PHYSICAL REVIEW B VOLUME 54, NUMBER 3 15 JULY 1996-I

## ARTICLES

# Generalized-gradient functionals in adaptive curvilinear coordinates

D. R. Hamann
*Bell Laboratories, Lucent Technologies, Murray Hill, New Jersey 07974*

(Received 5 February 1996)

The incorporation of generalized-gradient exchange-correlation functionals within the adaptive curvilinear coordinate method for electronic structure calculations is discussed. It is demonstrated that a revised formulation of the exchange-correlation potential yields greatly improved convergence. Modifications of the forces on the adaptive coordinates compared to the local density case are derived. A strategy for incorporating gradient corrections at minimal computational cost compared to adaptive local-density calculations is outlined. [S0163-1829(96)01227-1]

## I. INTRODUCTION

The adaptive curvilinear coordinate (ACC) method was recently introduced by Gygi as an enhancement of the plane-wave pseudopotential approach to *ab initio* calculations of materials properties.[1,2] This method allows materials with highly localized valence orbitals, such as those of first-row and transition-metal atoms, to be treated with a much smaller set of basis functions than possible in a conventional plane-wave calculation. Electron wave functions are expanded in plane waves in a generalized curvilinear coordinate system. Transformed to ordinary Euclidean coordinates, the spatial frequency of such a ''plane'' wave becomes a varying function, and surfaces of constant phase are no longer planes. When the parameters describing the coordinate transformation are allowed to adapt to minimize the energy of the system, spatial frequencies tend to be peaked where occupied atomic orbitals are most localized. The effective plane-wave kinetic energy cutoff is greatly enhanced only where necessary, which is the means by which the total number of basis functions can remain relatively small.

Gygi showed that an ACC formulation of the density-functional formalism within the local-density approximation[3] (LDA) retains all the most important advantages of ordinary plane waves, including basis orthogonality, absence of Pulay forces, and the ability to use fast Fourier transforms (FFT's) and the Car-Parrinello (CP) approach[4] to efficiently optimize wave functions and atomic geometries.[2] While his demonstration examples were isolated molecules in large supercells, we subsequently explored the application of this approach to $SiO_2$, and found that large effective cutoff enhancements were also achieved in the context of continuous solids.[5]

It was recently discovered that the LDA is totally inadequate for describing the structural phase transition of $SiO_2$ from the stable quartz structure to the metastable polymorph stishovite, which is formed at high pressures. The experimental energy difference between the two structures is 0.54 eV per formula unit, and the LDA result is zero, within a few hundredths of an eV.[6–10] Application of an improved exchange-correlation energy functional developed by Perdew and Wang[11] (PW91), however, yields an energy difference of 0.57 eV, and a coexistence pressure in excellent agreement with experiment.[10]

PW91 is one of a class of functionals known as generalized-gradient approximations (GGA's), which depend on both the local electron charge density and the magnitude of its gradient.[11–13] It was selected for these calculations because it was formulated solely on the basis of sum rules, scaling relations, known asymptotic limits, and quantum Monte Carlo electron-gas results.[11,14] It contains no adjustable parameters. Its functional form, however, is not unique, and further improvements of GGA functionals may be possible in the future. In the course of carrying out these calculations, it was necessary to develop efficient means for incorporating the GGA within the ACC method.

Convergence and stability problems associated with incorporation of the GGA into conventional plane-wave calculations are well known to specialists in the field, and are discussed in detail by White and Bird.[15] The common solution to these problems is to substantially increase the wave vector cutoff of the FFT's beyond that required to represent the charge density exactly for a given cutoff of the wave function basis set. The problems are caused by high spatial frequencies present in the exchange-correlation potential $v_{xc}$. In initial experiments with ACC-GGA, we found that it was necessary to at least double the average kinetic cutoff of the charge and potential expansion to obtain convergence comparable to the LDA, paralleling the experience of others with conventional plane waves. A significant computational penalty was associated with the increased size of the FFT's, whose execution dominates the ACC calculations.

An alternative solution to the stability problem was proposed by Bird and White. Conventionally, $v_{xc}$ is defined as the variational derivative of the integral defining the exchange-correlation energy $E_{xc}$ with respect to the charge density. These authors proposed varying the discrete sum on the FFT grid which approximates the $E_{xc}$ integral with respect to the densities on this set of points. Replacing the variational derivative with a set of ordinary derivatives leads to a modified $v_{\mathrm{xc}}$ defined only on the grid points. The modified $v_{\mathrm{xc}}$ converges to the standard $v_{\mathrm{xc}}$ in the large-cutoff limit.[15] This solution is not satisfactory for the ACC method. The total energy must be minimized not only with respect to the wave functions, but with respect to the coordinate transformation, so an energy functional defined continuously throughout space is necessary.

Nontrivial extensions of Gygi's formalism were necessary to incorporate the GGA in the ACC method. In the course of developing these extensions, we discovered that replacing the exchange-correlation potential with an alternative but mathematically equivalent operator greatly improves convergence and stability. The original formulation of density-functional theory demonstrates that the total energy is minimized by a charge density constructed from a set of orthogonal wave functions that satisfy a Schrödinger-like equation with a ''mean-field'' single-electron potential.[3] The exchange-correlation potential $v_{\mathrm{xc}}$ is one component of this potential, and the Schrödinger equation is the Euler equation of the energy functional, derived by following conventional procedures of variational calculus. For a functional with a GGA form of $E_{\mathrm{xc}}$, however, it is possible to formulate an alternative Euler equation. This form contains not just the Laplacian kinetic energy operator of the conventional equation, but gradient operators dotted with spatially varying vector coefficients, and a modified scalar exchange-correlation potential. The effective Hamiltonian operator remains Hermitian.

