Pseudopotentials with position-dependent electron masses

W. M. C. Foulkes and M. Schluter
AT&T Bell Laboratories, Murray Hill, New Jersey 07974-2070
(Received 21 May 1990)

Many of the computational limitations of the Green's-function quantum Monte Carlo (GFQMC) method could be overcome by using pseudopotentials. However, standard norm-conserving pseudopotentials are unsuitable because they contain nonlocal angular-momentum projection operators. This prompted Bachelet, Ceperley, and Chiocchetti (BCC) to put forward a new class of pseudopotential specifically for use in GFQMC calculations. The BCC pseudo-Hamiltonian does not involve any nonlocal integral operators, but consists of a local potential and a kinetic-energy term with a position-dependent electron effective-mass tensor. We find that sensible BCC pseudopotentials exist only when the atomic valence eigenvalues increase with increasing angular momentum, and even when this condition is met it can be difficult to ensure the correct scattering properties in more than two angular-momentum channels. Nevertheless, we have constructed good BCC pseudopotentials for a wide selection of atoms and have done density-functional calculations in atoms and solids to test their transferability. It turns out that the transferability depends very much on the method used to construct the pseudopotentials (different methods can lead to very different pseudopotentials with very different transferabilities) but can be comparable to that of ordinary nonlocal pseudopotentials in some cases.

I. INTRODUCTION

The use of pseudopotentials in solid-state electronic-structure calculations has been widespread for at least 20 years.[1,2] However, until recently they had always been combined with methods (such as the Hartree-Fock method which is variational, or the density-functional[3-5] method which is exact in principle although not in practice) involving some sort of independent-electron–mean-field approximation. Approaches which attempt a better treatment of the electron-electron interaction are few, but among them are the various quantum Monte Carlo[6-8] (QMC) schemes. These have been valuable in light atoms and small molecules and may also prove useful in solids. It turns out that many of the problems associated with QMC calculations are the same as those which led to the introduction of pseudopotential methods in independent electron calculations, and so it is not surprising that there has been a recent effort to combine the QMC and pseudopotential approaches. In this paper we investigate the properties and transferability of a new class of pseudopotential recently introduced by Bachelet, Ceperley, and Chiocchetti[9] (BCC) as particularly suitable for use in Green's-function and diffusion QMC calculations.[6-8] We find that there are strict mathematical bounds which make it impossible to construct BCC pseudopotentials for certain elements (among them the transition metals) unless one allows some of the pseudovalence wave functions to have a node remaining in the core region; in addition, even when the bounds are satisfied, it can be numerically difficult to find good potentials. However, there are many elements for which good pseudopotentials do exist, and we present results for a selection of these. Questions of transferability are also more complicated for BCC pseudopotentials than for ordinary nonlocal pseudopotentials, as we show using the results of a series of atomic and solid-state density-functional calculations. BCC potentials are usually comparable to or somewhat less transferable than ordinary nonlocal potentials, and can be *much* less transferable unless certain guidelines are observed in their construction.

With all these drawbacks, one could be forgiven for asking whether it is worth using BCC pseudopotentials at all. The answer is an emphatic yes, because BCC potentials have several great advantages to weigh against their disadvantages. The principal among these is simply that they are suitable for use with the Green's-function and diffusion QMC methods. Ordinary nonlocal pseudopotentials are not suitable and so the BCC approach is the best we have. Since the QMC method is a promising nonperturbative method for going beyond mean-field treatments of the electron-electron interaction in solids, and since some sort of pseudopotential (or damped core[10]) approach seems necessary if QMC calculations are to be practical, the BCC idea is an important one. In addition, it turns out that BCC pseudopotentials have particularly simple plane-wave matrix elements which may make them useful in conjunction with the Car-Parrinello[11] method or any other iterative scheme for solving the Kohn-Sham equations of density-functional theory. So, despite having been designed for QMC calculations, BCC pseudopotentials may have applications in ordinary density-functional electronic-structure calculations as well.

In Sec. II we explain the considerations which led BCC to propose the new type of pseudopotential (or pseudo-Hamiltonian, as perhaps it should more properly be called) for QMC calculations. We do not expect that most readers will be familiar with both pseudopotentials and QMC methods and so we have included brief intro-

ductions to the relevant aspects of these two fields in Appendixes A and B. A basic understanding of density-functional theory is assumed. In Sec. III we describe some of the properties of the BCC pseudo-Hamiltonian, including the mathematical bounds discussed earlier and the forms of the plane-wave matrix elements. In Sec. IV we discuss methods for constructing BCC pseudopotentials and give examples of potentials for Na, O, and Cu. In Sec. V we study the transferability of several different Si potentials in atoms and solids. All the calculations in Sec. V are done within density-functional theory and not using Monte Carlo methods; but since the BCC pseudopotentials were constructed within density-functional theory this is a perfectly consistent thing to do. If the BCC density-functional results do not agree with the full core density-functional results, then there will be similar (although not identical) errors in any BCC QMC calculations.

II. BACHELET, CEPERLEY, AND CHIOCCHETTI PSEUDOPOTENTIALS

A few years ago, Ceperley and Alder[12] successfully used the diffusion QMC method to study the electronic properties of jellium (an idealized solid in which the nuclear charges are smeared out into a uniform positive background). Since the valence electrons in nearly-free-electron solids behave very much like electrons in a jellium, this suggested strongly that QMC methods could be applied to real solids. This was later confirmed by the same authors,[13] who used the Green's-function QMC method to study solid hydrogen, and by Fahy, Wang, and Louie,[14] who used pseudopotentials and variational QMC (not discussed here) to calculate the binding energy, lattice parameter, and bulk modulus of diamond.

The biggest problem with diffusion and Green's-function QMC calculations is that they consume vast amounts of computer time. In particular, it turns out[15] that the time required to calculate the properties of an isolated atom scales with the fifth or sixth power of the atomic number Z, and rapidly becomes unreasonably large for Z greater than around 10. It is this atomic problem which limits the application of QMC to solids, and not any difficulties connected with arranging the atoms to make a lattice. The obvious solution is to replace the atoms by pseudoatoms with smooth pseudopotentials and no core electrons, thus reducing the effective atomic number and getting rid of the singularities in the nuclear Coulomb potential. A pseudopotential solid should be a little harder to deal with than a jellium or comparable density.

Before going on to discuss the type of pseudopotential put forward by BCC,[9] let us discuss briefly what we mean by a pseudopotential in a many-interacting-electron calculation. The full Hamiltonian for an atom with nuclear charge Z and N electrons is

$$\hat{H}=\sum_{i=1}^{N}\left[-\tfrac{1}{2}\nabla_i^2-\frac{Z}{r_i}\right]+\tfrac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1\\(j\neq i)}}^{N}\frac{1}{|\mathbf{r}_i-\mathbf{r}_j|}\ , \tag{2.1}$$

and we are hoping to replace this by a pseudo-Hamiltonian describing only the N_v valence electrons,

$$\hat{H}^{\rm ps}=\sum_{i=1}^{N_v}[-\tfrac{1}{2}\nabla_i^2+\hat{V}^{\rm ps}(\mathbf{r}_i)]+\tfrac{1}{2}\sum_{i=1}^{N_v}\sum_{\substack{j=1\\(j\neq i)}}^{N_v}\frac{1}{|\mathbf{r}_i-\mathbf{r}_j|}\ , \tag{2.2}$$

where $\hat{V}^{\rm ps}(\mathbf{r}_i)$ need not be a local potential and may involve derivatives and/or nonlocal integral operators. (We could also have altered the form of the electron-electron interaction, but that possibility will not be considered here.) Since there are no obvious many-electron equivalents of the partial norms and logarithmic derivatives used in one-electron pseudopotential theory (see Appendix A), it is not at all easy to come up with a convincing justification of the approximation involved in replacing $\hat{H}$ by $\hat{H}^{\rm ps}$. Indeed, it is not even clear how to divide a many-interacting-electron wave function into core and valence parts. But, nevertheless, we will take heart from the success of pseudopotentials in density-functional calculations (which do not, after all, provide a bad description of the gross features of the electron-electron interaction in most solids) and *assume* that the idea of a pseudopotential makes sense in many-electron theory as well. Since $\hat{V}^{\rm ps}(\mathbf{r})$ may contain differential operators, we often lump it together with the kinetic-energy operator and talk about a pseudo- (one-electron) Hamiltonian,

$$\hat{h}^{\rm ps}=-\tfrac{1}{2}\nabla^2+\hat{V}^{\rm ps}\ . \tag{2.3}$$

We now work out the most general form of the pseudo-Hamiltonian suitable for use in fixed-node Green's-function and diffusion QMC calculations. The arguments we use are basically those of BCC,[9] but their original letter was necessarily very terse and so we think it worthwhile to give a fuller explanation here.

If a pseudo-Hamiltonian,

$$\hat{H}^{\rm ps}=\sum_{i=1}^{N_v}\hat{h}^{\rm ps}(\mathbf{r}_i)+\tfrac{1}{2}\sum_{i=1}^{N_v}\sum_{\substack{j=1\\(j\neq i)}}^{N_v}\frac{1}{|\mathbf{r}_i-\mathbf{r}_j|}\ , \tag{2.4}$$

is to be useful in diffusion or Green's-function QMC calculations, it must satisfy two important criteria: it must have a non-negative imaginary-time Green's function,

$$G^{\rm ps}(\mathbf{R}',\mathbf{R};\tau)=\langle\mathbf{R}'|e^{-\hat{H}^{\rm ps}\tau}|\mathbf{R}\rangle\ , \tag{2.5}$$

and it must be compatible with the fixed-node approximation. Appendix B explains why these restrictions are necessary. The need for compatibility with the fixed-node approximation is more a practical matter than a matter of principle, but is pretty much inescapable nevertheless.

It should be clear from the discussion in Appendix B that nonlocal integral operators (such as $\hat{P}_l$, the operator which projects out the angular-momentum l component of any function on which it acts—see Appendix A) are incompatible with the fixed-node approximation. If $\hat{H}^{\rm ps}$ contains a nonlocal part, then the Schrödinger equation looks like

$$H^{\rm ps}_{\rm local}(\mathbf{R},\nabla_{\mathbf{R}})\psi(\mathbf{R})+\int H^{\rm ps}_{\rm nonlocal}(\mathbf{R},\mathbf{R}')\psi(\mathbf{R}')d\mathbf{R}' = E\psi(\mathbf{R}) , \quad (2.6)$$

and we see that the value of ψ at a point $\mathbf{R}$ depends on the values at all other points $\mathbf{R}'$, linked to $\mathbf{R}$ by $H^{\rm ps}_{\rm nonlocal}$. Unless, by some chance, all nonlocal matrix elements across nodes of the many-electron wave functions are zero, it is therefore impossible to find $\psi(\mathbf{R})$ in the regions where the trial function (see Appendix B) is positive without also involving the regions where the trial function is negative. (Perhaps symmetry arguments relating the solutions in the positive and negative regions might be useful here?) Any attempt to carry out an ordinary fixed-node QMC calculation would amount to setting all matrix elements across nodes to zero, which would neither be a variational approximation nor a very physical one.

Similar problems arise with pseudopotentials that contain higher than second-derivative operators. The simple boundary condition that $\psi=0$ on the nodal surface is then not enough to properly specify the eigenproblem in the enclosed region. Information about various normal derivatives on the nodal surface is also needed. It would probably be possible to follow a procedure analogous to that used in setting the nodes and choose these normal derivatives equal to the corresponding normal derivatives of the trial function, but this would not be an easy thing to do. At any rate, care would have to be taken since slope discontinuities across nodes can contribute nonvanishing surface terms to energy expectation values whenever the Hamiltonian contains higher than second derivatives.

There is another, more fundamental, reason why pseudopotentials involving higher than second derivatives are unacceptable, and that is because they can never have entirely non-negative imaginary-time Green's functions. We can explain this as follows.

First we define a $3N_v$-dimensional momentum operator,

$$\hat{\mathbf{P}}=\left[-i\frac{\partial}{\partial r_{1x}},-i\frac{\partial}{\partial r_{1y}},-i\frac{\partial}{\partial r_{1z}},-i\frac{\partial}{\partial r_{2x}},-i\frac{\partial}{\partial r_{2y}},\ldots,-i\frac{\partial}{\partial r_{N_v x}},-i\frac{\partial}{\partial r_{N_v y}},-1\frac{\partial}{\partial r_{N_v z}}\right]$$

corresponding to the $3N_v$-dimensional position vector, $\mathbf{R}=(r_{1x},r_{1y},r_{1z},r_{2x},r_{2y},\ldots,r_{N_v x},r_{N_v y},r_{N_v z})$. As long as it does not contain any local integral operators or (which is more or less equivalent) any infinitely high derivatives, we can then write the full many-electron pesudo-Hamiltonian [Eq. (2.4)] in the form

$$\hat{H}^{\rm ps}=C(\mathbf{R})+D_i(\mathbf{R})\hat{P}_i+E_{ij}(\mathbf{R})\hat{P}_i\hat{P}_j+\cdots +S_{ij\cdots n}(\mathbf{R})\hat{P}_i\hat{P}_j\cdots\hat{P}_n , \quad (2.7)$$

where the electron-electron interaction terms have been absorbed into $C(\mathbf{R})$. All subscripts range from 1 to N_v and there is an implied summation over these values whenever a subscript is repeated. Because the momentum operators are actually all in the one-electron part of the Hamiltonian (we *chose* not to alter the momentum-independent form of the electron-electron interaction), the tensors $E,\ldots,S$ have no nonzero elements linking components of $\hat{\mathbf{P}}$ corresponding to different electrons.

If the Green's function corresponding to the Hamiltonian in Eq. (2.7) is non-negative, then *all* non-negative wave functions must remain non-negative as they develop in imaginary time. We will show that whenever the Hamiltonian contains higher than second derivatives, there are always some initially non-negative wave functions which develop negative parts, and hence that the Green's function for such a Hamiltonian cannot be entirely non-negative.

Consider starting with a wave function that is non-negative but which has a quadratic zero at $\mathbf{R}=\mathbf{0}$. Such a wave function can be written in the form

$$\psi(\mathbf{R})=R_iR_j\psi_{ij}(\mathbf{R}) , \quad (2.8)$$

where all the eigenvalues of $\psi_{ij}(\mathbf{0})$ are greater than or equal to zero. If this wave function is to remain non-negative, then

$$\left.\frac{\partial\psi}{\partial\tau}\right|_{\mathbf{R}=0}=\{-\hat{H}^{\rm ps}[R_iR_j\psi_{ij}(\mathbf{R})]\}|_{\mathbf{R}=0} \quad (2.9)$$

must be greater than or equal to zero. Equation (2.9) is just the imaginary-time Schrödinger equation for $\psi(\mathbf{R})$ (see Appendix B) and can be simplified using the relation

$$\hat{H}^{\rm ps}R_j=R_j\hat{H}^{\rm ps}-i\partial\hat{H}^{\rm ps}/\partial\hat{P}_j , \quad (2.10)$$

which is easy enough to derive given the basic commutator, $[\hat{P}_i,R_j]=-i\delta_{ij}$. The "operator derivative," $\partial\hat{H}^{\rm ps}/\partial\hat{P}^j$, is really nothing more than the name of that function of $\mathbf{R}$ and $\hat{\mathbf{P}}$ which would be obtained by differentiating $\hat{H}^{\rm ps}(\mathbf{R},\hat{\mathbf{P}})$ if the components of $\hat{\mathbf{P}}$ were treated as ordinary independent variables rather than as operators. The Hamiltonian, Eq. (2.7), was written with all the momentum operators to the right of all the position operators and this convention must be maintained during the operator differentiation. Applying Eq. (2.10) to Eq. (2.9) gives

$$\left.\frac{\partial\psi}{\partial\tau}\right|_{\mathbf{R}=0}=i\left[\frac{\partial\hat{H}^{\rm ps}}{\partial\hat{P}_i}\right]R_j\psi_{ij}(\mathbf{R})\Bigg|_{\mathbf{R}=0} , \quad (2.11)$$

where the term with an R_i in front gives zero because of the evaluation at $\mathbf{R}=\mathbf{0}$. After a very similar manipulation, this then becomes

$$\left.\frac{\partial\psi}{\partial\tau}\right|_{\mathbf{R}=0}=\frac{\partial^2\hat{H}^{\rm ps}}{\partial\hat{P}_i\,\partial\hat{P}_j}\psi_{ij}(\mathbf{R})\Bigg|_{\mathbf{R}=0} . \quad (2.12)$$

If $\hat{H}^{\rm ps}$ involves third or higher powers of the momentum operator, then $\partial^2\hat{H}^{\rm ps}/\partial\hat{P}_i\,\partial\hat{P}_j$ has some derivatives left over. But since [at least in those cases where all eigenvalues of $\psi_{ij}(\mathbf{0})$ are definitely greater than zero] the first few derivatives of the $\psi_{ij}(\mathbf{R})$ functions can all be chosen arbitrarily, this means that $\partial\psi/\partial\tau|_{\mathbf{R}=0}$ cannot always be greater than or equal to zero. We therefore see that no pseudo-Hamiltonian involving higher than second powers of the momentum operator can have an entirely non-negative Green's function.

