 1494 (1979).

[16] G. B. Bachelet, D. R. Hamann, and M. Schluter, Phys. Rev. B **26**, 4199 (1982).

[17] E. L. Shirley, D. C. Allan, R. M. Martin, and J. D. Joannopoulos, Phys. Rev. B **40**, 3652 (1989).

[18] M. Y. Chou, Phys. Rev. B **45**, 11 465 (1992).

[19] A. M. Rappe, K. M. Rabe, E. Kaxiras, and J. D. Joannopoulos, Phys. Rev. B **41**, 1227 (1991).

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

[21] S. G. Louie, S. Froyen, and M. L. Cohen, Phys. Rev. B **26**, 1738 (1982).

[22] D. M. Bylander and L. Kleinman, Phys. Rev. B **43**, 12 070 (1991).

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

[24] P. E. Blochl, Phys. Rev. B **41**, 5414 (1990).

[25] X. Gonze, R. Stumpf, and M. Scheffler, Phys. Rev. B **44**, 8503 (1991).

[26] R. G. Parr and R. G. Pearson, J. Am. Chem. Soc. **105**, 7512 (1983).

[27] D. R. Hamann, Phys. Rev. B **40**, 2980 (1989).

[28] G. P. Kerker, J. Phys. C **13**, L189 (1980).

[29] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vettering, *Numerical Recipes* (Cambridge University Press, Cambridge, 1986), p. 523.

[30] L. Levien, C. T. Prewitt, and D. Weidner, Am. Mineral. **65**, 920 (1980).

[31] G. Herzberg, *Spectra of Diatomic Molecules,* 2nd ed. (Van Nostrand, New York, 1953), p. 560.