The advantage of this formulation derives from the fact that the spatial frequencies of the scalar and vector ''potential'' functions remain comparable to those of the charge density, so that extended cutoffs are not required. In a conventional plane-wave calculation, the additional FFT's required to evaluate the action of the vector GGA operator on the wave functions would obviate this advantage. In the ACC method, however, such vector operator terms are already required to describe the kinetic energy in curvilinear coordinates, so no significant additional computation is required.

This operator formulation is derived in Sec. II. In Sec. III we discuss the evaluation of the exchange-correlation contributions to the Hamiltonian matrix elements within the ACC method. We also provide formulas needed to modify the LDA ACC formalism[2] for the forces needed to allow the coordinate system to adapt. The convergence of the action of the exchange-correlation portion of the Hamiltonian on the wave functions is illustrated in Sec. IV, and compared to the conventional $v_{\mathrm{xc}}$ formulation. Uses of the GGA in other versions of the ACC method are discussed in Sec. V, and the present results are summarized. A generalization of expressions in Sec. III is given in the Appendix.

## II. EXCHANGE-CORRELATION VECTOR OPERATOR

Within any GGA, the exchange-correlation energy is given by an integral of the electron charge density $\rho(\mathbf{x})$ times an energy density $\varepsilon$,

$$E_{\mathrm{xc}}=\int d^3x\ \rho\varepsilon(\rho,\sigma), \tag{1}$$

where $\sigma=|\boldsymbol{\nabla}\rho|$. (For future use, we will define $\sigma$ with a subscript as a Cartesian component of the gradient, $\sigma_i=\partial\rho/\partial x_i$.) Various approximations differ in the functional form of $\varepsilon$. These functions typically employ a dimensionless combination of $\sigma$ and $\rho$ as their second argument, but the ensuing equations are simpler if $\varepsilon$ is treated as a function of independent $\rho,\sigma$ variables, and the translation to another form is straightforward. A spin-polarized version of these results should also be straightforward. The conventional exchange-correlation potential, defined as

$$v_{\mathrm{xc}}(\mathbf{x})\equiv\frac{\delta E_{\mathrm{xc}}}{\delta\rho(\mathbf{x})}, \tag{2}$$

is found using standard variational calculus to be

$$\begin{aligned} v_{\mathrm{xc}}=\varepsilon+\rho\varepsilon_{\rho}-\sigma\varepsilon_{\sigma}-\rho\sigma\varepsilon_{\rho\sigma}-\rho\sigma^{-1}\varepsilon_{\sigma}\nabla^2\rho \\ +(\rho\sigma^{-2}\varepsilon_{\sigma}-\rho\sigma^{-1}\varepsilon_{\sigma\sigma})\boldsymbol{\nabla}\rho\cdot\boldsymbol{\nabla}|\boldsymbol{\nabla}\rho|, \end{aligned} \tag{3}$$

where $\varepsilon_\rho=\partial\varepsilon/\partial\rho$, $\varepsilon_{\rho\sigma}=\partial^2\varepsilon/\partial\rho\,\partial\sigma$, etc. Examining Eq. (3), the sources of the convergence difficulties become clear. $|\boldsymbol{\nabla}\rho|$ has cusps at extrema of $\rho$, so $\boldsymbol{\nabla}|\boldsymbol{\nabla}\rho|$ is discontinuous, and contains arbitrarily high spatial frequencies even when the maximum spatial frequencies in $\rho$ are limited. While $\boldsymbol{\nabla}^2\rho$ has limited spatial frequencies, the Laplacian emphasizes the high-frequency terms. Since $\varepsilon$ is not a polynomial function of its arguments, high spatial frequencies can be generated even in the LDA case where only $\varepsilon$ and $\varepsilon_\rho$ occur in $v_{\mathrm{xc}}$. In practice, the functional dependence on $\rho$ is such that this is not a problem. The GGA $\varepsilon(\rho,\sigma)$, however, can be a rapidly varying function of its arguments in some ranges of their values, and the second derivatives in $v_{\mathrm{xc}}$ form another source of high spatial frequencies. Aliasing errors, due to the fact that the FFT grid cannot accurately represent the high spatial frequencies, occur both in the calculation of the higher-order gradients and in the evaluation of the wave-function matrix elements,

$$\langle\psi_j|H_{\mathrm{xc}}|\psi_i\rangle\equiv\langle\psi_j|\ \frac{\delta E_{\mathrm{xc}}}{\delta\psi_j^*}=\int d^3x\ v_{\mathrm{xc}}\psi_j^*\psi_i\,. \tag{4}$$

The alternative expression is derived by going back to the basic variational formulation,

$$\delta E_{\mathrm{xc}}=\int d^3x\left[\varepsilon+\rho\varepsilon_{\rho}+\rho\sum_{i=1}^{3}\left(\frac{\partial\varepsilon}{\partial\sigma_i}\right)\frac{\partial}{\partial x_i}\right]\delta\rho. \tag{5}$$