If $\hat{H}^{\rm ps}$ is quadratic in momentum, then $\partial^2\hat{H}^{\rm ps}/\partial\hat{P}_i\,\partial\hat{P}_j$ is just $E_{ij}(\mathbf{R})+E_{ji}(\mathbf{R})$ and Eq. (2.12) becomes

$$\left.\frac{\partial\psi}{\partial\tau}\right|_{\mathbf{R}=0}=[E_{ij}(\mathbf{0})+E_{ji}(\mathbf{0})]\psi_{ij}(\mathbf{0})\ . \tag{2.13}$$

ψ_{ij} (is any symmetric tensor which) has no negative eigenvalues, and so in this case we can be sure that $\partial\psi/\partial\tau|_{\mathbf{R}=0}\geq 0$ as long as $E_{ij}(\mathbf{0})+E_{ji}(\mathbf{0})$ also has no negative eigenvalues. This is a necessary condition for the Green's function to be non-negative but we have not shown that it is also a sufficient one. To prove sufficiency, one would have to show that the first nonvanishing imaginary-time derivative of ψ is always greater than zero wherever ψ has any sort of zero. This is easy enough in one dimension, but the tensor arithmetic gets complicated in the general case and anyway the result is clear from the explicit form of the short-time Green's function (which can be derived using the same sorts of methods as are used for the ordinary diffusion equation—see Appendix B and pages 57–60 of Ref. 16 for hints).

To summarize so far, we have shown that the most general form of the pseudo-Hamiltonian which has a non-negative Green's function and is consistent with the fixed-node approximation is

$$\hat{H}^{\rm ps}=C(\mathbf{R})+D_i(\mathbf{R})\hat{P}_i+E_{ij}(\mathbf{R})\hat{P}_i\hat{P}_j\ , \tag{2.14}$$

where the $3N_v\times 3N_v$ matrix, $E_{ij}(\mathbf{R})$, has no negative eigenvalues at any point $\mathbf{R}$. The corresponding one-electron part of this Hamiltonian is therefore of the form

$$\hat{h}^{\rm ps}=c(\mathbf{r})+[d_i(\mathbf{r})\hat{p}_i+\hat{p}_i d_i(\mathbf{r})]+\tfrac{1}{2}\hat{p}_i e_{ij}(\mathbf{r})\hat{p}_j\ , \tag{2.15}$$

where

$$\hat{\mathbf{p}}=\left(-i\frac{\partial}{\partial x},-i\frac{\partial}{\partial y},-i\frac{\partial}{\partial z}\right)$$

is the one-electron momentum operator, $e_{ij}(\mathbf{r})$ is an $\mathbf{r}$-dependent inverse effective-mass tensor which is symmetric and has no negative eigenvalues, and we have made sure that each term is explicitly Hermitian.

$\hat{H}^{\rm ps}$ is supposed to be an atomic pseudo-Hamiltonian and so the one-electron part, $\hat{h}^{\rm ps}$, must be spherically symmetric. This means that $c(\mathbf{r})$ must transform like a scalar, $d_i(\mathbf{r})$ $(i=1,2,3)$ like a vector, and $e_{ij}(\mathbf{r})$ $(i=1,2,3;j=1,2,3)$ like a second-rank tensor. The only vector we can make using only $\mathbf{r}$ and constants is $\mathbf{r}$ itself; and the only second-rank tensors are the unit tensor δ_{ij} and the dyadic $r_i r_j$. We therefore see that

$$c(\mathbf{r})=\bar{c}(r^2)\ , \tag{2.16}$$

$$d_i(\mathbf{r})=\bar{d}(r^2)r_i\ , \tag{2.17}$$

$$e_{ij}(\mathbf{r})=\bar{e}_1(r^2)\delta_{ij}+\bar{e}_2(r^2)r_i r_j\ , \tag{2.18}$$

where $\bar{c}(r^2)$, $\bar{d}(r^2)$, $\bar{e}_1(r^2)$, and $\bar{e}_2(r^2)$ depend only on the scalar r^2. The functions $\bar{e}_1(r^2)$ and $\bar{e}_2(r^2)$ must satisfy $\bar{e}_1(r^2)\geq 0$ and $\bar{e}_1(r^2)+r^2\bar{e}_2(r^2)\geq 0$ so that all the eigenvalues of $e_{ij}(\mathbf{r})$ are non-negative, but $\bar{c}(r^2)$ and $\bar{d}(r^2)$ are unrestricted. The rotationally invariant form of the pseudo-Hamiltonian is then

$$\hat{h}^{\rm ps}=\bar{c}(r^2)+[\bar{d}(r^2)\mathbf{r}\cdot\hat{\mathbf{p}}+\hat{\mathbf{p}}\cdot\mathbf{r}\bar{d}(r^2)]+\tfrac{1}{2}\hat{\mathbf{p}}\bar{e}_3(r^2)\hat{\mathbf{p}}-\tfrac{1}{2}\hat{\mathbf{L}}\bar{e}_2(r^2)\hat{\mathbf{L}}\ , \tag{2.19}$$

where $\hat{\mathbf{L}}=\mathbf{r}\times\hat{\mathbf{p}}$ is the angular-momentum operator, $\bar{e}_3(r^2)=\bar{e}_1(r^2)+r^2\bar{e}_2(r^2)$ (≥ 0), and use has been made of the identity

$$(\hat{\mathbf{p}}\cdot\mathbf{r})f(r^2)(\mathbf{r}\cdot\hat{\mathbf{p}})=\hat{\mathbf{p}}r^2f(r^2)\hat{\mathbf{p}}-\hat{\mathbf{L}}f(r^2)\hat{\mathbf{L}}\ . \tag{2.20}$$

It turns out that the terms linear in $\hat{\mathbf{p}}$ are not useful as they can be "gauged out" as follows. Suppose $\psi(\mathbf{r})$ is an eigenfunction of $\hat{h}^{\rm ps}(\mathbf{r},\hat{\mathbf{p}})$ with eigenvalue E,

$$\hat{h}^{\rm ps}(\mathbf{r},\hat{\mathbf{p}})\psi(\mathbf{r})=E\psi(\mathbf{r})\ . \tag{2.21}$$

Then if we define a gauge-transformed wave function $\tilde{\psi}(\mathbf{r})$ by

$$\psi(\mathbf{r})=e^{-i\phi(r^2)}\tilde{\psi}(\mathbf{r})\ , \tag{2.22}$$

we see that $\tilde{\psi}(\mathbf{r})$ satisfies the transformed Schrödinger equation

$$\hat{h}^{\rm ps}(\mathbf{r},\hat{\mathbf{p}}-\nabla\phi)\tilde{\psi}(\mathbf{r})=E\tilde{\psi}(\mathbf{r})\ . \tag{2.23}$$

If we choose ϕ such that

$$\frac{\partial\phi(r^2)}{\partial(r^2)}=\frac{\bar{d}(r^2)}{\bar{e}_3(r^2)}$$

then the new Hamiltonian is

$$\hat{h}^{\rm ps}(\mathbf{r},\hat{\mathbf{p}}-\nabla\phi)=\bar{c}(r^2)-\tfrac{1}{2}\bar{e}_3(r^2)|\nabla\phi(r^2)|^2+\tfrac{1}{2}\hat{\mathbf{p}}\bar{e}_3(r^2)\hat{\mathbf{p}}-\tfrac{1}{2}\hat{\mathbf{L}}\bar{e}_2(r^2)\hat{\mathbf{L}}\ , \tag{2.24}$$

which has no terms linear in $\hat{\mathbf{p}}$, but has the same eigenvalues as the original pseudo-Hamiltonian and eigenfunctions which differ only by a position-dependent phase factor. So we see that for every pseudo-Hamiltonian which does have terms linear in $\hat{\mathbf{p}}$, there is an essentially equivalent Hamiltonian with no linear terms. We lose nothing by assuming that there are no linear terms and writing the general pseudo-Hamiltonian in the form

$$\hat{h}^{\rm ps}=-\tfrac{1}{2}\nabla[1+a(r)]\nabla+\frac{b(r)}{2r^2}\hat{L}^2+V(r)\ , \tag{2.25}$$

where we have changed the notation to correspond with BCC (Ref. 9) and have used the fact that all components of $\hat{\mathbf{L}}$ commute with any function of r^2. The non-negativity constraints on $\bar{e}_3(r^2)$ and $\bar{e}_1(r^2)$ translate into

$$1+a(r)\geq 0\ , \tag{2.26a}$$

$$1+a(r)+b(r)\geq 0 \ . \tag{2.26b}$$

These inequalities make good sense since $1+a(r)$ and $1+a(r)+b(r)$ can be thought of as inverse effective radial and angular electron masses, respectively. If a Hamiltonian had regions of negative effective mass, it would not even have a lowest eigenvalue (the more oscillations a wave function had in the negative mass region, the lower would be its energy) and so would be very unphysical.

Equations (2.25), (2.26a), and (2.26b) specify the most general one-electron atomic pseudo-Hamiltonian which is spherically symmetric, has a non-negative imaginary-time Green's function, and is consistent with the usual form of the fixed-node approximation. This specification was the main theoretical result of BCC.[9]

III. BCC PSEUDOPOTENTIALS IN INDEPENDENT ELECTRON CALCULATIONS

So far, we have been considering the use of pseudopotentials in many-interacting-electron calculations. We assumed that it makes sense to replace all the effects of the core electrons on the valence electrons by a simple one-particle pseudopotential, and then we derived the most general form of the pseudopotential which can conveniently be used in diffusion and Green's-function QMC calculations. The obvious next question concerns pseudopotential construction: how do we use the results of full core atomic calculations to find good BCC pseudopotentials?

As mentioned earlier, we do not understand the proper many-electron generalizations of the one-electron logarithmic derivatives and partial norms which are so useful in constructing ordinary norm-conserving pseudopotentials (see Appendix A and Refs. 17, 18, and 19) and so do not have a good answer to this question. For lack of any better ideas, we follow Bachelet, Ceperley, and Chiocchetti[9] and choose to construct BCC pseudopotentials within density-functional theory and the local-density approximation. The assumption that such pseudopotentials can then be used in many-electron calculations is a big one (it relies, for example, on the supposition that the core-valence exchange and correlation interactions—which are absorbed into the one-electron pseudopotential—are reasonably well represented within the local-density approximation), but the work of Fahy, Wang, and Louie[14] gives some support to the idea. Pseudopotential construction within density-functional theory is also appropriate if the pseudopotentials are then to be used in density-functional calculations, and this is not unlikely given the simple forms of the plane-wave matrix elements which will be derived later in this section.

The details of the density-functional construction of BCC pseudopotentials and some examples are given in Sec. IV. In the rest of this section, we consider two other aspects of the behavior of BCC pseudopotentials in independent electron (density-functional) calculations. First we discuss some results on the ordering of the atomic one-electron eigenvalues. We show that the energy of the lowest s state must always be lower than the energy of the lowest p state, which in turn must be lower than the energy of the lowest d state, and so on. This means that (unless one is content with a pseudopotential which still binds a few core states) it is not possible to find BCC pseudopotentials for transition metals or other elements where the valence eigenvalues do not increase with increasing angular momentum. Finally, we show the forms of the plane-wave matrix elements of BCC pseudopotentials and give all the formulas needed to set up a BCC pseudopotential and plane-wave band-structure calculation.

A. Energy-level ordering in a BCC pseudopotential

If a BCC pseudopotential is used in a density-functional atomic calculation, the self-consistent one-electron eigenvalues and eigenfunctions satisfy a Schrödinger (Kohn-Sham) equation of the form

$$\left[-\tfrac{1}{2}\nabla[1+a(r)]\nabla+\frac{b(r)}{2r^2}\hat{L}^2+W_0(r)\right]\psi_i(\mathbf{r})=\epsilon_i\psi_i(\mathbf{r}) \ , \tag{3.1}$$

where $W_0(r)$ is the sum of $V(r)$ [from Eq. (2.25)] and the self-consistent Hartree and exchange-correlation potentials. The corresponding radial equation is

$$-\frac{1}{2}\frac{d}{dr}(1+a)\frac{du_l}{dr}+\left[\frac{l(l+1)}{2r^2}(1+a+b)+\frac{a'}{2r}+W_0\right]u_l =\epsilon_l u_l \ , \tag{3.2}$$

where l is the angular momentum, $a'=da/dr$, and $u_l(r)=rR_l(r)$ is r times the radial wave function. For each l, the lowest-energy radial wave function is nodeless and can be chosen to be real. We will now show that the eigenvalues corresponding to these wave functions must increase with increasing l.

The nodeless solutions u_{l_1} and u_{l_2} for two different angular momenta l_1 and l_2 satisfy

$$-\frac{1}{2}\frac{d}{dr}(1+a)\frac{du_{l_1}}{dr}+\left[\frac{a'}{2r}+W_0\right]u_{l_1} +\frac{l_1(l_1+1)}{2r^2}(1+a+b)u_{l_1}=\varepsilon_{l_1}u_{l_1} \ , \tag{3.3a}$$

$$-\frac{1}{2}\frac{d}{dr}(1+a)\frac{du_{l_2}}{dr}+\left[\frac{a'}{2r}+W_0\right]u_{l_2} +\frac{l_2(l_2+1)}{2r^2}(1+a+b)u_{l_1}=\varepsilon_{l_2}u_{l_2} \ . \tag{3.3b}$$

Multiply Eq. (3.3a) by u_{l_2} and Eq. (3.3b) by u_{l_1}, subtract and integrate from $r=0$ to infinity to get

$$\int_0^\infty\left[-\frac{1}{2}\frac{d}{dr}\left[u_{l_2}(1+a)\frac{du_{l_1}}{dr}\right]+\frac{1}{2}\frac{d}{dr}\left[u_{l_1}(1+a)\frac{du_{l_2}}{dr}\right] +\frac{l_1(l_1+1)-l_2(l_2+1)}{2r^2}(1+a+b)u_{l_1}u_{l_2}\right]dr =(\varepsilon_{l_1}-\varepsilon_{l_2})\int_0^\infty u_{l_1}u_{l_2}\,dr \ . \tag{3.4}$$

As long as $a(r)$, $b(r)$, and $V(r)$ are all regular at the origin and satisfy $1+a+b>0$ and $1+a>0$ for all r, it is not hard to show that $u_l(r)\sim r^\alpha$ for small r, with $\alpha=1$ for $l=0$ and $\alpha>1$ otherwise. This means that $u_{l_1}(1+a)(du_{l_2}/dr)$ and $u_{l_2}(1+a)(du_{l_1}/dr)$ are both zero at $r=0$ as well as at $r=\infty$, and so the first two contributions to the integral on the left-hand side of Eq. (3.4) vanish and we get

$$\varepsilon_{l_1}-\varepsilon_{l_2}=[l_1(l_1+1)-l_2(l_2+1)] \times \frac{\int_0^\infty \frac{1+a+b}{2r^2}u_{l_1}u_{l_2}\,dr}{\int_0^\infty u_{l_1}u_{l_2}\,dr} . \tag{3.5}$$

The two integrals in Eq. (3.5) have the same sign because both u_{l_1} and u_{l_2} are nodeless and $1+a+b>0$. It therefore follows that $\varepsilon_{l_1}\geq\varepsilon_{l_2}$ if $l_1>l_2$.

In real atoms, the energies of the important valence levels are not always ordered like this. Although the full core Hamiltonian has the BCC form (with $a=b=0$ and W_0 the full self-consistent atomic potential), the different valence states may have different numbers of radial nodes and so their energies do not have to increase with increasing l. A good example is copper, where the valence states (in ascending energy order) are the $3d$, $4s$, and $4p$ states, with 0, 3, and 2 radial nodes, respectively. The $3d$ and $4s$ states are actually very close in energy, with the extra angular $[l(l+1)]$ kinetic energy of the $3d$ state being offset by the extra radial kinetic energy of the $4s$ state (which has more radial nodes).

In any coreless pseudocopper atom, the difference in angular kinetic energies would remain, but both the d and s radial functions would be nodeless and so the radial kinetic-energy difference would not be so large. Reducing $1+a+b$ as much as possible would reduce the angular kinetic-energy difference and so bring the d level down closer to the s level, but $1+a+b$ always has to remain positive and so there is a limit to the advantage that can be gained this way. The result, as Eq. (3.5) shows, is that it is never possible to pull the d level down below the s level and so it is never possible to find a good coreless pseudopotential for copper.

The only way around this problem is to allow the pseudopotential to have some remaining core states. If one insists that the logarithmic derivatives of the full core $3d$, $4s$, and $4p$ states match the logarithmic derivatives of the pseudopotential $3d$, $2s$, and $3p$ states (rather than the $3d$, $1s$, and $2p$ states as usual), then one can find a reasonable BCC pseudopotential for copper. The cost is that the pseudoatom then has $1s$ and $2p$ core states which do not have much physical meaning but need to be carried along throughout all calculations anyway. This is disappointing, since the whole purpose of BCC pseudopotentials was to allow QMC calculations without any core electrons, but at least it is still far better than having to solve the full core atom.