Now we express $\delta\rho$ in terms of the occupied wave functions,

$$\delta\rho=\sum_k\ (\psi_k^*\,\delta\psi_k+\delta\psi_k^*\,\psi_k). \tag{6}$$

Following the conventional route, we would integrate by parts in Eq. (5), and recover Eq. (4) with $v_{\mathrm{xc}}$ given by Eq. (3). Suppose instead we substitute Eq. (6) in Eq. (5) and do not integrate by parts. We still have a perfectly valid and mathematically equivalent expression, and find

$$\begin{aligned} \langle\psi_j|H_{\mathrm{xc}}|\psi_i\rangle=\int d^3x[\tilde{v}_{\mathrm{xc}}\psi_j^*\psi_i+\psi_j^*(\mathbf{V}_{\mathrm{xc}}\cdot\boldsymbol{\nabla}\psi_i) \\ +(\mathbf{V}_{\mathrm{xc}}\cdot\boldsymbol{\nabla}\psi_j^*)\psi_i], \end{aligned} \tag{7}$$

where $\tilde{v}_{\mathrm{xc}}=\varepsilon+\rho\varepsilon_\rho$ and $\mathbf{V}_{\mathrm{xc}}=\rho\sigma^{-1}\varepsilon_\sigma\boldsymbol{\nabla}\rho$. $\mathbf{V}_{\mathrm{xc}}$ is smoothly varying at extrema of $\rho$, since $\varepsilon\rightarrow\varepsilon_0+\varepsilon_2\sigma^2+\cdots$ as $\sigma\rightarrow 0$,[11] and $\varepsilon_\sigma$ thus cancels the $\sigma^{-1}$ factor. By eliminating the $\nabla^2\rho$, $\boldsymbol{\nabla}|\boldsymbol{\nabla}\rho|$, and second derivatives of $\varepsilon$ which appear in Eq. (4), we obtain an integrand that is much more rapidly convergent on the FFT mesh. This will be demonstrated in Sec. IV.

## III. CURVILINEAR COORDINATE FORMULATION

Followng Gygi,[1,2] we introduce the curvilinear coordinate $\boldsymbol{\xi}$, which is a vector of three Cartesian components, and $\boldsymbol{\xi}$-space lattice defined by a set of vectors $\{\mathbf{R}\}$ which are identical to the ordinary lattice vectors of the solid (or supercell) under consideration. The transformation from curvilinear to Euclidean coordinates $\mathbf{x}$ is given by

$$\mathbf{x}(\boldsymbol{\xi})=\boldsymbol{\xi}+\sum_{\mathbf{G}}\mathbf{x}_{\mathbf{G}}\exp[i\mathbf{G}\cdot\boldsymbol{\xi}],\tag{8}$$

where $\mathbf{G}$ is a three-component reciprocal lattice vector in reciprocal space conjugate to $\{\mathbf{R}\}$, and $\mathbf{x}_{\mathbf{G}}$ are the (three-component vector) parameters defining the curvilinear coordinates. Note that while $\{\mathbf{G}\}$ are vectors in $\boldsymbol{\xi}$ (curvilinear) space, they are identical to the reciprocal lattice in ordinary Euclidean space, and that the dot product in the exponent of Eq. (8) is the usual sum of products of three pairs of Cartesian components. The $\mathbf{G}$ sum in Eq. (8) is cut off at a maximum value such that $\hbar^2G^2/2m\leqslant E_x$, where we introduce a cutoff energy for convenience in comparison with the other cutoffs of the problem. ($E_x$ has no meaning as a kinetic energy.) It is straightforward to evaluate the sum in Eq. (8) by FFT's for each Cartesian component of $\mathbf{x}_{\mathbf{G}}$, and thus to map the uniform FFT mesh in $\xi$ space onto a corresponding nonuniform mesh in $\mathbf{x}$ space.

The Reimannian metric tensor for the curvilinear system is given by

$$g_{ij}=\sum_k\frac{\partial x_k}{\partial\xi_i}\frac{\partial x_k}{\partial\xi_j},\tag{9}$$

which is straightforward to evaluate for the transformation given in Eq. (8). We generalized Gygi's set of basis functions to include a Bloch wave vector $\mathbf{k}$ in the first Brillouin zone.[16] These functions are given by

$$\chi_{\mathbf{k},\mathbf{G}}(\mathbf{x})=\frac{1}{\sqrt{\Omega}}\,g^{-1/4}(\mathbf{x})\exp[i(\mathbf{k}+\mathbf{G})\cdot\boldsymbol{\xi}(\mathbf{x})],\tag{10}$$

where $g=\det g_{ij}$, and $\Omega$ is the unit cell volume. The $\chi$ basis is orthonormal and complete.[1,2,16]

A Bloch function is expanded as a sum over the $\chi$ basis set,

$$\psi_{n\mathbf{k}}(\mathbf{x})=\frac{g^{-1/4}(\mathbf{x})}{\sqrt{\Omega}}\sum_{\mathbf{G}}c_{n\mathbf{k},\mathbf{G}}\exp[i(\mathbf{k}+\mathbf{G})\cdot\boldsymbol{\xi}(\mathbf{x})].\tag{11}$$

We terminate the expansion so that $\hbar^2|\mathbf{k}+\mathbf{G}|^2/2m\leqslant E_{\mathrm{cut}}$, where $E_{\mathrm{cut}}$ has the interpretation of an average kinetic energy cutoff.

The charge density $\rho(\mathbf{x})$ is given by the sum over occupied states $n$ and Brillouin zone sample points $\mathbf{k}$ with appropriate weights $w_{n\mathbf{k}}$. All computation in the ACC method is actually carried out in $\boldsymbol{\xi}$ space, and it is more convenient to compute a modified charge density,

$$\tilde{\rho}(\xi)=1/\Omega\sum_{n\mathbf{k}}w_{n\mathbf{k}}\left|\sum_{\mathbf{G}}c_{n\mathbf{k},\mathbf{G}}e^{i\mathbf{G}\cdot\boldsymbol{\xi}}\right|^2,\tag{12}$$

and its Fourier transform in $\boldsymbol{\xi}$ space,

$$\tilde{\rho}_{\mathbf{G}}=\int\tilde{\rho}(\boldsymbol{\xi})e^{-i\mathbf{G}\cdot\boldsymbol{\xi}}d^3\xi.\tag{13}$$

We note that if the Brillouin zone sum is over an irreducible wedge of the zone, $\tilde{\rho}_{\mathbf{G}}$ needs to be appropriately symmetrized, and back-transformed to produce a symmetric $\tilde{\rho}(\boldsymbol{\xi})$. The charge density itself is the product $\rho=g^{-1/2}\tilde{\rho}$. The components $\sigma_j$ of $\boldsymbol{\nabla}\rho$ are given by

$$\sigma_j(\boldsymbol{\xi})=g^{-1/2}\sum_{k=1}^{3}\frac{d\xi_k}{dx_j}\left(\tilde{\rho}A_k+\sum_{\mathbf{G}}iG_k\tilde{\rho}_{\mathbf{G}}e^{i\mathbf{G}\cdot\boldsymbol{\xi}}\right).\tag{14}$$

The $\mathbf{G}$ sum is evaluated by three FFT's, and the frequently needed quantity $d\xi_k/dx_j$ is most conveniently computed as

$$\frac{d\xi_i}{dx_j}=\sum_{k=1}^{3}g^{ik}\frac{dx_j}{d\xi_k},\tag{15}$$

where $g^{ij}$ is the conjugate (inverse) of the metric tensor $g_{ij}$ defined in Eq. (9). The geometric "vector potential" $A_i$ is

$$A_i=\frac{1}{4}\frac{\partial}{\partial\xi_i}\ln g=\frac{1}{2}\sum_{jkn=1}^{3}g^{kj}\frac{\partial^2x_n}{\partial\xi_i\partial\xi_k}\frac{\partial x_n}{\partial\xi_j}.\tag{16}$$

$A_i$ and the other geometric quantities are needed for many parts of the ACC calculations, and are best calculated after each update of the coordinate transformation parameters $\mathbf{x}_{\mathbf{G}}$ and saved.

After the $\sigma_j$ have been calculated, it is straightforward to evaluate the GGA exchange-correlation energy,

$$E_{\mathrm{xc}}=\int\tilde{\rho}\varepsilon(\rho,\sigma)d^3\xi,\tag{17}$$

where the integral is over the $\xi$ unit cell. Since $d^3x=g^{1/2}d^3\xi$, $\rho$ in the $\mathbf{x}$ space integral, Eq. (1), is converted into $\tilde{\rho}$ in the $\boldsymbol{\xi}$ space integral. In transforming Eq. (7) for the matrix elements of $H_{\mathrm{xc}}$ for efficient evaluation in the $\chi_{\mathbf{k},\mathbf{G}}$ basis, we introduce a modified form of the vector potential $\mathbf{V}_{\mathrm{xc}}$ which incorporates one of the geometric functions associated with the calculation of $\boldsymbol{\nabla}\chi$,

$$\tilde{V}_{\mathrm{xc},j}=\rho\sigma^{-1}\varepsilon_\sigma\sum_{k=1}^{3}\frac{d\xi_j}{dx_k}\sigma_k.\tag{18}$$

The matrix element is now given by

$$\langle \mathbf{k},\mathbf{G}|H_{\mathrm{xc}}|\mathbf{k},\mathbf{G}'\rangle = \frac{1}{\Omega}\int d^3\xi\Bigg[\tilde{v}+\sum_{p=1}^{3} \times(k_p+G_p-iA_p)(-i\tilde{\mathbf{V}}_p) + \sum_{q=1}^{3}(i\tilde{V}_q)(k_q+G'_q+iA_q)\Bigg]e^{i(\mathbf{G}'-\mathbf{G})\cdot\boldsymbol{\xi}}. \tag{19}$$

In presenting Eq. (19), we have omitted some simplifying cancellations and inserted appropriate factors of $i$ to emphasize the resemblance of this expression to the matrix element of the kinetic energy, which was given by Gygi[1,2,17] as

$$\langle \mathbf{k},\mathbf{G}|-\tfrac{1}{2}\nabla^2|\mathbf{k},\mathbf{G}'\rangle = \frac{1}{2\Omega}\int d^3\xi\sum_{pq=1}^{3}(k_p+G_p-iA_p)g^{pq} \times(k_q+G'_q+iA_q)e^{i(\mathbf{G}'-\mathbf{G})\cdot\boldsymbol{\xi}}. \tag{20}$$

For each time step in the fictitious molecular dynamics of the CP method, we must evaluate the matrix elements in the $\chi$ basis set of the Hamiltonian acting on each of the occupied wave functions.[4] With our new formulation for $H_{\mathrm{xc}}$, the evaluation of this term and the kinetic energy operating on each wave function are combined,

$$\begin{aligned}\langle \mathbf{k},\mathbf{G}|-\tfrac{1}{2}\nabla^2+H_{\mathrm{xc}}|\psi_{\mathbf{k}n}\rangle &= \frac{1}{2\Omega}\sum_{p,q=1}^{3}\int d^3\xi\sum_{\mathbf{G}'}c_{\mathbf{k}n,\mathbf{G}'}\\ &\quad\times[2\tilde{v}/9+(k_p+G_p-iA_p)(-2i\tilde{V}_p/3)\\ &\quad+(2i\tilde{V}_q/3)(k_q+G'_q+iA_q)+(k_p+G_p-iA_p)g^{pq}\\ &\quad\times(k_q+G'_q+iA_q)]e^{i(\mathbf{G}'-\mathbf{G})\cdot\boldsymbol{\xi}}.\end{aligned} \tag{21}$$