B. Plane-wave matrix elements

If BCC pseudopotentials are to be used in density-functional calculations in place of ordinary nonlocal pseudopotentials,[17–19] it is necessary to work out their plane-wave matrix elements. It turns out that these have very simple forms which may be advantageous in iterative band-structure methods such as the one pioneered by Car and Parrinello.[11]

In any density-functional plane-wave pseudopotential calculation, irrespective of the kind of pseudopotential used, the one-electron eigenvalues at each $\mathbf{k}$ point are given by solving a secular equation of the form

$$\sum_{\mathbf{G}'} H^{\rm ps}(\mathbf{k};\mathbf{G},\mathbf{G}')c_k(\mathbf{G}')=E(\mathbf{k})c_{\mathbf{k}}(\mathbf{G}) , \tag{3.6}$$

where $\mathbf{G}$ and $\mathbf{G}'$ are reciprocal-lattice vectors and $E(\mathbf{k})$ is an eigenvalue at point $\mathbf{k}$ in the Brillouin zone. The corresponding Bloch eigenfunctions are then

$$\psi_{\mathbf{k}}(\mathbf{r})=\sum_{\mathbf{G}} c_{\mathbf{k}}(\mathbf{G})e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}} . \tag{3.7}$$

The matrix elements are given by[20]

$$H^{\rm ps}(\mathbf{k};\mathbf{G},\mathbf{G}')=\tfrac{1}{2}(\mathbf{k}+\mathbf{G}')^2\delta_{\mathbf{G},\mathbf{G}'}+V_{\rm sc}(\mathbf{G}-\mathbf{G}') +\frac{1}{\Omega_C}\sum_{j=1}^{n_{\rm type}} S^j(\mathbf{G}-\mathbf{G}')V_j^{\rm ps}(\mathbf{k};\mathbf{G},\mathbf{G}') , \tag{3.8}$$

where $V_{\rm sc}(\mathbf{G}-\mathbf{G}')$ is a Fourier component of the Hartree and local exchange-correlation potential due to the valence electrons, Ω_C is the volume of the unit cell, $S^j(\mathbf{G}-\mathbf{G}')$ is the structure factor for atoms of type j, and

$$V_j^{\rm ps}(\mathbf{k};\mathbf{G},\mathbf{G}')=\int e^{-i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}}\hat{V}_j^{\rm ps}(\mathbf{r}) \times e^{i(\mathbf{k}+\mathbf{G}')\cdot\mathbf{r}}d^3r , \tag{3.9}$$

where $\hat{V}_j^{\rm rs}(\mathbf{r})$ is the bare atomic pseudopotential operator for an atom of type j. The form of the pseudopotential matrix element, $V_j^{\rm ps}(\mathbf{k};\mathbf{G},\mathbf{G}')$, is the only thing which depends on the type of pseudopotential used. Everything else (including the procedures for dealing with the infinities due to the Coulomb interaction, the definition of the "α energy," etc.) is exactly as described in Ref. 20.

For BCC pseudopotentials, $\hat{V}^{\rm ps}(\mathbf{r})$ takes the form

$$\hat{V}^{\rm ps}(\mathbf{r})=-\tfrac{1}{2}\nabla a(r)\nabla+\frac{1}{2r^2}b(r)\hat{L}^2+V(r) , \tag{3.10}$$

where $\hat{\mathbf{L}}=-i(\mathbf{r}\times\nabla)$ is the angular-momentum operator. The expression for the plane-wave matrix elements of $V(r)$ will not be explained since it is exactly the same as for any local potential. However, the other terms are less familiar and so we will sketch out the derivation of the form of the $b\hat{L}^2$ matrix elements. The steps needed to deal with the $\nabla a\nabla$ term are almost the same.

We would like to evaluate

$$I_b(\mathbf{k};\mathbf{G},\mathbf{G}')=\int e^{-i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}}\hat{\mathbf{L}}\frac{b(r)}{2r^2}\hat{\mathbf{L}}e^{i(\mathbf{k}+\mathbf{G}')\cdot\mathbf{r}}d^3r , \tag{3.11}$$

where $b(r)$ vanishes outside the core region. Because $\hat{\mathbf{L}}$

is Hermitian, this becomes

$$I_b(\mathbf{k};\mathbf{G},\mathbf{G}')=\int(\hat{\mathbf{L}}e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}})^*\frac{b(r)}{2r^2}(\hat{\mathbf{L}}e^{i(\mathbf{k}+\mathbf{G}')\cdot\mathbf{r}})d^3r \tag{3.12a}$$

$$=\int[\mathbf{r}\times(\mathbf{k}+\mathbf{G})]\cdot[\mathbf{r}\times(\mathbf{k}+\mathbf{G}')]\frac{b(r)}{2r^2}\times e^{i(\mathbf{G}'-\mathbf{G})\cdot\mathbf{r}}d^3r \tag{3.12b}$$

$$=\tfrac{1}{2}(\mathbf{k}+\mathbf{G})^T\cdot\mathbf{b}(\mathbf{G}-\mathbf{G}')\cdot(\mathbf{k}+\mathbf{G}')\ , \tag{3.12c}$$

where

$$\mathbf{b}(\mathbf{G}-\mathbf{G}')=\int(r^2\mathbf{I}-\mathbf{r}\cdot\mathbf{r}^T)\frac{b(r)}{r^2}e^{i(\mathbf{G}'-\mathbf{G})\cdot\mathbf{r}}d^3r\ , \tag{3.13}$$

is $\mathbf{k}$ independent and the last step used the identity

$$(\mathbf{r}\times\mathbf{A})\cdot(\mathbf{r}\times\mathbf{B})=r^2(\mathbf{A}\cdot\mathbf{B})-(\mathbf{r}\cdot\mathbf{A})(\mathbf{r}\cdot\mathbf{B})\ .$$

If we choose the z axis along $\mathbf{G}'-\mathbf{G}$, it is clear that the tensor $\mathbf{b}$ has only two independent components, $b_{zz}=b_\parallel$ and $b_{xx}=b_{yy}=b_\perp$, where

$$b_\parallel(|\mathbf{G}-\mathbf{G}'|)=\int(x^2+y^2)\frac{b(r)}{r^2}e^{i|\mathbf{G}'-\mathbf{G}|z}d^3r \tag{3.14a}$$

and

$$b_\perp(|\mathbf{G}-\mathbf{G}'|)=\int[\tfrac{1}{2}(x^2+y^2)+z^2]\frac{b(r)}{r^2}e^{i|\mathbf{G}'-\mathbf{G}|z}d^3r\ . \tag{3.14b}$$

The angular parts of these integrals can be done analytically, leaving only radial integrals (see below for the expressions) involving various spherical Bessel functions. The full $\mathbf{b}$ tensor in a general coordinate system can then be expressed in terms of the two independent components, using

$$\mathbf{b}(\mathbf{G}-\mathbf{G}')=b_\perp(|\mathbf{G}-\mathbf{G}'|)\mathbf{I}+\mathbf{g}[b_\parallel(|\mathbf{G}-\mathbf{G}'|)-b_\perp(|\mathbf{G}-\mathbf{G}'|)]\mathbf{g}^T\ , \tag{3.15}$$

where $\mathbf{g}$ is a unit vector in the $\mathbf{G}-\mathbf{G}'$ direction.

Having sketched the derivation of the form of the plane-wave matrix elements of the $b\hat{L}^2$ part of the pseudopotential, and stated that the matrix elements of the $\nabla a\nabla$ part can be evaluated in a similar way, we now just list all the formulas needed to work out $V^{\rm ps}(\mathbf{k};\mathbf{G},\mathbf{G}')$ given $a(r)$, $b(r)$, and $V(r)$,

$$V^{\rm ps}(\mathbf{k};\mathbf{G},\mathbf{G}')=\int e^{-i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}}\left[-\tfrac{1}{2}\nabla a\nabla+\frac{1}{2r^2}b\hat{L}^2+V\right]\times e^{i(\mathbf{k}+\mathbf{G}')\cdot\mathbf{r}}d^3r \tag{3.16}$$

$$=\tfrac{1}{2}(\mathbf{k}+\mathbf{G})^T\cdot\mathbf{F}(\mathbf{G}-\mathbf{G}')\cdot(\mathbf{k}+\mathbf{G}')+V(|\mathbf{G}-\mathbf{G}'|)\ , \tag{3.17}$$

where

$$\mathbf{F}(\mathbf{G}-\mathbf{G}')=[a(|\mathbf{G}-\mathbf{G}'|)+b_\perp(|\mathbf{G}-\mathbf{G}'|)]\mathbf{I}+\mathbf{g}[b_\parallel(|\mathbf{G}-\mathbf{G}'|)-b_\perp(|\mathbf{G}-\mathbf{G}'|)]\mathbf{g}^T\ , \tag{3.18}$$

$$V(|\mathbf{G}-\mathbf{G}'|)=\int_{r=0}^{\infty}V(r)j_0(|\mathbf{G}'-\mathbf{G}|r)4\pi r^2dr\ , \tag{3.19a}$$

$$a(|\mathbf{G}-\mathbf{G}'|)=\int_{r=0}^{\infty}a(r)j_0(|\mathbf{G}'-\mathbf{G}|r)4\pi r^2dr\ , \tag{3.19b}$$

$$b_\perp(|\mathbf{G}-\mathbf{G}'|)=\int_{r=0}^{\infty}b(r)[\tfrac{2}{3}j_0(|\mathbf{G}'-\mathbf{G}|r)-\tfrac{1}{3}j_2(|\mathbf{G}'-\mathbf{G}|r)]4\pi r^2dr\ , \tag{3.19c}$$

$$b_\parallel(|\mathbf{G}-\mathbf{G}'|)=\int_{r=0}^{\infty}b(r)[\tfrac{2}{3}j_0(|\mathbf{G}'-\mathbf{G}|r)+\tfrac{2}{3}j_2(|\mathbf{G}'-\mathbf{G}|r)]4\pi r^2dr\ , \tag{3.19d}$$

and j_0 and j_2 are spherical Bessel functions.

For a band-structure calculation, the V, a, $b_\perp$, and $b_\parallel$ functions in Eqs. (3.19a)–(3.19d) are worked out beforehand, and this is only a little harder than Fourier transforming four spherically symmetric local potentials. The evaluation of any given pseudopotential matrix element, $V^{\rm ps}(\mathbf{k};\mathbf{G},\mathbf{G}')$, is then a simple matter involving just a few multiplications.

In iterative schemes, such as the Car-Parrinello method,[11] one is interested in evaluating expressions such as

$$d_k(\mathbf{G})=\sum_{\mathbf{G}'=1}^{N_G}V^{\rm ps}(\mathbf{k};\mathbf{G},\mathbf{G}')c_k(\mathbf{G}') \tag{3.20}$$

to find the $N_{\mathbf{G}}$-component vector $d_k(\mathbf{G})$ given the $N_{\mathbf{G}}$-component vector $c_k(\mathbf{G})$. When using ordinary nonlocal pseudopotentials, such an operation requires of order $N_{\mathbf{G}}^2$ multiplications. With BCC pseudopotentials, however, one can use fast-Fourier-transform techniques to evaluate all the convolutions and only a few times $N_{\mathbf{G}}\ln N_{\mathbf{G}}$ multiplications are needed. Iterative band-structure calculations using BCC pseudopotentials should therefore be quicker than similar calculations using conventional nonlocal pseudopotentials.

IV. BCC PSEUDOPOTENTIAL CONSTRUCTION AND EXAMPLES

In principle, the construction of BCC pseudopotentials within local-density-functional theory looks fairly straightforward; in practice, it does not turn out that way. The first step is obviously to solve the full core atom. Relativistic effects may not be negligible in heavy atoms because of the singular nature of the Coulomb potential near the nucleus, and so we partially include them by using a scalar relativistic radial equation.[21–23] (In the pseudopotential atom, the relativistic effects are absorbed into the phase shifts and the ordinary radial equation suffices.) Since we will not consider f electron metals, all the valence states will be s, p, and d states and the pseudopotential ought to have the correct scattering properties for $l=0$, 1, and 2 (although the details of the $l=2$

scattering are not very important for first row elements).

We choose initial reference energies ε_l and core radii r_{cl}, according to Hamann's[24] prescription. The reference energies remain unaltered throughout the calculation, but it is often necessary to increase some or all of the core radii before reasonable BCC pseudopotentials can be found. For angular momenta corresponding to occupied valence states, the reference energies are equal to the corresponding atomic eigenvalues; for other angular momenta, the highest among the occupied valence eigenvalues is used. The reference energies for unoccupied states do not correspond to atomic or ionic bound states with the same angular momentum, and so the radial wave functions from which the partial norms and logarithmic derivatives are calculated diverge at large r.

The most obvious approach to BCC pseudopotential construction is the direct analogue of the Kerker[18] (see Appendix A) scheme for ordinary nonlocal norm-conserving pseudopotentials. One simply invents smooth and nodeless pseudo-wave-functions u_s, u_p, and u_d (u_l is actually r times the radial pseudo-wave-function in our notation), and then "inverts" the pseudoradial equation to find the pseudopotential which would have generated them. The lth pseudo-wave-function u_l must be the same as the true wave function for $r \geq r_{cl}$, and should have the same partial norm

$$N_l = \frac{1}{u_l^2(r_{cl})} \int_0^{r_{cl}} u_l^2(r_{cl}) dr \ , \tag{4.1}$$

and the same logarithmic derivative at r_{cl}. [The relation between the partial norm and the energy derivative of the logarithmic derivative, Eq. (A9), works exactly as for ordinary nonlocal potentials).] In order to be sure that neither the pseudopotentials generated nor their first radial derivatives have step discontinuities at r_{cl}, it is also necessary to insist that u_l''/u_l and u_l'''/u_l match the corresponding derivatives of the true wave function there.

For ordinary nonlocal pseudopotentials, the "inversion" of the radial equation to find the pseudopotential from the wave functions can be done separately for each l and requires nothing more than some simple algebra [see Eq. (A7)]. For BCC pseudopotentials, however, it is necessary to solve three simultaneous differential equations,

$$-\frac{1}{2}\left[u_s' - \frac{u_s}{r}\right]\frac{dA}{dr} - \tfrac{1}{2}u_s'' A = (\varepsilon_s - V)u_s \ , \tag{4.2a}$$

$$-\frac{1}{2}\left[u_p' - \frac{u_p}{r}\right]\frac{dA}{dr} - \tfrac{1}{2}u_p'' A + \frac{2u_p}{2r^2}B = (\varepsilon_p - V)u_p \ , \tag{4.2b}$$

$$-\frac{1}{2}\left[u_d' - \frac{u_d}{r}\right]\frac{dA}{dr} - \tfrac{1}{2}u_d'' A + \frac{6u_d}{2r^2}B = (\varepsilon_d - V)u_d \ , \tag{4.2c}$$

for the functions $A = 1 + a$, $B = 1 + a + b$, and V. When r is greater than the largest of the three core radii, all the pseudo-wave-functions are equal to the true wave functions and so V is equal to the full self-consistent potential and $A = B = 1$. Inside the core, however, V should be smoother and weaker than the full potential and a and b will be nonzero.

Faced with the simultaneous equations (4.2a)–(4.2c), one first eliminates B and V to get a first-order differential equation for A,

$$-\frac{1}{2}\left[\frac{2u_s'}{u_s} - \frac{3u_p'}{u_p} + \frac{u_d'}{u_d}\right]\frac{dA}{dr} - \frac{1}{2}\left[\frac{2u_s''}{u_s} - \frac{3u_p''}{u_p} + \frac{u_d''}{u_d}\right]A = (2\varepsilon_s - 3\varepsilon_p + \varepsilon_d) \ . \tag{4.3}$$

This can be solved by integrating inwards from the largest of the three core radii, r_c [$= \max(r_{cs}, r_{cp}, r_{cd})$], with the boundary condition that $A(r_c) = 1$. The A potential can then be substituted back into Eqs. (4.2b) and (4.2c) to find B and V. In general, the A and B potentials will not satisfy the "positivity" conditions, $A \geq 0$ and $B \geq 0$ for all r, and so will not be acceptable. It is therefore necessary to resort to some sort of iterative procedure during which the input pseudo-wave-functions are varied (subject to the conditions on their partial norms and derivatives) until acceptable A and B potentials are found. Because of the energy-level ordering restriction discussed in Sec. III, it is clear that acceptable solutions do not always exist. But at least this procedure will enable us to find out whether they exist or not, and to find them if they do.

Or so it seems. The problem is that for most reasonable choices of pseudo-wave-functions, the coefficient of the dA/dr term in Eq. (4.3) passes through zero somewhere in the core region. (We found zeros in all the examples we looked at, but apparently[25] it is sometimes possible to choose the reference energies so as to avoid them.) The differential equation therefore has a singular point and the inward integration almost inevitably produces an $A(r)$ which diverges there. Pseudopotentials with divergences are unacceptable, and so the direct inversion procedure is no good.

It is worth examining this problem a little more closely. If the coefficient of the dA/dr term in Eq. (4.3) vanishes at $r = r_0$ then there are two possibilities: either $A(r_0)$ satisfies

$$A(r_0) = \left.\frac{-2(2\varepsilon_s - 3\varepsilon_p + \varepsilon_d)}{\left[\frac{2u_s''}{u_s} - \frac{3u_p''}{u_p} + \frac{u_d''}{u_d}\right]}\right|_{r=r_0} \ , \tag{4.4}$$

or dA/dr diverges at r_0. Equation (4.4) can be viewed as an extra, "internal," boundary condition which must be satisfied if $A(r)$ is to be well behaved and the singular point is to be rendered harmless. If this boundary condition had been imposed, then Eq. (4.3) could have been integrated outwards or inwards from the singular point to obtain a well-behaved $A(r)$. The solution would not, in general, have satisfied the external boundary condition [$A(r) = 1$ at $r = r_c$], but this is not an impossibility and so the presence of a singular point does not *necessarily* preclude the existence of well-behaved pseudopotentials.

Indeed, the eigenfunctions of most of the BCC pseudopotentials which will be constructed (by other methods) later on do produce harmless singular points of this type. The problem with direct inversion, then, is not the existence of singular points, but the difficulty of guaranteeing that they are harmless ones and that the internal and external boundary conditions are consistent. Some sort of iterative procedure might be designed to accomplish this, but we decided it was probably easier to take an altogether different approach.

In fact, we considered two different approaches, which are summarized in Figs. 1(a) and 1(b). In method 1, the chosen inputs are the two most important pseudo-wavefunctions (the s and p wave functions in most cases) and the A potential (≥ 0). The B and V potentials are then obtained from Eqs. (4.2), which takes a little algebra but does not mean solving any differential equations. An integration of the pseudoradial equation then gives the third pseudo-wave-function. The inputs are varied until $B(r) \geq 0$ and the partial norm and logarithmic derivative of the third pseudo-wave-function are close enough to the required values. A pseudopotential which does not have $B \geq 0$ does not even have a lowest eigenvalue and so is useless, but a pseudopotential with slightly the wrong scattering properties in the third (and least important) angular-momentum channel may still be useful. It is therefore sensible to impose the $B \geq 0$ constraint first, and only to worry about optimizing the scattering in the third angular-momentum channel when the B constraint is satisfied. If it proves impossible to satisfy the B constraint then no BCC pseudopotential exists; if the B constraint is satisfied but the scattering in the third l channel is not perfect, then numerical tests are needed to see just how good or bad the pseudopotential is.

In method 2 [Fig. 1(b)], the inputs are the A, B, and V potentials. These are varied subject to the constraints ($A \geq 0$, $B \geq 0$) until the scattering properties in all three angular-momentum channels are acceptable. This method has the advantage of simplicity, since the search over inputs involves optimizing only scattering properties rather than a confusing mixture of scattering properties and "hard-wall" constraints ($B \geq 0$) as in method 1. However, it is necessary to optimize more quantities simultaneously (six rather than two and a hard-wall constraint) and to solve the radial equation more times each iteration, and so it is not clear which method is the best. For reasons which are mostly historical (we started by trying direct inversion), we chose method 1; Bosin and Bachelet[26] are currently implementing method 2. Since both methods require complicated optimizations to find input functions which generate good pseudopotentials, it is important to use efficient representations of those functions. For method 1, we need the two input wave functions, $u_{l_1}(r)$ and $u_{l_2}(r)$ (on $0 \leq r \leq r_{cl_1}$ and $0 \leq r \leq r_{cl_2}$, respectively) and the input $A(r)$ potential (on $0 \leq r \leq r_{cA}$, where r_{cA} is the chosen cutoff radius beyond which the A potential is set equal to 1 and is usually set equal to the largest of the two wave-function cutoff radii). The wave functions must have the correct values of $u_l'/u_l|_{r_{cl}}$, $u_l''/u_l|_{r_{cl}}$, and $u_l'''/u_l|_{r_{cl}}$, and should satisfy the norm-conservation condition, Eq. (4.1). The A potential should be equal to 1 at r_{cA} and should have at least one derivative equal to zero there.

We write the two input wave functions for $r \leq r_{cl}$ in the form

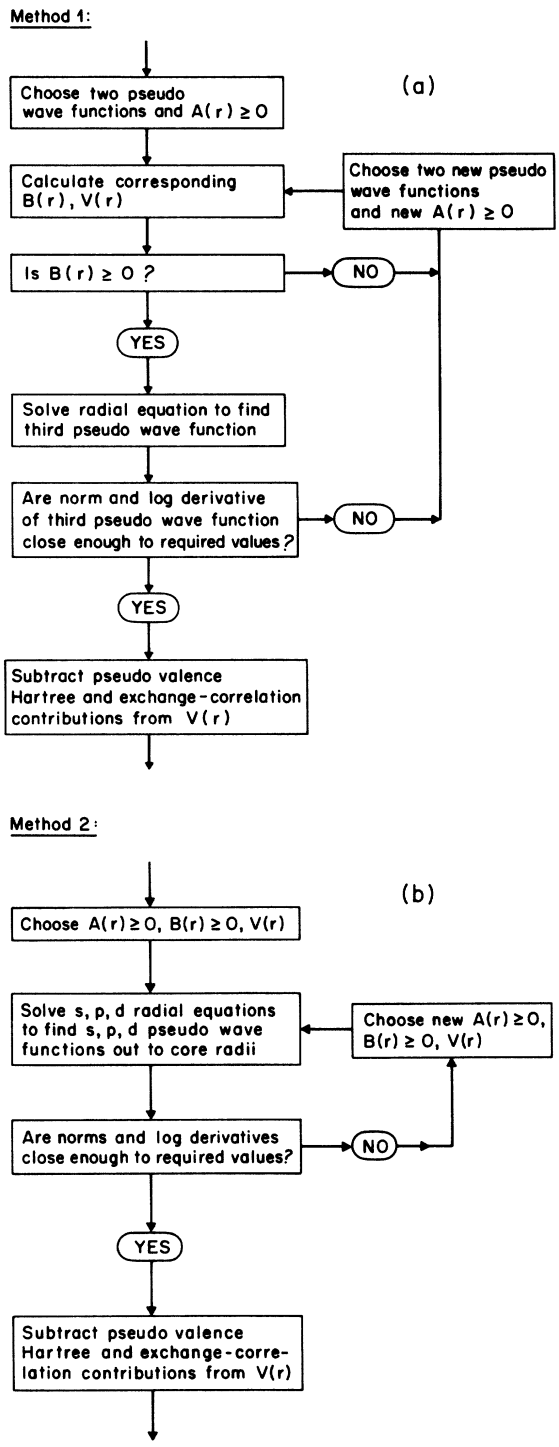

FIG. 1. (a) and (b) Flow diagrams showing two possible methods for generating BCC pseudopotentials.

$$u_l(r)=r^{l+1}S_l(r)\ , \tag{4.5}$$

where $S_l'(0)=0$. In order to ensure that $u_l(r)$ is explicitly positive (nodeless), we then set

$$S_l^2(r)=\alpha_l^2+P_l^2(r)\ , \tag{4.6}$$

where α_l^2 is a small positive constant (typically between 0.05^2 and 0.1^2) and $P_l(r)$ is a polynomial. We use a similar representation to ensure that $A(r)$ is positive on $0\le r\le r_{cA}$,

$$A^2(r)=\alpha_A^2+P_A^2(r)\ , \tag{4.7}$$

and impose the boundary condition $A'(0)=0$.

This choice of boundary conditions at the origin [$u_l\sim r^{l+1}$, $S_l'(0)=0$, $A'(0)=0$] is only one of many possible choices, but it is a convenient one and the results do not seem to depend very much on the particular choice made.[26] [Our boundary conditions imply that $b(r)\sim r^2$ as $r\to 0$ and so the mass tensor becomes isotropic near the origin.] In fact, to obtain smoother pseudopotentials, we often insist that higher derivatives of $S_l(r)$ and $A(r)$ must be zero at the origin as well.

The polynomials P_l and P_A are expanded in Chebyshev series,

$$P(r)=\sum_{i=0}^{n} c_i T_i(x)\ , \tag{4.8}$$

where $T_i(x)$ is the ith Chebyshev polynomial, and $x=2r/r_{\rm cut}-1$ (where $r_{\rm cut}$ is r_{cl} or r_{cA}, whichever is appropriate) lies between -1 and 1. It is the various c_i coefficients which are the "coordinates" of the representations and which are varied during the optimization. The boundary conditions at $x=-1$ and $+1$ translate into linear relations among the c_i coefficients and can be used to reduce the numbers of independent variables. The normalization condition then imposes a simple quadratic constraint on the remaining coefficients for each wave function.

Our experience suggests that there is little advantage to be gained by using very high-order Chebyshev expansions. If the function to be expanded is subject to m constraints (m is usually between 2 and 7), then we rarely use a Chebyshev series with more than $m+10$ coefficients. The full optimization involves three functions (two input wave functions and the A potential) and no more than about 30 independent variables altogether. Typical values are only 10 to 15.

Before starting the optimization, it is necessary to choose initial P_l and P_A polynomials and we usually

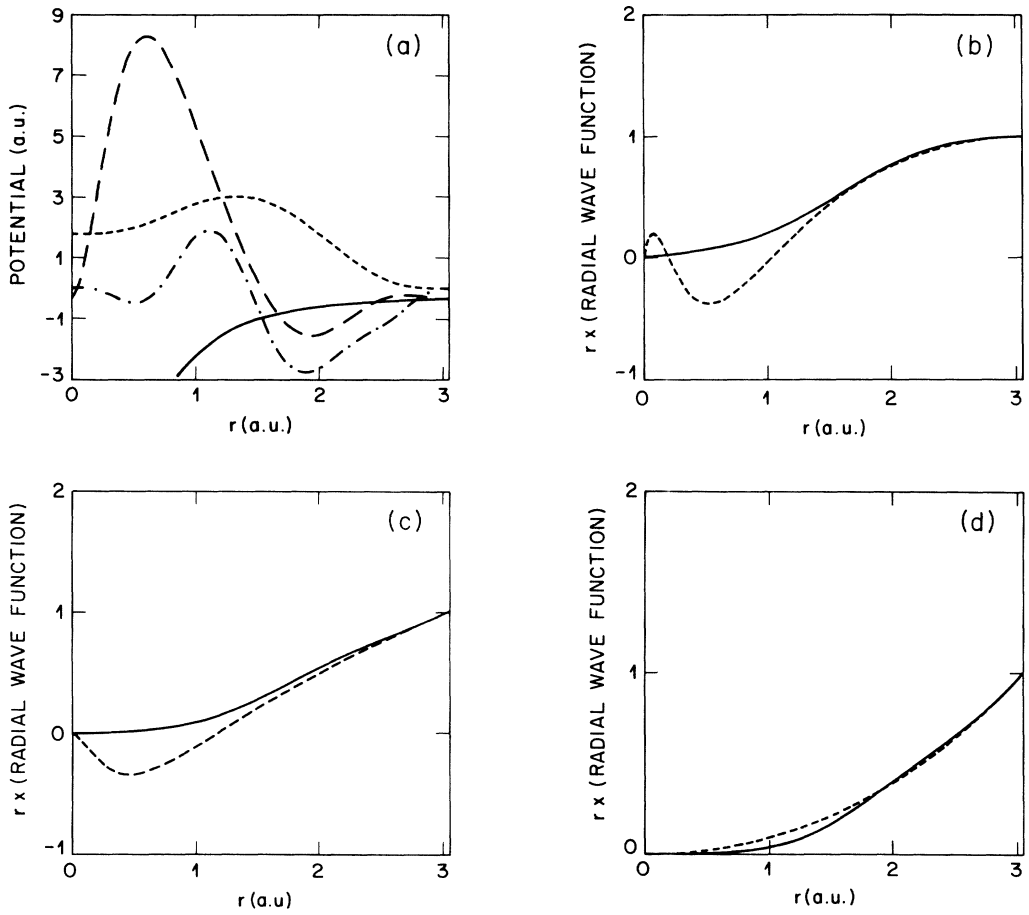

FIG. 2. (a) BCC pseudopotential for Na. Solid line, full potential; long-dashed line, $V(r)$ potential; short-dashed line, $a(r)$ potential; dashed-dotted line, $b(r)$ potential. (b)–(d) $l=0$, 1, and 2 full core and pseudo-wave-functions for the Na pseudopotential shown in (a). Solid line, pseudo-wave-function; dashed line, full core wave function.

choose the lowest-order polynomials that can be made to satisfy all the boundary and normalization conditions. [The initial A potential is thus a constant, $A(r)=1$.] The c_i coefficients are then changed gradually: first to try and find a region of configuration space where $B(r)\geq 0$; and then (if the first part succeeds) to optimize the partial norm and logarithmic derivative of the third (noninput) wave function. During the first part of this procedure, we look for regions where the cost function

$$\int_0^{r_c}[|B(r)|-B(r)]dr \ , \tag{4.9}$$

is zero; and during the second part, we minimize the sum of the squares of the proportional errors in the partial norm and logarithmic derivative subject to the condition that the cost function in Eq. (4.9) remains equal to zero.

Having little idea of the complexity of the cost functions we were trying to optimize, we initially used a simulated annealing method. However, we found that (although there were often several minima) the best and smoothest solutions usually lay straight downhill from the starting guesses, and so we now use a downhill simplex method which is much quicker (but still easy to adapt to the complicated "hierarchical" optimization required).

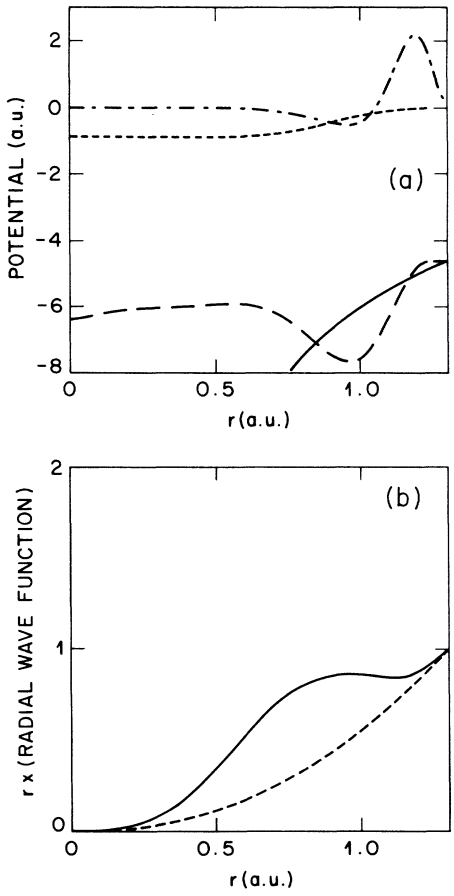

FIG. 3. (a) BCC pseudopotential for O. Solid line, full potential; long-dashed line, $V(r)$ potential; short-dashed line, $a(r)$ potential; alternating dashed line, $b(r)$ potential. (b) $l=2$ full core and pseudo-wave-functions for the O pseudopotential shown in (a). Solid line, pseudo-wave-function; dashed line, full core wave function.

In Fig. 2(a) we show a, b, and (bare) V BCC pseudopotentials for sodium. The first thing to notice is that the outermost core radius, $r_c=3.05a_0$, is chosen considerably larger than would be usual for an ordinary nonlocal pseudopotential (although the discrepancy is not as large as it may at first appear, because r_c is here defined as the radius beyond which the true and pseudo-wave-functions are identical, which is somewhat larger than the radius which Bachelet, Hamann, and Schluter[19] choose to call r_c). Half the interatomic distance in sodium metal is $3.5a_0$ and so this pseudopotential may still work reasonably well, but transferability and core-valence exchange and correlation problems[27] are bound to be increased by having the core radius so large. It is often the case that it is necessary to choose large core radii in order to find reasonable BCC pseudopotentials, and the transferability of such potentials must always be suspect and should be tested.

The second thing to notice is that the pseudopotentials have large oscillations even though the pseudo-wave-functions [shown in Figs 2(b), 2(c), and 2(d)] all look fairly smooth. This is because a, b, and V are related (via the radial equation) to up to second derivatives of the pseudo-wave-functions, so that small oscillations in the wave functions correspond to large oscillations in the potentials. Very different looking pseudopotentials can produce very similar sets of pseudo-wave-functions, and the features of any one pseudopotential are as much a reflection of the construction procedure as of the physics. These oscillations in the pseudopotentials do not matter very much in QMC calculations, but are undesirable in plane-wave band-structure calculations. However, it would not be very difficult to include an extra cost function (probably the one used in Ref. 28) in the optimization to ensure that the pseudo-wave-functions have better plane-wave expansion properties and hence that the BCC potentials generated are "softer."

Figures 3(a) and 3(b) show an oxygen pseudopotential and its d wave function (the s and p wave functions are smooth and well behaved). Oxygen is a very nonlocal element and we did not succeed in reducing the errors in the d wave-function partial norm and logarithmic derivative to zero as we did for sodium. The logarithmic derivative was 1.48 instead of the correct value of 1.79, and the partial norm was 0.50 instead of 0.24. These are large errors, but one would not expect d scattering to be very important for oxygen (which is in the first row of the Periodic Table, after all) and so the pseudopotential should still be useful. The core radius, $r_c=1.3a_0$, is also large (half the interatomic distance in O_2 is $1.15a_0$) which may be more of a problem. After the discussion of the energy-level ordering constraint in Sec. III, it should not come as a surprise that $1+a+b$ is small throughout most of the core (as is $1+a$; all the kinetic-energy terms are small).

Figures 4(a)–4(d) show a copper pseudopotential and its s, p, and d valence wave functions. As explained earlier, the only way to construct a copper pseudopotential

is to allow the s and p pseudovalence wave functions [Figs. 5(b) and 5(c)] each to have one remaining node so that the pseudopotential binds s and p core states. The core radius is then very small, which is an advantage, but the pseudopotential is very hard and the wave functions cannot be expanded in a reasonably small number of plane waves. The least important angular momentum in this case is $l=1$, and so the $l=0$ and $l=2$ wave functions were taken as input. However, there was little difficulty optimizing the p partial norm and logarithmic derivative and the scattering properties of the pseudopotential are correct in all three important angular-momentum channels.

V. TRANSFERABILITY

In this section we will describe a series of atomic and solid-state transferability tests for three different silicon pseudopotentials, Si I, Si II, and Si III. The third potential, Si III, gives the correct partial norm and logarithmic derivative at the reference energy in all three ($l=0,1,2$) angular-momentum channels, but neither Si I nor Si II scatters correctly for $l=2$. The $l=2$ partial norms of Si I and Si II are both 0.70 instead of the correct value of 0.75, and the logarithmic derivatives are both 0.63 instead of 0.60. The results calculated using these two "two-channel" pseudopotentials are included because they clarify transferability issues which are not obvious from the "three-channel" results alone. Also, on a more practical level, the two-channel potentials are easier to converge in plane waves.