The $\mathbf{G}'$ sum in Eq. (21) can be evaluated by performing an inverse FFT of the $c_{\mathbf{k}n,\mathbf{G}'}$ and three inverse FFT's of $G'_q c_{\mathbf{k}n,\mathbf{G}'}$ for $q=1,3$. The $q$ sum can then be performed, and terms combined to give four functions of $\xi$. The $\xi$ integral over these functions can now be evaluated as four FFT's, and the $p$ sum performed, efficiently yielding the complete set of $\langle\mathbf{k},\mathbf{G}|$ matrix elements. [The factors 2/3 and 2/9 are included in Eq. (21) so the unsummed, single-summed, and double-summed quantities can be combined for more compact notation.]

The introduction of $|\nabla\rho|$ into $\varepsilon$ invalidates the derivation given by Gygi for the exchange-correlation contribution to the forces acting on the coordinate transformation coefficients $\mathbf{x}_{\mathbf{G}}$. The derivation of the required equations for the GGA is simplified by the fact that $\tilde{\rho}$ is only a function of $\xi$, and is independent of variations of the coordinate transformation itself, $\delta x_p$. $\rho=g^{-1/2}\tilde{\rho}$ depends only on $\xi$ derivatives of these variations, and $\sigma=|\nabla\rho|$ depends on first and second $\xi$ derivatives of the variation. It is convenient to split the variation of $E_{\mathrm{xc}}$ into three terms,

$$\delta E_{\mathrm{xc1}} = -\sum_{ip=1}^{3}\int d^3\xi\;\tilde{\rho}\rho\varepsilon_{\rho}\frac{\partial\xi_i}{\partial x_p}\frac{\partial\delta x_p}{\partial\xi_i}, \tag{22}$$

$$\delta E_{\mathrm{xc2}} = -\sum_{ijp=1}^{3}\int d^3\xi\;\tilde{\rho}\varepsilon_{\sigma}\sigma^{-1}\sigma_j\Bigg[\sigma_p\frac{\partial\xi_i}{\partial x_j}+\sigma_j\frac{\partial\xi_i}{\partial x_p} -\rho\sum_{kmn=1}^{3}\frac{\partial\xi_k}{\partial x_j}\frac{\partial\xi_m}{\partial x_p}\frac{\partial^2 x_n}{\partial\xi_k\partial\xi_m}\frac{\partial\xi_i}{\partial x_n}\Bigg]\frac{\partial\delta x_p}{\partial\xi_i}, \tag{23}$$

$$\delta E_{\mathrm{xc3}} = -\frac{1}{2}\sum_{ijkp=1}^{3}\int d^3\xi\;\tilde{\rho}\rho\varepsilon_{\sigma}\sigma^{-1}\sigma_j \times\left[\frac{\partial\xi_i}{\partial x_j}\frac{\partial\xi_k}{\partial x_p}+\frac{\partial\xi_k}{\partial x_j}\frac{\partial\xi_i}{\partial x_p}\right]\frac{\partial^2\delta x_p}{\partial\xi_i\partial\xi_k}. \tag{24}$$

$\delta E_{\mathrm{xc1}}$ is present in the LDA, and Eq. (22) can be manipulated into the form given in Ref. 2 by integration by parts. Most of the multiple sum in the last term in brackets in Eq. (23) can be collapsed into a third-rank tensor which is already present in LDA calculations, where it is needed to evaluate the variation of the kinetic energy with the coordinate transformation. We note that Eqs. (22)–(24) must be modified in the presence of model core charges,[18] and this case is discussed in the Appendix.

The total variation of the energy with the coordinate transformation can be written as

$$\delta E = \int d^3\xi\Bigg[\sum_{p=1}^{3}B_p\delta x_p+\sum_{pj=1}^{3}C_{pj}\frac{\partial\delta x_p}{\partial\xi_j} +\sum_{pjk=1}^{3}D_{pjk}\frac{\partial^2\delta x_p}{\partial\xi_j\partial\xi_k}\Bigg]. \tag{25}$$

Contributions to the three tensors $B$, $C$, and $D$ from the kinetic energy and separable nonlocal pseudopotentials must be accumulated during the computational loop over occupied states to compute $\langle\mathbf{k},\mathbf{G}|H|\psi_{\mathbf{k}n}\rangle$. When properly embedded in the matrix element calculations, these contributions require only a small additional computational effort. Contributions from the Coulomb interaction, local ion pseudopotentials, and exchange-correlation energy are calculated once, outside the state loop. While the $\delta E_{\mathrm{xc}}$ only contributed to $B$ in the LDA as formulated in Ref. 2, the GGA form requires contributions to $C$ and $D$. This does not increase the complexity of evaluating the forces on the $\mathbf{x}_{\mathbf{G}}$, since other terms present in the LDA already contribute to these tensors. When $\delta x^p$ is expressed in Eq. (25) in terms of the Fourier series for the coordinate transformation, Eq. (8), the $\xi$ derivatives simply bring down $\mathbf{G}$ components from the exponentials, the $\xi$ integrals over the independent components of $B$, $C$, and $D$ can be evaluated by 30 FFT's, and the various coordinate index sums performed to evaluate the complete set of coordinate forces $\partial E/\partial x_{\mathbf{G},p}$.