Before going on to look at particular cases, let us consider the factors which affect transferability. All norm-conserving pseudopotentials (including BCC potentials) are constructed so that they have the correct logarithmic derivatives,

$$D_l(E)=\left.\frac{u'_{l,E}(r)}{u_{l,E}(r)}\right|_{r=r_{cl}}, \tag{5.1}$$

and partial norms,

$$N_l(E)=\frac{1}{u^2_{l,E}(r_{cl})}\int_0^{r_{cl}} u^2_{l,E}(r)dr\ , \tag{5.2}$$

at the reference energies E_l. The first energy derivatives,

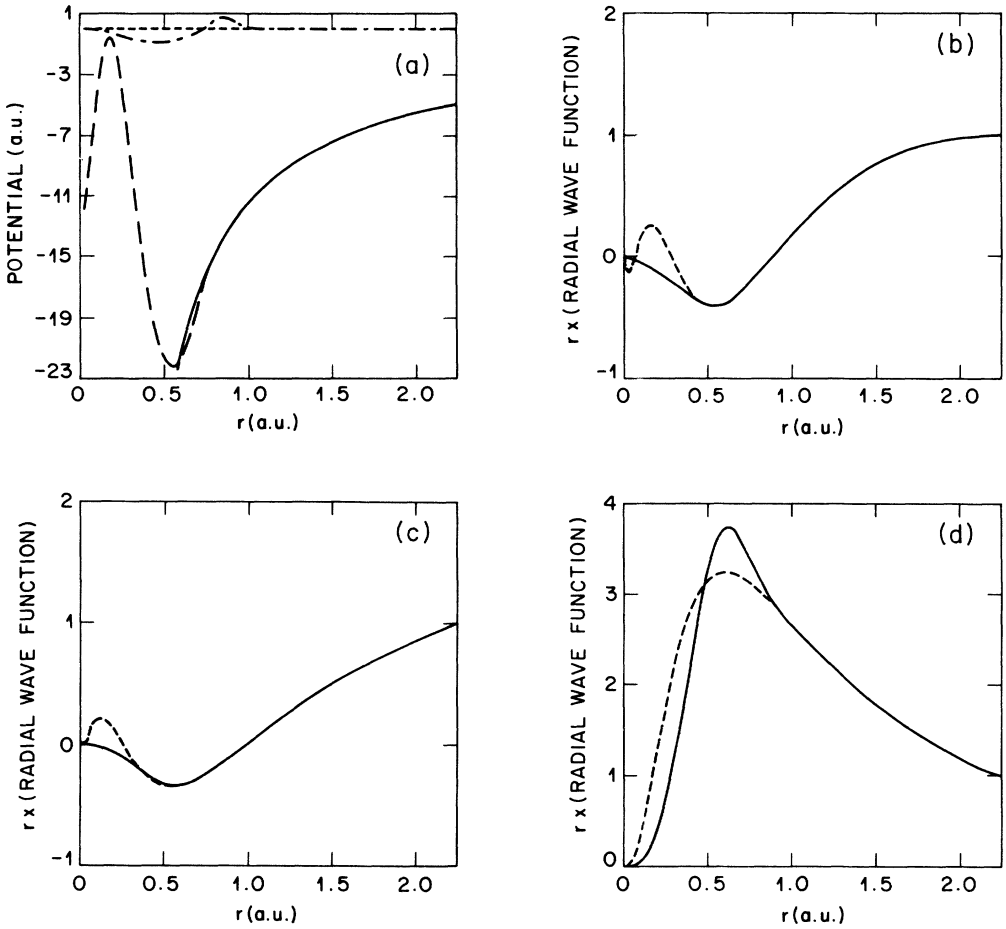

FIG. 4. (a) BCC pseudopotential for Cu. Solid line, full potential; long-dashed line, $V(r)$ potential; short-dashed line, $a(r)$ potential; alternating dashed line, $b(r)$ potential. (b)–(d) $l=0$, 1, and 2 full core and pseudo-wave-functions for the Cu pseudopotential shown in (a). Solid line, pseudo-wave-function; dashed line, full core wave function.

dD_l/dE, of the logarithmic derivatives are then also correct at E_l. A pseudopotential is usually assumed to be transferable if its logarithmic derivatives, $D_l(E)$, remain close to the full core logarithmic derivatives throughout the interesting energy range. Because of the link between the energy derivative of the logarithmic derivative and the partial norm, quickly increasing errors in $D_l(E)$ are reflected in quickly increasing errors in the partial norm, $N_l(E)$.

This is as far as most discussions of transferability go, but there is another aspect which is less often pointed out. The logarithmic-derivative argument addresses the

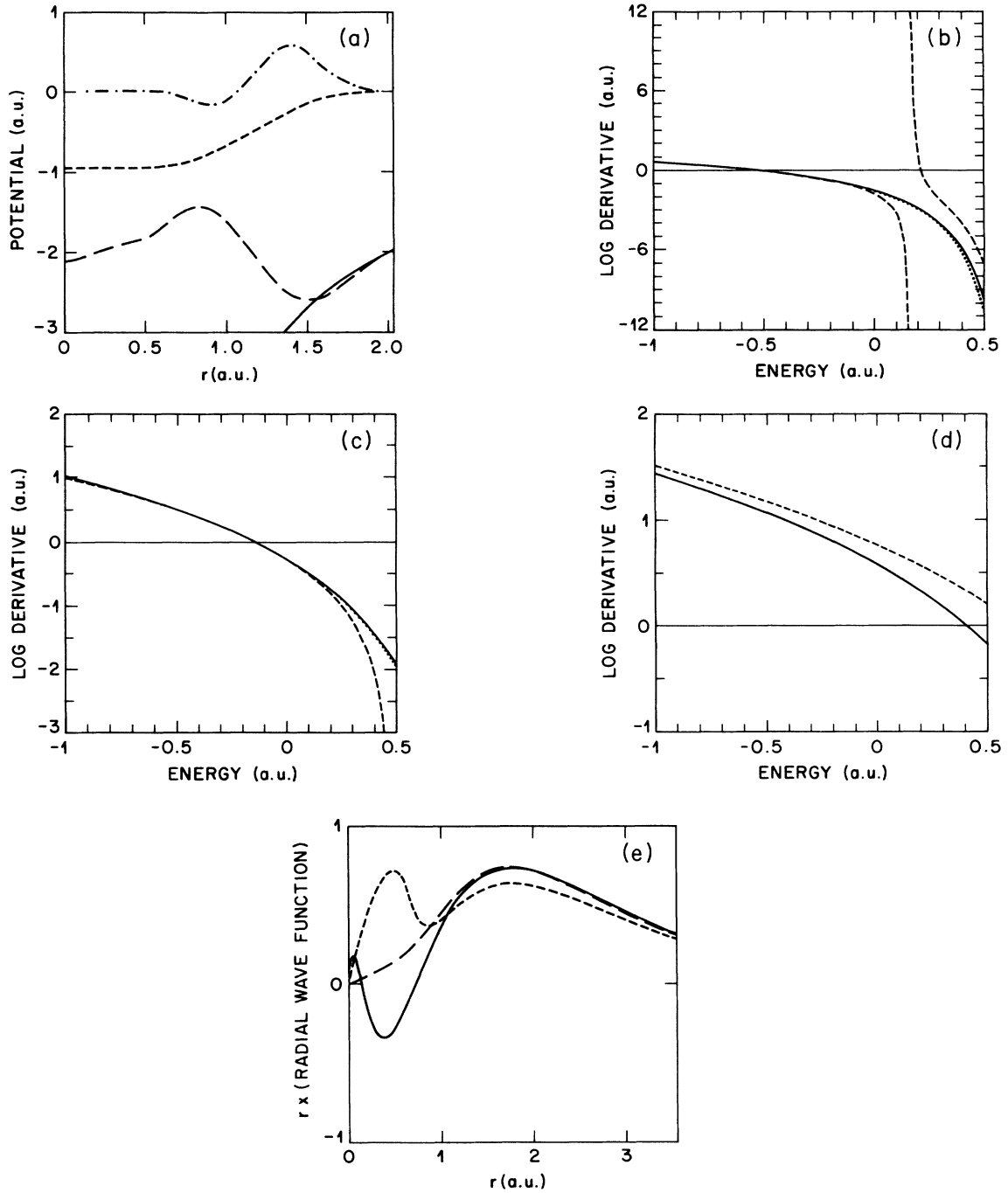

FIG. 5. (a) BCC pseudopotential Si I for Si. Solid line, full potential; long-dashed line, $V(r)$ potential; short-dashed line, $a(r)$ potential; dashed-dotted line, $b(r)$ potential. (b)–(d) $l=0$, 1, and 2 logarithmic derivatives at $r=2.1a_0$ for the Si pseudopotential shown in (a). The $l=0$, 1, and 2 reference energies were -0.4, -0.15, and -0.15 hartree, respectively. Solid line, full potential logarithmic derivative; dashed line, BCC pseudopotential logarithmic derivative; dotted line, corresponding nonlocal pseudopotential logarithmic derivative. (e) $3s$ wave functions when a perturbing potential $-e^{-2r^2}$ is added to the Si pseudopotential shown in (a). Solid line, unperturbed full core wave function; short-dashed line, perturbed BCC pseudo-wave-function; long-dashed line, perturbed wave function for corresponding nonlocal pseudopotential.

energy dependence of the scattering properties of the pseudoatom with the potential (pseudopotential + Hartree potential + exchange-correlation potential) frozen equal to the self-consistent potential. But the potential in the core region really depends on the environment, both because of contributions from the tails of the potentials on other atoms and because of the self-consistent rearrangement of the valence charge. A pseudopotential will only be transferable if the changes in scattering properties and pseudovalence charge density due to these changes in potential are not too different from the corresponding changes in the full core atom. Ordinary logarithmic-derivative transferability refers only to the response of the pseudoatom when constants are added to the potential (mathematically equivalent to changes in E) and so is a special case of this more general sort of transferability where the potential is allowed to change arbitrarily.

The pseudopotential which we have called Si I is shown in Fig. 5(a) and looks smooth and well behaved. However, the $l=0,1,2$ logarithmic derivatives, shown in Figs. 5(b), 5(c), and 5(d), are not well behaved at all. (The reference energies for the s, p, and d states were -0.4, -0.15, and -0.15 hartree, respectively). The solid lines show the full core logarithmic derivatives and the long-dashed lines the BCC logarithmic derivatives. [Figure 5(d) shows clearly that the potential does not have the correct d scattering at the reference energy.] The short-dashed lines show the logarithmic derivatives of an ordinary nonlocal potential (*not* a BCC potential) constructed by direct inversion from the wave functions of the BCC potential at the reference energies. This nonlocal potential therefore has exactly the same partial norms and logarithmic derivatives as the BCC potential at the reference energies, but may have different transferability characteristics and therefore different logarithmic derivatives at other energies.

It can be seen that the logarithmic derivatives of the BCC potential are much worse than the logarithmic derivatives of the standard nonlocal potential away from the reference energies, and so the BCC potential is less transferable. A better understanding of the origins of this problem can be obtained by examining the responses of the pseudo-wave-functions to small perturbations in the one-electron potential (the more general idea of transferability discussed above). In Fig. 5(e) we show the unperturbed full core $3s$ wave function and the perturbed BCC and standard nonlocal pseudo-wave-functions when an attractive perturbing potential ($-e^{-2r^2}$) is applied in the core region. The wave function of the standard nonlocal potential is little changed by the perturbing potential, but the BCC wave function develops a large peak in the region where the attractive potential is strongest. Clearly a pseudopotential with wave functions which respond in such an unphysical way is not likely to be transferable.

This example may seem rather artificial, but similar behavior can be seen in standard atomic transferability tests. If electrons are transferred from the $3p$ level to the $4s$ level (which is just bound), the Hartree potential becomes less repulsive in the core region and the $4s$ wave function develops just such a hump. The self-consistent effects of the $4s$ charge in this hump alter all the valence states so that the BCC eigenvalues are all much less accurate that the eigenvalues of the corresponding nonlocal potential. For configuration $3s^2 3p^1 4s^1$ (the pseudopotentials were generated in the ground-state configuration, $3s^2 3p^2 4s^0$), the full core $3s$ and $3p$ eigenvalues are -0.54 and -0.27 hartree, respectively, and the eigenvalues of the nonlocal pseudopotential agree with these almost exactly. The BCC eigenvalues are -0.49 and -0.24 hartree, and so are wrong by amounts of order 1 eV.

The root of the transferability problems for the Si I pseudopotential can be traced to the very low value of $1+a(r)$ throughout most of the core region. This corresponds to a very high radial mass and so the radial wave function is able to develop all sorts of bumps and oscillations without paying the usual high kinetic-energy cost. Similar problems would occur if $1+a+b$ was very small (except that then the bumps and oscillations would be functions of θ and ϕ rather than of r) and so any very high effective electron masses will have deleterious effects on transferability.

Despite the poor transferability of this pseudopotential, it still gives remarkably good physical properties when used in plane-wave calculations for bulk silicon. In Table I, we show the lattice parameter, a model cohesive energy (the total energy of a non-spin-polarized atom minus the total energy per atom in the solid—note that a real silicon atom is spin polarized and so the number we are quoting should not be the same as the experimental cohesive energy), and the bulk modulus and its first pressure derivative for three different (but related) silicon pseudopotentials. The first column of results is for the BCC potential, Si I; the second column is for the nonlocal potential with the same scattering properties and wave functions as Si I at the reference energies; and the third column is for a standard three-channel nonlocal potential with the correct scattering properties at the reference en-

TABLE I. Physical properties (Refs. 30 and 31) of bulk silicon calculated with the Si-I BCC pseudopotential, the corresponding two-channel nonlocal potential, and a three-channel nonlocal potential. The cohesive energy is relative to a non-spin-polarized Si atom.

	BCC Si-I	Two-channel nonlocal	Three-channel nonlocal
Lattice parameter (Å)	5.45	5.47	5.39
Cohesive energy (eV/atom)	5.78	5.67	5.91
Bulk modulus, B (kbar)	920	900	940
Pressure derivative of B	4.0	4.0	4.1

TABLE II. Γ-point eigenvalues (Refs. 30 and 31) (in eV) of bulk silicon calculated with the Si-I BCC pseudopotential, the corresponding two-channel nonlocal potential, and a three-channel nonlocal potential.

State	Symmetry	BCC Si-I	Two-channel nonlocal potential	Three-channel nonlocal potential
Γ_1	s	−6.03	−5.95	−5.93
$\Gamma_{25'}$	p,d	6.23	6.25	6.02
Γ^c_{15}	p,d	8.86	8.87	8.58
$\Gamma^c_{2'}$	s	8.88	9.18	9.21

ergies for $l=0$, 1, and 2. All three potentials have the same s and p wave functions at the reference energies and the same (large) core radii. The differences between columns 1 and 2 are mostly attributable to the extra transferability problems associated with the BCC potential, and the differences between columns 2 and 3 are attributable to the errors in the d scattering (which are roughly the same for the BCC pseudopotential and the corresponding two-channel nonlocal potential). The three-channel results are very close to those obtained using an ordinary Bachelet, Hamann, and Schluter[19] (BHS) nonlocal potential and may be taken as a standard. Questions about the accuracy of the local-density approximation are important but not relevant here and so we purposely avoid comparison with experimental results.

All the calculations used the Ceperley-Alder[12,29] form for the exchange-correlation potential (in the local-density approximation) and a plane-wave cutoff of 20 Ry. The plane-wave convergence properties of the BCC potential are similar to those of the two-channel nonlocal potential (this is not surprising since they have the same wave functions at the reference energies) and are slightly better than those of a standard BHS potential so the calculations are all well converged. Following Yin and Cohen,[30] the physical properties were determined by fitting a Murnaghan equation of state through the total energies calculated for a series of different lattice parameters (but the same cutoff energies). The bulk moduli are the least well determined of the fitted parameters, being subject to uncertainties of up to 10%.

In Table II we show the first four eigenvalues at Γ for the same three potentials. The simple symmetry decompositions of the eigenstates at Γ help to distinguish the errors due to the poor d scattering from those due to poor transferability. Neither Γ_1 nor $\Gamma^c_{2'}$ has any d character and so the two- and three-channel nonlocal potentials give almost identical eigenvalues. The errors in the BCC eigenvalues for these states are all attributable to transferability problems in the s channel. The other two states, $\Gamma_{25'}$ and Γ^c_{15}, have no s character and the BCC eigenvalues are almost the same as the eigenvalues of the two-channel nonlocal potential. Both potentials suffer from the same poor d scattering, but the transferability problems of the BCC potential are far less severe than they were in the s channel. This is all exactly as would have been predicted from the atomic logarithmic derivatives shown in Figs. 5(b), 5(c), and 5(d).

Si I seems to suffer from two distinct problems: the error in the $l=2$ scattering and the transferability problems which stem from the very high radial mass throughout most of the core. The effects of these two problems combine and the result is a poor band structure, although the total energies and associated physical (mechanical) properties still come out quite well.

The potential called Si II [Fig. 6(a)] is almost the same as Si I except that $1+a$ never falls below 0.2. The d scattering problem remains, but the transferability problems are almost completely cured as can be seen from the logarithmic derivatives in Figs. 6(b), 6(c), and 6(d). Tables III and IV show that the errors in physical properties and eigenvalues are now almost entirely attributable to the errors in the d scattering and that the BCC pseudopotential is just as transferable as the nonlocal potential. The two- and three-channel nonlocal potentials are slightly different from those corresponding to Si I, of course, but the results are all very similar as would have been expected.

Si II is already a good pseudopotential, but there is still the remaining problem of the poor d scattering. In an attempt to cure this, we constructed a fully optimized pseu-

TABLE III. Physical properties (Refs. 30 and 31) of bulk silicon calculated with the Si-II BCC pseudopotential, the corresponding two-channel nonlocal potential, and a three-channel nonlocal potential. The cohesive energy is relative to a non-spin-polarized Si atom.

	BCC Si-II	Two-channel nonlocal potential	Three-channel nonlocal potential
Lattice parameter (Å)	5.49	5.48	5.39
Cohesive energy (eV/atom)	5.64	5.63	5.92
Bulk modulus, B (kbar)	880	890	940
Pressure derivative of B	4.0	4.0	4.2

TABLE IV. Γ-point eigenvalues (Refs. 30 and 31) (in eV) of bulk silicon calculated with the Si-II BCC pseudopotential, the corresponding two-channel nonlocal potential, and a three-channel nonlocal potential.

State	Symmetry	BCC Si-II	Two-channel nonlocal potential	Three-channel nonlocal potential
Γ_1	s	−5.97	−5.95	−5.93
$\Gamma_{25'}$	p,d	6.27	6.27	6.01
Γ^c_{15}	p,d	8.90	8.89	8.57
$\Gamma^c_{2'}$	s	9.14	9.17	9.20

dopotential, Si III, with the correct partial norms and logarithmic derivatives at the reference energies in all three angular-momentum channels, $l=0,1,2$. The potential is shown in Fig. 7(a) and, as can be seen, it is very different from the two-channel BCC potentials, Si I and Si II. It is much less smooth and much harder to converge in plane waves (this is not surprising, since there was no smoothness criterion in the optimization scheme used), and $1+a$ goes to a large value (low radial effective mass) throughout most of the core rather than a small one (high radial effective mass). The logarithmic derivatives are shown in Figs. 7(b)–7(d), and it is apparent that this potential is much less transferable than Si II. The problem here is exactly the opposite of the problem for Si I: the low effective radial mass means that the wave functions are too rigid in response to external perturbations or to changes in energy and are not able to adapt properly to different environments.

The Γ-point eigenvalues for Si III and the corresponding nonlocal potential are shown in Table V and the poor transferability is easily seen. In order to obtain reasonably well converged (accurate to a few tenths of an elec-

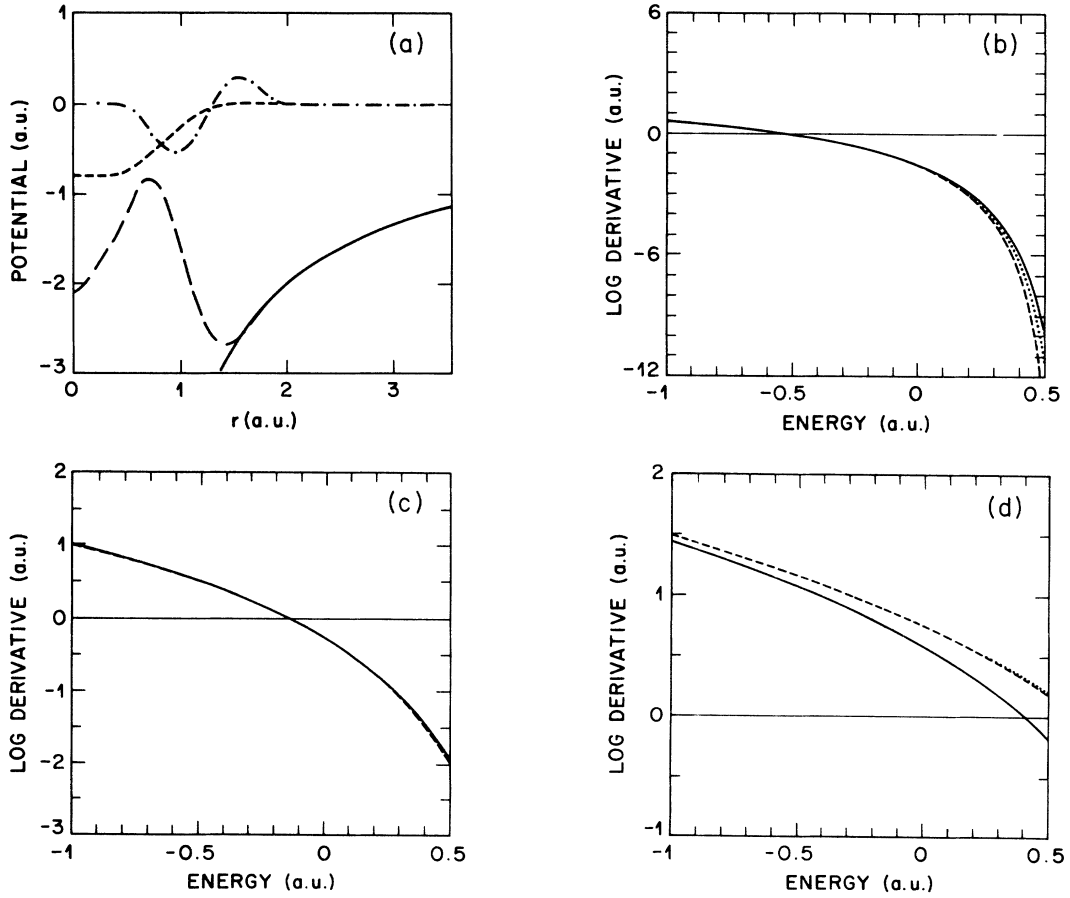

FIG. 6. (a) BCC pseudopotential Si II for Si. Solid line, full potential; long-dashed line, $V(r)$ potential; short-dashed line, $a(r)$ potential; dashed-dotted line, $b(r)$ potential. (b)–(d) $l=0$, 1, and 2 logarithmic derivatives at $r=2.1a_0$ for the Si pseudopotential shown in (a). The $l=0$, 1, and 2 reference energies were −0.4, −0.15, and −0.15 hartree, respectively. Solid line, full potential logarithmic derivative; dashed line, BCC pseudopotential logarithmic derivative; dotted line, corresponding nonlocal pseudopotential logarithmic derivative.

tron volt) eigenvalues, the plane-wave cutoff had to be increased to 30 Ry. The bulk modulus, lattice parameter, and other physical properties are not well converged at this cutoff and are not given here, but they appear to be converging towards values no worse than those obtained using Si I and Si II.

To summarize, we find that BCC pseudopotentials are only transferable as long as none of the effective masses is too far from 1, and this can be difficult to accomplish if the scattering properties are to be correct in all angular-momentum channels. In particular, we saw large decreases in transferability when the radial mass was above about 5 or below about $\frac{1}{3}$ in most of the core region. The freedom to vary the effective masses is all that distinguishes BCC pseudopotentials from ordinary local potentials, but this freedom is severely curtailed by the requirement that any useful BCC pseudopotential must be transferable.

VI. CONCLUSIONS

In this paper we have investigated the properties of a new class of pseudopotential suggested by Bachelet, Ceperley, and Chiocchetti[9] (BCC) as uniquely suitable for use in Green's-function and diffusion quantum Monte Carlo (QMC) calculations. After carefully explaining the reasoning which led BCC to propose this new type of pseudopotential, we studied the construction and transferability of BCC pseudopotentials within local-density-functional theory. Of course, the primary interest in these pseudopotentials is because of their uses in QMC calculations and not because of their uses in density-functional theory, but any pseudopotential which does not work well in density-functional calculations will

TABLE V. Γ-point eigenvalues (Refs. 30 and 31) (in eV) of bulk silicon calculated with the Si-III BCC pseudopotential and the corresponding nonlocal potential.

State	Symmetry	BCC Si-III	Nonlocal
Γ_1	s	−5.76	−5.95
$\Gamma_{25'}$	p,d	6.09	6.02
Γ^c_{15}	p,d	8.57	8.58
$\Gamma^c_{2'}$	c	9.57	9.16

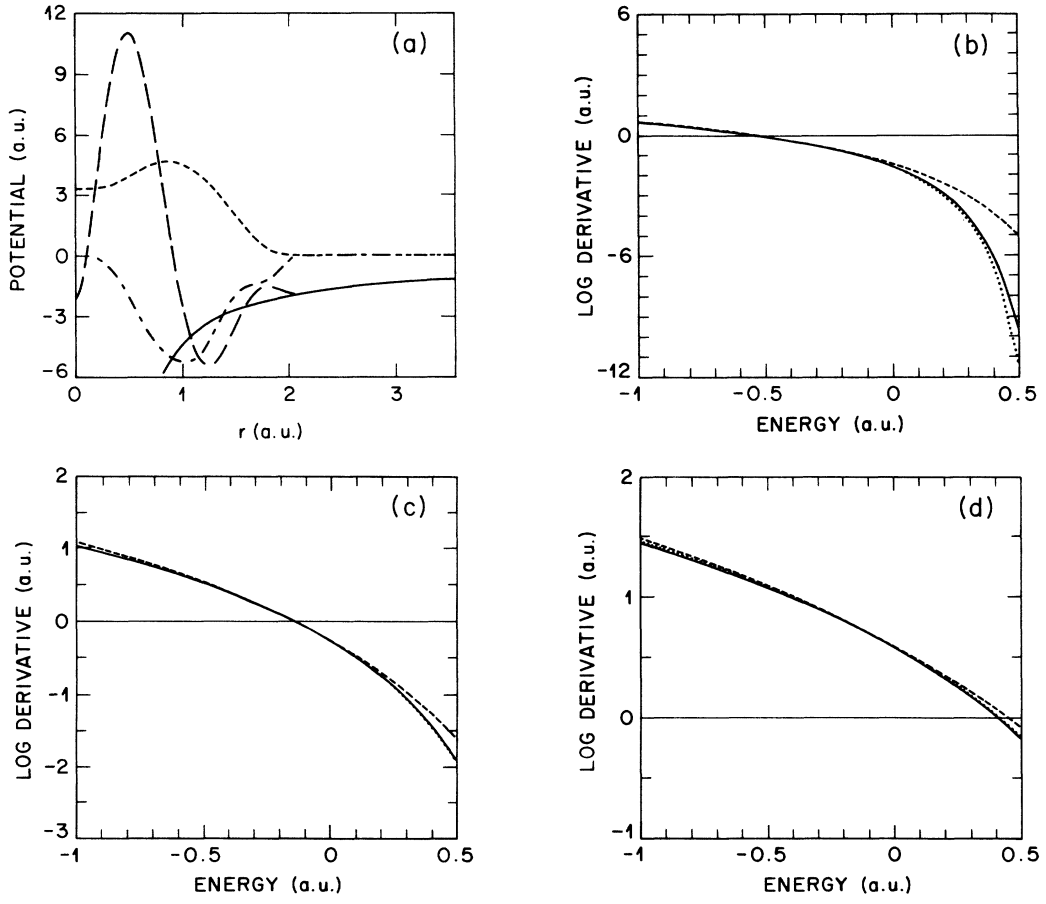

FIG. 7. (a) BCC pseudopotential Si III for Si. Solid line, full potential; long-dashed line, $V(r)$ potential; short-dashed line, $a(r)$ potential; alternating dashed line, $b(r)$ potential. (b)–(d) $l=0$, 1, and 2 logarithmic derivatives at $r=2.1a_0$ for the Si pseudopotential shown in (a). The $l=0$, 1, and 2 reference energies were −0.4, −0.15, and −0.15 hartree, respectively. Solid line, full potential logarithmic derivative; dashed line, BCC pseudopotential logarithmic derivative; dotted line, corresponding nonlocal pseudopotential logarithmic derivative.

not work well in QMC calculations either, and so the density-functional results (besides being much easier to obtain than the corresponding QMC results) are useful indicators of quality.

The construction of BCC pseudopotentials is not straightforward even within density-functional theory. We considered two possible approaches and implemented one of these. Because of the energy-level ordering constraint discussed in Sec. III, it turns out that BCC pseudopotentials with nodeless valence functions (i.e., pseudopotentials which bind no core states) do not exist for some elements, among them the transition metals. Even for elements where the energy-level ordering constraint does not absolutely rule it out, it can sometimes be numerically difficult to find BCC potentials. In some cases (oxygen being one), we are able to succeed only by accepting potentials which have the correct partial norms and logarithmic derivatives in the two most important angular-momentum channels (s and p for oxygen) but not in the third.

Questions of transferability are also more complicated for BCC pseudopotentials than for ordinary nonlocal pseudopotentials, and we find that BCC potentials are only transferable as long as all the effective electron masses are fairly close to 1. Together with the energy-level ordering constraint, this considerably limits the number of elements for which we are able to find useful BCC potentials. The more "nonlocal" (in the sense of ordinary nonlocal pseudopotentials) an element is, the more difficult it is to find a good BCC potential.

To set against these disadvantages, BCC pseudopotentials have two principal advantages. They have very simple plane-wave matrix elements of a form which is useful in iterative band-structure methods,[11] and they are the most general type of pseudopotential suitable for use in Green's-function and diffusion QMC calculations. For pseudopotential QMC calculations, at any rate, there seems little alternative but to put up with the problems of BCC pseudopotentials for the time being.

APPENDIX A: NONLOCAL NORM-CONSERVING PSEUDOPOTENTIALS

There are several reasons why pseudopotential methods have been so popular in electronic-structure calculations based on density-functional theory. The first is that they enable one to escape calculating the core electronic wave functions and their associated energies. Since the core electrons are tightly bound near the nuclei and do not participate in chemical bonding, their wave functions are not usually of much interest and the use of pseudopotentials avoids a lot of unnecessary work. A second reason is that conventional pseudopotentials, although nonlocal, are smooth and do not have Coulomb singularities at the nuclei. The pseudo-wave-functions are therefore smoother and have fewer nodes than the true valence wave functions and can often be expanded accurately using a fairly small basis of plane waves. Plane-wave basis sets are so simple and convenient that this can be reckoned a considerable advantage.

The pseudopotentials most frequently used in density-functional electronic-structure calculations are called norm-conserving pseudopotentials.[17–19] To understand how they work, think first about an isolated atom. The Schrödinger equation (including all the core electrons) can be solved within density-functional theory, yielding a self-consistent one-electron potential,

$$V([n(\mathbf{r})],\mathbf{r})=V_{\text{nuc}}(\mathbf{r})+V_{\text{Hartree}}([n(\mathbf{r})],\mathbf{r}) + V_{\text{xc}}([n(\mathbf{r})],\mathbf{r}) , \tag{A1}$$

and a density, $n(\mathbf{r})$. The density is the sum of the densities associated with each of the occupied one-electron functions,

$$n(\mathbf{r})=\sum_{i \text{ occupied}} \psi_i^*(\mathbf{r})\psi_i(\mathbf{r}) , \tag{A2}$$

and these are the solutions of the Kohn-Sham equation

$$\{-\tfrac{1}{2}\nabla^2+V([n(\mathbf{r})],\mathbf{r})\}\psi_i(\mathbf{r})=\varepsilon_i\psi_i(\mathbf{r}) . \tag{A3}$$

The Kohn-Sham equation looks like a one-electron Schrödinger equation, but the potential depends self-consistently on the density it generates, and the one-electron eigenfunctions and eigenenergies are nothing more than mathematical constructs with no strict physical meaning.

Let us look at the properties of the radial atomic Kohn-Sham equation for angular momentum l when the potential is held equal to the self-consistent potential. The radial equation has a regular singular point at the origin and a power-series expansion about that point shows that the radial wave functions must start either like r^l or $r^{-(l+1)}$ as r increases from zero. Throwing away the $r^{-(l+1)}$ solution as unphysical, we are left with the solution which starts like r^l. So, if we integrate the radial equation out from the origin at an energy E (which need not be an eigenenergy), we obtain a radial function $R_{l,E}(r)$, which is unique except for a normalization (it is not even normalizable in the usual sense unless E is an eigenvalue). It is the rejection of solutions which diverge at the singular point at the origin which guarantees this uniqueness.

Now imagine drawing a sphere of radius r_{cl} centered on the atom. The uniqueness of the radial function within the sphere implies that the radial logarithmic derivative,

$$D_l(E)=\left.\frac{d\ln[R_{l,E}(r)]}{dr}\right|_{r=r_{cl}} ,$$

evaluated at r_{cl} is also unique. But $D_l(E)$ is the only information (boundary condition) needed to continue the radial integration beyond r_{cl}. Given $D_l(E)$, we could construct (again only up to a normalization) the radial wave function outside r_{cl} and hence find the eigenvalues [energies at which $R_{l,E}(r)\to 0$ as $r\to\infty$] and other quantities of interest. Solutions of the radial equation outside the sphere which do not have the correct logarithmic derivative at r_{cl} would blow up when integrated in towards the origin and so are not acceptable.

So, as long as we are only interested in angular momentum l solutions of the Kohn-Sham equation in the region outside the sphere, then the function $D_l(E)$ is all we need

to know about the inside of the sphere—details of the potential and wave functions are irrelevant. The idea of a pseudopotential follows immediately. For a given l, we simply replace the full self-consistent atomic one-electron potential inside the sphere by a smoother and weaker potential, the pseudopotential for that l, chosen so that its logarithmic-derivative function, $D_l^{\rm ps}(E)$, agrees reasonably well with $D_l(E)$ over the energy range of interest.

In solid-state physics, the energies of interest typically span the highest few occupied and lowest few unoccupied energy bands, corresponding roughly to the highest few occupied and lowest few unoccupied atomic-energy levels or scattering resonances. Near any particular atom, the wave functions of the solid may contain important angular-momentum components corresponding to any or all of these atomic levels, and so a good pseudopotential must mimic the scattering properties (logarithmic derivatives) of the fully self-consistent potential for several different angular momenta at once. This is accomplished by combining the separate pseudopotentials, $V_l^{\rm ps}(r)$, for each important l, to give a total pseudopotential,

$$\hat{V}^{\rm ps}(r)=V^{\rm local}(r)+\sum_l [V_l^{\rm ps}(r)-V^{\rm local}(r)]\hat{P}_l \ , \tag{A4}$$

where $\hat{P}_l$ is a nonlocal operator which projects out the l component of any function on which it acts,

$$\hat{P}_l f(r,\theta,\phi)=\sum_{m=-l}^{l} Y_{lm}(\theta,\phi)\int_{\phi'=0}^{2\pi}\int_{\theta'=0}^{\pi} Y_{lm}^*(\theta',\phi')f(r,\theta',\phi')\sin\theta'\,d\theta'\,d\phi' \ , \tag{A5}$$

and $V^{\rm local}(r)$ is any smooth function which tends to the full potential at large r. $V^{\rm local}(r)$ is usually set equal to one of the $V_l^{\rm ps}(r)$ functions and its presence ensures that angular momenta not specifically included in the summation over l see the correct potential far from the nucleus and a reasonable potential near the nucleus. As will be shown later, the $V_l^{\rm ps}(r)$ functions can be chosen so that their logarithmic derivatives agree reasonably well with those of the full potential over the energy range of interest but so that the pseudopotential does not bind any core states. The close agreement of the logarithmic derivatives means that the valence wave functions of the true and pseudopotential radial equations are very similar outside r_{cl} (they are identical when the logarithmic derivatives agree exactly, of course); but they differ inside r_{cl} since the true wave functions usually have several nodes whereas the pseudo-wave-functions have no core states below them and so are nodeless.