## IV. CONVERGENCE

The convergence of the standard and vector operator forms of the GGA Hamiltonian operating on the wave functions was evaluated by comparing numerical and analytic derivatives of $E_{\mathrm{xc}}$. While this procedure was first employed simply as a test of the programs during their development, it provides strong evidence for the superiority of the vector formulation.

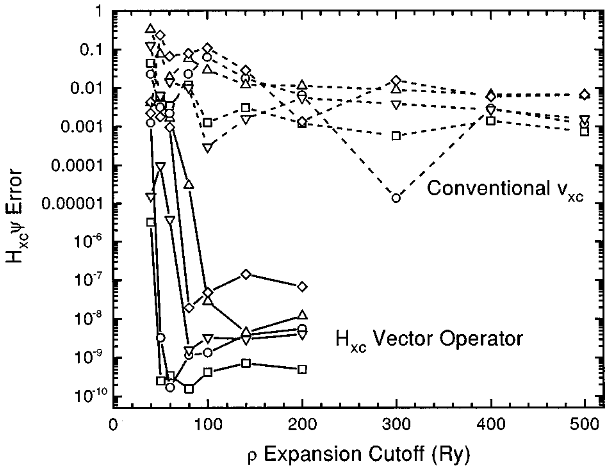

FIG. 1. Magnitude of the relative error in the evaluation of $H_{xc}\psi$ for five randomly selected wave functions $\psi$ and $\chi_{\mathbf{k},\mathbf{G}}$ components for $\alpha$-quartz. The same symbol denotes the same component in conventional and vector calculations.

To illustrate this comparison, we use $\alpha$-quartz, with the pseudopotentials and other parameters of Ref. 5. The average cutoff for the wave-function expansion is 25 Ry, and the adaptation of the coordinates boosts this to a maximum effective cutoff of 70 Ry. The cutoff for the coordinate transformation is 5 Ry. Two **k** points were used to sample the irreducible wedge of the Brillouin zone. As a measure of the error, we evaluate

$$\mathrm{Err}_{\mathbf{k}n,\mathbf{G}} = \left|\mathrm{Re}(\langle \mathbf{k},\mathbf{G}|H_{xc}|\psi_{kn}\rangle)/\partial E_{xc}/\partial\, \mathrm{Re}(c_{\mathbf{k}n,\mathbf{G}}) - 1\right|, \tag{26}$$

or a similar expression with the ratio of imaginary parts. The derivative is evaluated using a five-point formula. Five sets of **k**, $n$, **G**, and Re or Im were selected more or less randomly for illustration, and the error was calculated for a series of cutoffs of the FFT's used to evaluate the numerator and denominator. The results are shown in Fig. 1. An average cutoff of 100 Ry is the minimum necessary to calculate $\rho$ with full accuracy given the 25 Ry average wave-function cutoff. It was necessary to go far below this value to see any systematic cutoff dependence in the error of the vector operator calculation. By 100 Ry, the results had saturated at what are essentially roundoff error values. If there is any trend, it is that the rate of convergence and roundoff error are better for the smaller $|\mathbf{k}+\mathbf{G}|$'s in the sample.

By comparison, the convergence of the conventional $v_{xc}$ form is barely noticeable on the scale of ten decades shown in the figure, even out to a 500 Ry cutoff, which would be impossibly costly for practical calculations. In evaluating $v_{xc}$ for the conventional tests, both a straightforward calculation of $\nabla|\nabla\rho|$, and the form suggested in Ref. 15, $\nabla|\nabla\rho|^2/2|\nabla\rho|$, were tried. The former gave notably superior performance until the cutoff reached the 4–500 Ry range, and these results are shown in Fig. 1.

Initial experience with the standard GGA $v_{xc}$ indicated that reasonably stable convergence of CP dynamics[4] could be obtained with a 200–250 Ry cutoff for the FFT's for $SiO_2$ systems. For the vector operator form, the "standard" (average) cutoff of four times the wave-function cutoff gives GGA stability and convergence completely comparable to LDA calculations.

## V. GENERALIZATION AND SUMMARY

More recent alternatives to Gygi's adaptive plane-wave ACC method are being studied by several groups. One variation is to fix the coordinate deformation by a sum of atom-centered functions that have been optimized in an atomic calculation, or by some other criterion, but are not allowed to adapt otherwise.[19–22] In these cases, the strategy for calculating $H_{xc}$ matrix elements discussed here remains unchanged. The mathematics of calculating the various geometric functions must be somewhat modified, but as discussed above, these same functions are also needed for the LDA. In the atom-centered deformation case, one need not calculate forces on independent coordinate parameters. However, unlike the adaptive case considered here, the explicit dependence of the coordinate transformation on atomic positions gives rise to Pulay-like forces[23] that must be calculated using essentially the same formulas that are required in the adaptive case. The tensors $B$, $C$, and $D$ in Eq. (25) would be accumulated as discussed above, but $\delta x_p$ would be expressed in terms of motions of the atom centers, and the various derivatives of $\delta x^p$ calculated from the atom-centered analytic functions describing the coordinate transformation.

The other variation on the curvilinear coordinate method discussed here is the use of finite-difference formulations for derivatives, and the representation of wave functions and other quantities strictly by their numerical values on a uniform grid in $\boldsymbol{\xi}$ space.[20–22] Plane-wave expansions and FFT's are not utilized. While it is no longer appropriate to discuss aliasing errors in this case, the concept of the spatial frequency content of the exchange-correlation operators remains valid. The analog of the average cutoff in the ACC plane-wave case is the spacing of points $h$ on the $\boldsymbol{\xi}$ grid. The errors of the finite-difference calculations of $H_{xc}|\psi\rangle$ should be more rapidly convergent functions of $h$ with the vector operator form than the conventional $v_{xc}$ form. As in the plane-wave case, $\xi$ first derivatives of the wave functions must already be required by the Laplacian in curvilinear coordinates, so that the computational cost of implementing the vector operator form of $H_{xc}$ should be minimal.