So far, we have suggested that it may be possible to design pseudopotentials which mimic the scattering of the valence electrons from the full potential of a free atom. However, this will be useful only if the pseudopotentials constructed can then be used to replace the atomic potentials in other environments, thereby simplifying electronic-structure calculations. If this cannot be done—if a pseudopotential is not transferable—then it is useless. Unfortunately, it is clear that a pseudopotential like $\hat{V}^{\rm ps}(r)$, which implicitly includes Hartree and exchange-correlation contributions from both the core and valence atomic electron densities, cannot be transferable. Valence (but not core) electron densities may be strongly environment dependent and so it would not make sense to try using $\hat{V}^{\rm ps}(r)$ to replace an atomic potential in a molecule or solid. To obtain a pseudopotential which one expects to be more transferable, the Hartree and exchange-correlation contributions due to the atomic pseudovalence charge density must be subtracted from $\hat{V}^{\rm ps}(r)$ [or, more specifically, from $V^{\rm local}(r)$] to obtain a "bare" or "unscreened" pseudopotential:

$$\hat{V}_{\rm bare}^{\rm ps}(r)=V_{\rm bare}^{\rm local}(r)+\sum_l [V_l^{\rm ps}(r)-V^{\rm local}(r)]\hat{P}_l \ . \tag{A6}$$

This can then be used in pseudopotential density-functional calculations in which only the valence electrons appear explicitly and need to be treated self-consistently.

The construction of the pseudopotential $V_l^{\rm ps}(r)$ for a particular l goes as follows.[17] First one chooses a reference energy E_l, which should be somewhere in the middle of the range of interesting energies but is otherwise unrestricted (this has only recently been realized[24]). For angular momenta corresponding to occupied valence states, it is convenient to set E_l to the valence atomic eigenvalue. For unoccupied angular momenta, a good choice,[24] although not the most common one, is to set E_l to the highest of the occupied reference energies.

The next step is to choose the core radius r_{cl}. For occupied angular momenta, this should be somewhere inside the last maximum but outside the last node of the corresponding atomic valence wave function. In general, r_{cl} should be regarded as a "quality" parameter[19] and kept as small as possible, but values outside the last node are sometimes permissible as long as they are considerably less than half the interatomic distance in the solid. The choice of r_{cl} for unoccupied angular momenta is more problematic, although Hamann's[24] prescription seems to work well.

Given E_l, r_{cl}, and the self-consistent potential, it is straightforward to integrate the radial Schrödinger equation out from the origin to r_{cl} and hence to obtain $R_{l,E_l}(r)$ (the radial wave function at energy E_l for $r \le r_{cl}$) and $D_l(E_l)$ (the logarithmic derivative evaluated at r_{cl} and energy E_l). The radial wave function $R_{l,E_l}(r)$ shows the effects of the strong Coulomb potential near $r=0$ and has as many nodes within r_{cl} as there are core wave functions for that l. Different procedures have been proposed,[17–19] but the crucial step in the construction of any pseudopotential is the replacement of the wave function within r_{cl} by a pseudo-wave-function, $R_{l,E_l}^{\rm ps}(r)$. The pseudo-wave-function can be almost any convenient function which is smooth and nodeless and has the correct logarithmic derivative, $D_l^{\rm ps}(E_l)=D_l(E_l)$, at r_{cl}. The pseudopotential $V_l^{\rm ps}(r)$ is then *defined* as that potential which would give rise to $R_{l,E_l}^{\rm ps}(r)$ if the corresponding ra-

dial equation was integrated out from the origin at energy E_l. Because the pseudo-wave-function is nodeless, the pseudopotential may easily be found by inverting the radial pseudo-Schrödinger (Kohn-Sham) equation,

$$V_l^{\rm ps}(r)=E_l-\frac{l(l+1)}{2r^2}+\frac{1}{2}\left[\frac{d^2u_{l,E_l}^{\rm ps}(r)/dr^2}{u_{l,E_l}^{\rm ps}(r)}\right], \tag{A7}$$

where

$$u_{l,E_l}^{\rm ps}(r)=rR_{l,E_l}^{\rm ps}(r)\ . \tag{A8}$$

Note that $u_{l,E_l}^{\rm ps}(r)$ must behave like r^{l+1} near $r=0$ and must have at least two radial derivatives (not just the logarithmic derivative) correct at r_{cl} if a continuous nonsingular pseudopotential is sought.

Given a sensible choice for the pseudo-wave-function, the pseudopotential will be smooth and much weaker than the full potential. Nevertheless, it gives the correct logarithmic derivative at the reference energy E_l, and so has the correct scattering properties at that energy. Since the pseudo-wave-function is nodeless by construction, the pseudopotential has no core states.

So far, we have only ensured that the pseudopotential logarithmic derivative matches the true logarithmic derivative at one energy. We can do better than this by making sure that the first energy derivative of the pseudopotential logarithmic derivative is also correct at E_l. Pseudopotentials for which this is true are called norm-conserving pseudopotentials,[17] and their construction is suggested by a version of the Friedel sum rule,

$$-\frac{1}{2}\frac{\partial}{\partial E}\frac{\partial}{\partial r}\ln[R_{l,E}(r)]\Big|_{r=r_{cl}}=\frac{1}{r_{cl}^2R_{l,E}^2(r_c)}\int_0^{r_c}R_{l,E}^2(r)r^2\,dr\ , \tag{A9}$$

which holds for any radial wave function satisfying a Schrödinger-like equation, for any energy E, and for any r_{cl} (for a proof of this and other related equalities, see Ref. 31). Equation (A9) says that the first energy derivative of the pseudopotential radial logarithmic derivative at energy E_l is automatically correct whenever the pseudo-wave-function partial norm,

$$\frac{1}{u_{l,E_l}^2(r_c)}\int_0^{r_c}u_{l,E_l}^2(r)dr\ ,$$

is also correct. As long as we choose pseudo-wave-functions which have the the same partial norms as the true wave functions, the pseudopotentials generated will always have the correct first energy derivatives of the logarithmic derivatives and therefore improved transferability.

Having the correct partial norms is important for another reason as well. If the partial norms are wrong, then the proportion of the total valence charge inside the spheres is also wrong and this alters the Hartree and exchange-correlation potentials in a way which adversely affects transferability.

It turns out that fixing the logarithmic derivatives and their first energy derivatives is usually good enough, although there has been some recent work[31] investigating whether there is much advantage to be gained by fixing higher-energy derivatives of the logarithmic derivatives as well. Within the limits imposed by the use of the local-density approximation, pseudopotential calculations usually give a good description of the interatomic forces and the valence and lower conduction (density-functional) bands in solids. The main problem is that there are a number of elements (those for which there is an important valence angular-momentum channel—the p channel in oxygen, for example—but no corresponding core state) for which it is not possible to find pseudo-wave-functions which can be expanded in a reasonably small number of plane waves. This is a technical matter which has nothing to do with the existence or quality of the pseudopotentials themselves, but which nevertheless severely limits the application of plane-wave band-structure methods. Attempts to ameliorate this problem fall into two classes: one approach[28] is to try to optimize the plane-wave convergence properties of pseudopotentials for the difficult elements by improvements in the construction procedure; and the other[32,33] is to increase the number of plane waves that can be used by replacing the angular-momentum nonlocality of standard norm-conserving pseudopotentials by something which is more efficient when combined with the Car-Parrinello[11] scheme for solving the eigenproblem. Kleinman and Bylander[32] replace the P_l operators by projections onto local basis functions in the core region. The resulting pseudopotentials are very efficient and so allow many plane waves to be used, but seem to suffer from subtle transferability problems[34] which are not yet completely solved. Vanderbilt[33] has suggested a form of energy-dependent pseudopotential which seems to be both efficient and quickly convergent in plane waves.

APPENDIX B: THE GREEN'S-FUNCTION AND DIFFUSION MONTE CARLO METHODS

This is a paper about pseudopotentials for QMC and not about QMC itself, and so the descriptions of the Green's-function and diffusion QMC methods[6–8] will be very brief and much more simplified. In particular, we will give little attention to the methods[6–8] for carrying out the sampling of probability distributions and Green's functions required in all QMC calculations. Although these methods constitute the heart of the Monte Carlo approach, the question of how to combine pseudopotentials with QMC can be addressed without considering them in much detail.

Consider the "imaginary-time" Schrödinger equation,

$$\hat{H}|\psi\rangle=-\frac{\partial}{\partial\tau}|\psi\rangle\ , \tag{B1}$$

where $\hat{H}$ is the full interacting electron Hamiltonian. The phrase "imaginary time" sounds profound, but signifies nothing more than that the usual Schrödinger equation,

$$H|\psi\rangle=i\frac{\partial}{\partial t}|\psi\rangle\ , \tag{B2}$$

would look like Eq. (B1) if $t=-i\tau$ was an imaginary variable. The imaginary-time Schrödinger equation is a standard partial-differential equation and the imaginary time τ is an ordinary real variable despite its name.

The formal solution of Eq. (B1) is

$$|\psi(\tau_2)\rangle = e^{-\hat{H}(\tau_2-\tau_1)}|\psi(\tau_1)\rangle \ , \tag{B3}$$

and so $e^{-\hat{H}\tau}$ is the (imaginary-) time evolution operator. Suppose that the full interacting electron Hamiltonian has many electron eigenfunctions $|\psi_i\rangle$ with eigenenergies E_i,

$$H|\psi_i\rangle = E_i|\psi_i\rangle \ . \tag{B4}$$

Then an arbitrary starting function $|\psi(0)\rangle$ can be expanded in terms of the eigenfunctions

$$|\psi(0)\rangle = \sum_i c_i|\psi_i\rangle \ , \tag{B5}$$

and application of Eq. (B3) shows that

$$|\psi(\tau)\rangle = \sum_i e^{-E_i\tau}c_i|\psi_i\rangle \ . \tag{B6}$$

At large enough times, the contribution from the eigenstate with the lowest eigenvalue dominates and so, assuming that the starting state $|\psi(0)\rangle$ was not orthogonal to the ground state $|\psi_0\rangle$, we see that

$$\lim_{\tau\to\infty} |\psi(\tau)\rangle = e^{-E_0\tau}c_0|\psi_0\rangle \ , \tag{B7}$$

where E_0 is the ground-state energy. In the $\mathbf{R}$ representation, Eq. (B7) becomes

$$\lim_{\tau\to\infty} \psi(\mathbf{R},\tau) = e^{-E_0\tau}c_0\psi_0(\mathbf{R}) \ , \tag{B8}$$

where

$$\psi(\mathbf{R},\tau) = \langle \mathbf{R}|\psi(\tau)\rangle \ . \tag{B9}$$

It follows that the quantity

$$E_G(\tau) = -\frac{\partial}{\partial\tau}\left[\ln\left|\int \psi(\mathbf{R},\tau)d\mathbf{R}\right|\right] \tag{B10}$$

converges to the ground-state energy as $\tau\to\infty$. For all starting states which are not orthogonal to the ground state, the dynamics of Eq. (B1) eventually projects out the ground-state component.

Both the diffusion and Green's-function QMC methods are based on a combination of this imaginary-time projection and the observation that Eq. (B1) looks like a diffusion equation. If one wanted to describe particles diffusing while multiplying or being absorbed at a rate $s(\mathbf{R})$ which depends on position [negative $s(\mathbf{R})$ corresponds to absorption, positive $s(\mathbf{R})$ to multiplication], then the appropriate equation would be

$$\nabla_R^2\rho + s(\mathbf{R})\rho = \frac{\partial\rho}{\partial\tau} \ . \tag{B11}$$

But if we define the $3N$- (where N is the number of electrons) dimensional vector

$$\mathbf{R} = (r_{1x}, r_{1y}, r_{1z}, r_{2x}, r_{2y}, r_{2z}, \ldots, r_{Nx}, r_{Ny}, r_{Nz}) \ ,$$

then the imaginary-time Schrödinger equation can be written as

$$\tfrac{1}{2}\nabla_{\mathbf{R}}^2\psi(\mathbf{R},\tau) - V(\mathbf{R})\psi(\mathbf{R},\tau) = \frac{\partial\psi(\mathbf{R},\tau)}{\partial\tau} \ , \tag{B12}$$

which looks exactly the same. [In this equation, $V(\mathbf{R})$ contains all the electron-electron interactions as well as the one-electron parts of the potential.] We see, therefore, that the imaginary-time Schrödinger equation describes the diffusion of particles which are multiplying or being absorbed at a rate $-V(\mathbf{R})$. These particles or "walkers" diffuse around in a $3N$-dimensional space and are *not* the electrons, but after a long enough (imaginary) time, the density of diffusers at a point $\mathbf{R}$ in the $3N$-dimensional space becomes proportional to the value of the many-electron wave function there. It should be noted here that the Pauli principle causes problems when dealing with many fermion systems such as the electrons in a solid. The density of diffusers is a positive quantity, and so the wave function evaluated by solving the diffusion equation is always nodeless and cannot be the many-electron ground state (which is totally antisymmetric in the electron coordinates and can never be nodeless). In fact, the solution of the diffusion problem is the boson ground state rather than the fermion ground state. We will return to this problem later on and explain how the fixed-node approximation can be used to build the required antisymmetry into the solution of the diffusion equation. For the time being, however, let us forget the Pauli principle and imagine that the nodeless ground state is the solution we want. The hope, of course, is that it might be possible to solve the diffusion equation by simulation on a computer, and hence find the ground state of the many-interacting-electron system.

We can get another view of the diffusion idea by going back to Eq. (B3) and rewriting it in the $\mathbf{R}$ representation,

$$\psi(\mathbf{R},\tau_2) = \int G(\mathbf{R},\mathbf{R}';\tau_2-\tau_1)\psi(\mathbf{R}',\tau_1)d\mathbf{R}' \ , \tag{B13}$$

where

$$G(\mathbf{R},\mathbf{R}';\tau) = \langle \mathbf{R}|e^{-\hat{H}\tau}|\mathbf{R}'\rangle \ , \tag{B14}$$

and we have used the completeness of the $\mathbf{R}$ basis,

$$\int |\mathbf{R}'\rangle\langle\mathbf{R}'|d\mathbf{R}' = 1 \ . \tag{B15}$$

Equation (B13) has a simple physical interpretation: $\psi(\mathbf{R}',\tau_1)$ is the number density of particles at point $\mathbf{R}'$ at the initial time τ_1, $G(\mathbf{R},\mathbf{R}';\tau_2-\tau_1)$ is a transition probability, giving the number density of particles at $(\mathbf{R},\tau_2)$ arising from a particle at $(\mathbf{R}',\tau_1)$, and $\psi(\mathbf{R},\tau_2)$ is the number density of particles at $(\mathbf{R},\tau_2)$. We might, therefore, hope to simulate the dynamics specified by Eq. (B13) as follows.

(i) Choose a set of points $\{\mathbf{R}_1',\mathbf{R}_2',\ldots,\mathbf{R}_n'\}$ distributed according to some initial number density $\psi(\mathbf{R}',\tau_1)$. This could be accomplished using the Metropolis[6,35] algorithm, for example. The expected number of points in a small volume δV at $\mathbf{R}'$ is then $C\psi(\mathbf{R}',\tau_1)\delta V$, where C is a constant.

(ii) Replace each point $\mathbf{R}_i'$ by m_i other points,

$\{\mathbf{R}_1^{(i)}, \mathbf{R}_2^{(i)}, \ldots, \mathbf{R}_{m_i}^{(i)}\}$, distributed according to the number density $G(\mathbf{R}, \mathbf{R}_i', \tau_2 - \tau_1)$. Make sure that

$$\langle m_i \rangle = \int G(\mathbf{R}, \mathbf{R}_i'; \tau_2 - \tau_1) d\mathbf{R} \ . \tag{B16}$$

(iii) The points generated (all of them together) are then distributed according to the density function $\psi(\mathbf{R}, \tau_2)$, with the expected number in a small volume δV at $\mathbf{R}$ being $C\psi(\mathbf{R}, \tau_2)\delta V$.

As long as the starting state $\psi(\mathbf{R}', \tau_1)$ is not orthogonal to the ground state, then the distribution of walkers gives the ground-state wave function for large enough $\tau_2 - \tau_1$. The expected value of the total number of walkers at time τ_2 is proportional to $\int \psi(\mathbf{R}, \tau_2) d\mathbf{R}$, and so an estimate (biased) of E_G [see Eq. (B10)] can be obtained from the rate of change of the total number of walkers with time. For large $\tau_2 - \tau_1$, $E_G \to E_0$ and so this estimate (which is called the growth estimate and is only one of several possible energy estimates) gives the ground-state energy. It is obvious that neither the growth estimate of the ground-state energy nor the sampling of the ground-state wave function will converge quickly unless the starting state $\psi(\mathbf{R}', \tau_1)$ is close to the true ground-state wave function.

The rate at which $E_G(\tau_2)$ (and hence the statistical growth estimate) converges to E_0 can be greatly increased by using an "importance" weighting in Eq. (B13). Suppose $I(\mathbf{R})$ is some arbitrary positive function. Then Eq. (B13) can be rewritten

$$\tilde{\psi}(\mathbf{R}, \tau_2) = \int \tilde{G}(\mathbf{R}, \mathbf{R}'; \tau_2 - \tau_1) \tilde{\psi}(\mathbf{R}', \tau_1) d\mathbf{R}' \ , \tag{B17}$$

where

$$\tilde{\psi}(\mathbf{R}, \tau) = I(\mathbf{R}) \psi(\mathbf{R}, \tau) \tag{B18a}$$

and

$$\tilde{G}(\mathbf{R}, \mathbf{R}'; \tau) = \frac{I(\mathbf{R}) G(\mathbf{R}, \mathbf{R}'; \tau)}{I(\mathbf{R}')} \ . \tag{B18b}$$

Equation (B17) is of the same form as Eq. (B13) and can be solved on a computer in a similar way. We now work out the choice of $I(\mathbf{R})$ which optimizes the imaginary-time convergence rate of

$$\tilde{E}_G(\tau_2) = -\frac{\partial}{\partial \tau_2} \left[\ln \left| \int \tilde{\psi}(\mathbf{R}, \tau_2) d\mathbf{R} \right| \right] . \tag{B19}$$

Suppose that both $\psi(\mathbf{R}, \tau_1)$ and $I(\mathbf{R})$ are expanded in the eigenfunctions

$$\psi(\mathbf{R}, \tau_1) = \sum_i c_i \psi_i(\mathbf{R}) \ , \tag{B20a}$$

$$I(\mathbf{R}) = \sum_i d_i \psi_i(\mathbf{R}) \ . \tag{B20b}$$

Then $\tilde{E}_G(\tau_2)$ is given by

$$\tilde{E}_G(\tau_2) = -\frac{\partial}{\partial \tau_2} \left[\ln \left| \sum_i c_i d_i e^{-E_i(\tau_2 - \tau_1)} \right| \right] , \tag{B21}$$

and we see that if we choose $I(\mathbf{R}) = \psi_0(\mathbf{R})$ (which corresponds to $d_0 = 1$ and $d_i = 0$ for $i \neq 0$), then $\tilde{E}_G(\tau_2)$ gives the ground-state energy exactly at all times irrespective of the starting function, $\psi(\mathbf{R}, \tau_1)$. We no longer have to wait for $\tilde{E}_G(\tau_2)$ to converge, but can look at it immediately at $\tau_2 = \tau_1$ and still get the exact ground-state energy.

Perhaps it is not surprising that we can calculate the ground-state energy exactly if we use the exact ground-state wave function as the importance function (it turns out that it is also possible to get rid of the statistical errors associated with the computer simulation in this case). But the point about importance sampling is that even using an importance function, which is a reasonable guess at the true ground-state wave function (the usual choice is a Hartree-Fock determinant multiplied by a Jastrow factor), can make the simulation much more efficient. The convergence rate of the wave function is not improved (in the sense that the components along eigenfunctions of energy $E > E_0$ still decay relative to the ground-state component at a rate $e^{-(E - E_0)\tau}$), but the fluctuations in the rate of change of population are reduced so that the growth estimate of the energy converges more quickly and this and many other statistical quantities can be calculated more accurately. (The overall rate of change of population can be kept near zero by regularly redefining the zero of energy during the calculation.) It is not difficult to write down the equivalent of the diffusion equation for $\tilde{\psi}(\mathbf{R})$ (see, e.g., Ref. 36), and one finds that it looks rather like Eq. (B12) except that the form of the potential is altered and there are now drift terms as well as diffusion and multiplication and/or absorption terms. The simulation is a little more complicated than before, but is no more difficult in principle and converges much more quickly.

Although importance sampling is of great practical value, it has little relevance to the question of how to combine the pseudopotential and QMC approaches. We therefore ignore it in the bulk of the paper and the rest of this appendix; the generalizations of the equations to include an importance weighting are straightforward.

In describing the algorithm for solving Eq. (B13) earlier, it was assumed that the Green's function $G(\mathbf{R}, \mathbf{R}', \tau)$ was known, and of course this is not the case. There are two ways to get around this problem, one of which leads to the diffusion Monte Carlo method, and the other to the Green's-function Monte Carlo method. These two methods are the same in all other respects.

In the diffusion Monte Carlo method, use is made of the known form of the Green's function at short enough times,

$$G(\mathbf{R}', \mathbf{R}; \Delta\tau) = \frac{1}{(2\pi \Delta\tau)^{3N/2}} \times \exp\left[-\frac{1}{2} \frac{(\mathbf{R}' - \mathbf{R})^2}{\Delta\tau} \right] \times \exp[-V(\mathbf{R}) \Delta\tau] \ . \tag{B22}$$

A straightforward derivation of this equation may be based on the techniques explained on pages 57–60 of Ref. 16, but the form should come as no surprise in light of our earlier identification of the imaginary-time

Schrödinger equation as a diffusion equation with branching (birth and death) terms. The first exponential is just the usual spreading Gaussian to be expected in any diffusion problem, and the second exponential is a multiplicative factor (independent of $\mathbf{R}'$) to take account of the creation and annihilation of particles (if importance sampling is included, there is also a drift contribution). Since $\Delta\tau$ is small, the particles never get much further than $\sqrt{\Delta\tau}$ from $\mathbf{R}$, and the assumption behind Eq. (B22) is that the potential is constant throughout this region. If the potential varies very rapidly, then shorter time steps are needed for the same accuracy.

We can now make repeated use of Eq. (B13) to propagate the wave function (particle-density distribution) from τ_1 to $\tau_1+\Delta\tau$ and then to $\tau_1+2\Delta\tau$ and so on. The algorithm is just as described earlier: first we choose a set of points from the distribution $\psi(\mathbf{R},\tau_1)$ and then repeatedly replace each point $\mathbf{R}_i$ by m_i other points chosen from $G(\mathbf{R},\mathbf{R}_i;\Delta\tau)$. Since the form of the short-time Green's function is so simple, this replacement step may be described in a more concrete fashion as follows.

(a) Choose a distance d sampled from the Gaussian distribution,

$$\frac{1}{(2\pi\Delta\tau)^{3N/2}}\exp\left[-\frac{1}{2}\frac{d^2}{\Delta\tau}\right].$$

(b) Move distance d away from $\mathbf{R}_i$ in a random direction.

(c) Replace the particle at $\mathbf{R}_i$ by m_i particles at this new point, where the integer m_i is chosen so that its expectation value is $e^{-V(\mathbf{R}_i)\Delta\tau}$.

After enough steps, the density distribution should approach the exact ground state with the only errors being those due to the finite time step.

The Green's-function Monte Carlo method manages to avoid this time-step error and sample the Green's function exactly. The details are complicated and not relevant here, but the method is based on the use of a Dyson equation to relate the Green's function of the full problem to the known Green's function of some simpler problem. The result is an integral equation which can be solved using Monte Carlo methods not too different from those used in the solution of Eq. (B13).

Both the diffusion and Green's-function QMC methods rely on a probabilistic interpretation of Eq. (B13), and this in turn relies on the assumption that both ψ and G are non-negative. The short-time approximation, Eq. (B22), shows that G really is greater than zero, but (as mentioned earlier) antisymmetry implies that all many-electron wave functions must have both positive and negative regions and so the assumption that $\psi\geq 0$ is not a good one. At first sight this objection seems trivial. After choosing an antisymmetric starting function ψ, we could simply write

$$\psi=\tfrac{1}{2}(\psi+|\psi|)+\tfrac{1}{2}(\psi-|\psi|)\ , \tag{B23}$$

and then propagate $\phi_+=\frac{1}{2}(\psi+|\psi|)$ and $\phi_-=\frac{1}{2}(\psi-|\psi|)$, which are both non-negative, separately according to the imaginary-time Schrödinger equation. The fermion ground state would then be the difference between these two positive states as τ goes to infinity. Unfortunately, both ϕ_+ and ϕ_- separately converge to the boson ground state, which is the lowest nodeless eigenfunction of $\hat{H}$. The components along the fermion ground state (which we extract when we do the subtraction) decay relative to the boson ground-state components like $e^{-\Delta E\tau}$, where ΔE is the difference between the fermion and boson ground-state energies [see Eq. (B6)]. So when we do the subtraction to obtain the fermion ground state, we are looking for the tiny difference between two enormous quantities, and are swamped by numerical noise.

One way around the sign problem is to make the fixed-node approximation[36,37] (other modes—transient estimation, the release-node method—are perhaps best viewed as ways of improving a fixed-node solution once it has been obtained). The idea is first to guess the nodal structure of the ground-state wave function, and then to get as close to the true ground-state wave function as possible with the constraint that these nodes remain fixed. For this procedure to make sense, the nodal structure must correspond to some antisymmetric trial function, which could be the same Hartree-Fock determinant times a Jastrow factor used as the importance weighting, for example.

Suppose that the nodes, and hence the regions they enclose, have been chosen, and that the time-independent Schrödinger equation has been solved in each of the regions separately. Because of the simple quadratic form, $-\frac{1}{2}\nabla_{\mathbf{R}}^2+V(\mathbf{R})$, of the Hamiltonian, the eigenproblem in any particular region is defined completely by the boundary condition that the wave function must be zero everywhere on the enclosing nodal surface. The eigenfunctions and eigenvalues are therefore independent of what is happening in other regions. (This would not be true if the Hamiltonian contained higher derivatives or a nonlocal integral operator such as the angular-momentum projector $\hat{P}_l$ from Appendix A—hence the attempt to find practical BCC-type pseudo-Hamiltonians). Nodal surfaces in $3N$ dimensions are so complicated that finding the solutions in the individual regions would be a formidable task, but at least the lowest eigenfunction inside each region is nodeless and so there would be no sign problem if a Monte Carlo method was used.

The lowest eigenfunction in any given region is not totally antisymmetric, but the ground states from all the different regions can be assembled (with the appropriate signs and weights) to give an antisymmetric function ψ_{FN}, which must satisfy the variational principle

$$\frac{\langle\psi_{FN}|\hat{H}|\psi_{FN}\rangle}{\langle\psi_{FN}|\psi_{FN}\rangle}\geq E_0\ , \tag{B24}$$

where E_0 is the fermion ground-state energy of the whole system. If the fixed nodal structure is exactly the same as the true ground-state nodal structure, then the ground states in each region (which are completely specified given the form of the Hamiltonian and the shape of the surrounding nodal surface) must all equal the corresponding pieces of the full ground-state wave function.

ψ_{FN} is then the same as the true fermion ground state ψ_0 and the equality holds in Eq. (B24). If the fixed nodal structure is not quite right, then ψ_{FN} may have slope discontinuities across the nodal surfaces and these become δ functions under the action of the $\nabla_{\mathbf{R}}^2$ operator in the Hamiltonian. Luckily, the δ functions all occur at places where $\psi_{FN}=0$ and so contribute nothing to the expectation value, $\langle \psi_{FN}|\hat{H}|\psi_{FN}\rangle$. It is therefore still possible to consider the expectation value as a sum of contributions from the interiors of all the regions and there is no need to worry about boundary terms.

It is not very difficult to show that the various regions bounded by the nodal surfaces of some antisymmetric trial function can always be classified into one or more classes. If two regions are in the same class, then there is at least one permutation of the electron coordinates which brings about a one-to-one mapping from every point in one region to every point in the other. If two regions are in different classes, then there are no permutations of the electron coordinates which map any point in one region to any point in the other. Since the Hamiltonian is invariant under all permutations, it is clear that the solution of the fixed-node eigenproblem is essentially equivalent in all regions belonging to the same class: the eigenvalues are all the same, and the eigenfunctions differ only by permutations of the electron coordinates.

We can write ψ_{FN} as a sum of contributions from each class,

$$\psi_{FN}(\mathbf{R})=\sum_c \psi_c(\mathbf{R}) , \tag{B25}$$

where $\psi_c(\mathbf{R})$ equals $\psi_{FN}(\mathbf{R})$ in all regions belonging to class c but is zero otherwise. Because there are no permutations which map points from regions in one class to regions in another, each $\psi_c(\mathbf{R})$ is separately antisymmetric and so is an acceptable variational function,

$$\frac{\langle \psi_c|\hat{H}|\psi_c\rangle}{\langle \psi_c|\psi_c\rangle} \geq E_0 . \tag{B26}$$

Again the boundary terms vanish, and so we see that the lowest eigenvalue for each and every class (and hence for each and every region) satisfies its own variational principle. This is a stronger statement than Eq. (B24): if we solve the fixed-node eigenproblem in only one region belonging to each class, then even the lowest of the eigenvalues obtained is greater than or equal to the ground-state energy. [N.B. If the nodal structure corresponds to a trial function which is the ground state of some sort of fermion eigenproblem (the Hartree-Fock problem, for example), then it can be shown that there is only one class and so Eqs. (B24) and (B26) are equivalent. This is a generalization of the statement that the ground state of a one-electron system is nodeless and the proof is similar].

Now that we have explained all this, it is easy to understand how a fixed-node QMC calculation works and why it gives a variational approximation to the ground-state energy. The initial points (QMC walkers) are scattered through *all* the regions where the trial function is positive, with a density distribution proportional to the trial function itself (times the importance weighting, which is probably the same as the trial function). These points then diffuse around and die and multiply in the usual way, but with the boundary condition that any point which diffuses into a region where the trial function is negative is removed from the simulation. This is equivalent to putting infinite repulsive potentials in all the negative regions [cf. Eq. (B22)], and so ensures that the wave function remains equal to zero at the fixed nodes. The fixed-node eigenvalue problem is thus solved simultaneously in all the regions where the trial function is positive, and the nodal surfaces need never be determined explicitly. After a long enough (imaginary) time, the density distribution in each region converges to the nodeless ground state, with the population growth or decay rate given by the corresponding eigenvalue. The highest growth and lowest decay rates are in regions belonging to the class with the lowest ground-state energy, and so eventually the vast majority of particles are in these regions. The final variational estimate of the energy is then the lowest among the ground-state eigenvalues for all the different classes.

[1] W. A. Harrison, *Pseudopotentials in the Theory of Metals* (Benjamin, New York, 1966).

[2] V. Heine, in *Solid State Physics*, edited by H. Ehrenreich, F. Seitz, and D. Turnbull (Academic, New York, 1970), Vol. 24.

[3] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).

[4] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).

[5] C. J. Callaway and N. H. March, in *Solid State Physics*, edited by H. Ehrenreich, F. Seitz, and D. Turnbull (Academic, New York, 1984), Vol. 38.

[6] M. H. Kalos and P. A. Whitlock, *Monte Carlo Methods* (Wiley, New York, 1986), Vol. 1.

[7] D. M. Ceperley and M. H. Kalos, in *Monte Carlo Methods in Statistical Physics*, edited by K. Binder (Springer-Verlag, Berlin, 1979).

[8] K. E. Schmidt and M. H. Kalos, in *Applications of the Monte Carlo Method in Statistical Physics*, edited by K. Binder (Springer-Verlag, Berlin, 1984).

[9] G. B. Bachelet, D. M. Ceperley, and M. G. B. Chiocchetti, Phys. Rev. Lett. **62**, 2088 (1989).

[10] B. L. Hammond, P. J. Reynolds, and W. A. Lester, Jr., Phys. Rev. Lett. **61**, 2312 (1988).

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

[12] D. M. Ceperley and B. J. Alder, Phys. Rev. Lett. **45**, 566 (1980).

[13] D. M. Ceperley and B. J. Alder, Phys. Rev. B **36**, 2092 (1987).

[14] S. Fahy, X. W. Wang, and S. G. Louie, Phys. Rev. Lett. **61**, 1631 (1988).

[15] D. M. Ceperley, J. Stat. Phys. **43**, 815 (1986).

[16] J. W. Negele and H. Orland, *Quantum Many Particle Systems* (Addison-Wesley, Reading, 1988).

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

[18]G. P. Kerker, J. Phys. C **13**, 189 (1980).
[19]G. B. Bachelet, D. R. Hamann, and M. Schluter, Phys. Rev. B **26**, 4199 (1982).
[20]J. Ihm, A. Zunger, and M. L. Cohen, J. Phys. C **12**, 4409 (1979); see also corrigendum: J. Ihm, A. Zunger, and M. L. Cohen, *ibid.* **13**, 3095 (1980).
[21]D. D. Koelling and B. N. Harmon, J. Phys. C **10**, 3107 (1977).
[22]H. Gollisch and L. Fritsche, Phys. Status Solidi B **86**, 145 (1978).
[23]G. B. Bachelet and M. Schluter, Phys. Rev. B **25**, 2103 (1982).
[24]D. R. Hamann, Phys. Rev. B **40**, 2980 (1989).
[25]G. B. Bachelet (private communication).
[26]A. Bosin and G. B. Bachelet (unpublished).
[27]S. G. Louie, S. Froyen, and M. L. Cohen, Phys. Rev. B **26**, 1738 (1982).
[28]A. M. Rappe, K. M. Rabe, E. Kaxiras, and J. D. Joannopoulos, Phys. Rev. B **41**, 1227 (1990).
[29]J. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).
[30]M. T. Yin and M. L. Cohen, Phys. Rev. B **26**, 5668 (1982).
[31]E. L. Shirley, D. C. Allan, R. M. Martin, and J. D. Joannopoulos, Phys. Rev. B **40**, 3652 (1989).
[32]L. Kleinman and D. M. Bylander, Phys. Rev. Lett. **48**, 1425 (1982).
[33]D. Vanderbilt, Phys. Rev. B **41**, 7892 (1990).
[34]X. Gonze (private communication).
[35]N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
[36]P. J. Reynolds, D. M. Ceperley, B. J. Alder, and W. A. Lester, J. Chem. Phys. **77**, 5593 (1982).
[37]J. W. Moskowitz, K. E. Schmidt, M. A. Lee, and M. H. Kalos, J. Chem. Phys. **77**, 349 (1982).