In summary, we have outlined the incorporation of generalized-gradient approximation forms of density-functional theory within the framework of adaptive curvilinear coordinate electronic structure calculations. Recent results suggest a need for a GGA to correctly describe the physics of an important class of systems, $SiO_2$ and related materials.[10] The highly localized O orbitals of these materials place computational demands on electronic structure methods that are efficiently met by the ACC approach.[5] The forced merger of ACC and GGA in the author's work, two additions to standard plane-wave LDA methods which relate to very different aspects of the problems being investigated, led to the discovery of an unexpected synergy. The improved convergence of the reformulated GGA $H_{xc}$ operator, and the computational efficiency with which it can be incorporated into ACC calculations, make ACC-GGA an attractive combination for many difficult electronic structure problems.

## ACKNOWLEDGMENTS

The author wishes to thank J. P. Perdew for FORTRAN routines for calculating the PW91 functional. He also thanks F. Gygi, E. Kaxiras, and E. B. Stechel for valuable discussions and copies of their unpublished work.

## APPENDIX

In this appendix, we generalize the results of Sec. III to include model core charges. Such charges were introduced by Louie, Froyen, and Cohen to improve transferrability by representing the ''unscreening'' and ''rescreening'' processes in the derivation and application of *ab initio* pseudopotentials more accurately.[18] The usual approach is equivalent to a linearization of $v_{\mathrm{xc}}$ about the reference atomic pseudocharge density, and including the model core charge gives a more accurate representation of the nonlinear $\rho$ dependence of $\varepsilon_{\mathrm{xc}}$ and $v_{\mathrm{xc}}$.

The total charge density is

$$\begin{gathered} \rho=\rho_c+g^{-1/2}\tilde{\rho}, \\ \rho=\rho_c+\rho_e, \end{gathered} \tag{A1}$$

where $\rho_c$ denotes the sum of all the model core charges $\rho_{cj}$ over the atoms $j$, and $\rho_e$ denotes the electronic charge. The components of the gradient have two contributions,

$$\sigma_i=\sigma_{ci}+\sigma_{ei}, \tag{A2}$$

where $\sigma_{ei}$ are given by Eq. (14), and $\sigma_{ci}$ can be calculated in a straightforward manner from radial derivatives of the spherically symmetric $\rho_{cj}$. $\rho_c$ and $\sigma_{ci}$ must be evaluated on each point of the real-space mesh, which is best done in conjunction with the calculation of pseudopotential components and their derivatives on the mesh. Second derivatives $\partial^2\rho_c/\partial x_i\partial x_j$ are also needed.

The $\boldsymbol{\xi}$-space integral for $E_{\mathrm{xc}}$ is now given by

$$E_{\mathrm{xc}}=\int d^3\xi\; g^{1/2}\rho\varepsilon(\rho,\sigma). \tag{A3}$$

The potentials $\tilde{v}$ and $\tilde{V}_j$ needed to compute the $H_{\mathrm{xc}}$ matrix elements are given by the same expressions as in Sec. II, simply substituting the total $\rho$ and $\sigma$. The variation of $E_{\mathrm{xc}}$ with the coordinate transformation, unfortunately, cannot be found simply by substituting the total quantities in Eq. (22)–(24). The reason is that the two components of $\rho$ in Eq. (A1) depend on the coordinate transformation in different ways. We recall that $\tilde{\rho}$ has no explicit $\mathbf{x}$ dependence (only $\boldsymbol{\xi}$ dependence), and that $g$ depends only on $\partial x_i/\partial\xi_j$. $\rho_c$ depends explicitly on $\mathbf{x}$, and not on $\boldsymbol{\xi}$ derivatives. Corresponding differences are true of the electron and core contributions to $\sigma$.

The presence of $\rho_c$ introduces an additional component of $\delta E_{\mathrm{xc}}$ which depends on $\delta x_p$,

$$\begin{aligned} \delta E_{\mathrm{xc}0}=\sum_{p=1}^{3}\int d^3\xi\; g^{1/2}\bigg[&(\varepsilon+\rho\varepsilon_\rho)\sigma_{cp} \\ &+\rho\sigma^{-1}\varepsilon_\sigma\sum_{i=1}^{3}\sigma_i\,\frac{\partial^2\rho_c}{\partial x_p\partial x_i}\bigg]\delta x_p. \end{aligned} \tag{A4}$$

Corresponding to Eq. (22) we have the modified expression

$$\delta E_{\mathrm{xc}1}=\sum_{ip=1}^{3}\int d^3\xi(\rho_c g^{1/2}\varepsilon-\tilde{\rho}\rho\varepsilon_\rho)\,\frac{\partial\xi_i}{\partial x_p}\,\frac{\partial\,\delta x_p}{\partial\xi_i}. \tag{A5}$$

Equation (23) retains its form,

$$\begin{aligned} \delta E_{\mathrm{xc}2}=-\sum_{ijp=1}^{3}\int d^3\xi\; g^{1/2}\rho\varepsilon_\sigma\sigma^{-1}\sigma_j\bigg[&\sigma_{ep}\,\frac{\partial\xi_i}{\partial x_j}+\sigma_{ej}\,\frac{\partial\xi_i}{\partial x_p} \\ &-\rho_e\sum_{kmn=1}^{3}\frac{\partial\xi_k}{\partial x_j}\,\frac{\partial\xi_m}{\partial x_p}\,\frac{\partial^2x_n}{\partial\xi_k\partial\xi_m}\,\frac{\partial\xi_i}{\partial x_n}\bigg]\,\frac{\partial\,\delta x_p}{\partial\xi_i}, \end{aligned} \tag{A6}$$

but the absence or presence of $e$ subscripts in various terms must be noted carefully. Equation (24) for $\delta E_{\mathrm{xc}3}$ survives intact.

Suppose the $\rho_e$ and $\sigma_e$ were set to zero in Eq. (A1). The resulting $E_{\mathrm{cxc}}$ is then the energy of a charge density which is fixed in $\mathbf{x}$ space, and must be invariant under any coordinate transformation. The $\delta E_{\mathrm{xc}2}$ and $\delta E_{\mathrm{xc}3}$ terms are explicitly zero in this limit. The sum of $\delta E_{\mathrm{xc}0}$ and $\delta E_{\mathrm{xc}1}$ can be shown to vanish using various identities and performing an integration by parts. As is often the case, the numerical convergence of mathematically equivalent integrals can be inequivalent, and the forms given here appear to have better convergence properties. The model core charges can be another source of aliasing errors since they may contain higher spatial frequencies than the pseudopotential or electron charge density. We have found that the explicit subtraction of the ''core only'' quantity $E_{\mathrm{cxc}}$ and its (nominally zero) contribution to the forces on the transformation parameters minimizes these errors.

Another effect which must be considered in the presence of the model core charges is their contribution to the forces on the atoms. Such forces are present in the LDA as well as the GGA, although the author is unaware of any discussion of this effect in the literature. It is straightforward to show that in the LDA case, the force is given by the integral of the gradient of $v_{\mathrm{xc}}$ times the core charge of the atom. In the GGA case, we find

$$\begin{aligned} \frac{\partial E_{\mathrm{xc}}}{\partial R_{ki}}=-\int d^3\xi\; g^{1/2}\bigg[&(\varepsilon+\rho\varepsilon_\rho)\,\frac{\partial\rho_{ck}}{\partial x_i} \\ &+\rho\sigma^{-1}\varepsilon_\sigma\sum_{j=1}^{3}\sigma j\,\frac{\partial^2\rho_{kc}}{\partial x_i\partial x_j}\bigg], \end{aligned} \tag{A7}$$

where $\mathbf{R}_k$ and $\rho_{ck}$ are the coordinate and core charge of atom $k$. This expression depends on the coordinate transformation only through the volume element $g^{1/2}d^3\xi=d^3x$, and can be used directly in conventional plane-wave calculations. In the LDA limit, the second term vanishes, and the first term can be transformed to the form stated above by integration by parts.

Because of the presence of second derivatives of $\rho_c$ in Eqs. (A4) and (A7), it is important to ensure that these are continuous. In Ref. 18, the actual core charge is used outside a matching radius, where the value and first derivative are fitted with an inner model function. We use a three-term polynomial of even powers of the radius, $\rho_{ck}(r)=a_k+b_kr^2++c_kr^4$, and fit the second derivative as well.

[1] F. Gygi, Europhys. Lett. **19**, 617 (1992).

[2] F. Gygi, Phys. Rev. B **48**, 11 692 (1993).

[3] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964); W. Kohn and L. J. Sham, *ibid*. **140**, A1133 (1965).

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

[5] D. R. Hamann, Phys. Rev. B **51**, 7337 (1995).

[6] J. L. Holm, O. J. Kleppa, and E. F. Westrum, Jr., Geochim. Cosmochim. Acta **31**, 2289 (1967).

[7] M. Akaogi and A. Navrotsky, Phys. Earth Planet. Inter. **36**, 124 (1984).

[8] N. R. Keskar and J. R. Chelikowsky, Phys. Rev. B **46**, 1 (1992).

[9] F. Liu, S. H. Garofalini, D. King-Smith, and D. Vanderbilt, Phys. Rev. B **49**, 12 528 (1994).

[10] D. R. Hamann, Phys. Rev. Lett. **76**, 660 (1996).

[11] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jaceson, M. R. Pederson, D. J. Singh, and C. Fiolhais, Phys. Rev. B **46**, 6671 (1992); J. Perdew, Phys. Rev. Lett. **55**, 1665 (1985); J. P. Perdew and Y. Wang, Phys. Rev. B **33**, 8800 (1986); J. P. Perdew, in *Electronic Structure of Solids '91*, edited by P. Ziesche and H. Eschrig (Akademie Verlag, Berlin, 1991).

[12] D. C. Langreth and M. J. Mehl, Phys. Rev. B **28**, 1809 (1983).

[13] A. D. Becke, Phys. Rev. A **38**, 3098 (1988).

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

[15] J. A. White and D. M. Bird, Phys. Rev. B **50**, 4954 (1994).

[16] D. R. Hamann, Phys. Rev. B **51**, 9508 (1995).

[17] The signs of the $iA$ terms given here are correct. Those in Eq. (7) of Ref. 2 are misprints. [F. Gygi (private communication).]

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

[19] F. Gygi, Phys. Rev. B **51**, 11 190 (1995).

[20] F. Gygi and G. Galli, Phys. Rev. B **52**, R2229 (1995).

[21] E. Kaxiras (unpublished).

[22] E. B. Stechl (private communication).

[23] P. Pulay, Mol. Phys. **17**, 197 (1969); C. Satoko, Phys. Rev. B **30**, 1754 (1984).