Optimally smooth norm-conserving pseudopotentials

David Vanderbilt
Lyman Laboratory of Physics, Harvard University, Cambridge, Massachusetts 02138
(Received 15 July 1985; revised manuscript received 13 September 1985)

Modern norm-conserving pseudopotentials are constructed to satisfy a set of criteria for the matching of pseudo- and all-electron eigenvalues and wave functions. In practice, it is also desirable that they be as smooth as possible, so that their reciprocal-space representation decays as quickly as possible. To this end, a simple modification of a standard pseudopotential generation scheme is developed. The new, smoother potentials are shown to decay significantly faster in reciprocal space, with no loss of transferability.

The availability of first-principles norm-conserving pseudopotentials[1–6] has greatly contributed to the practicality of realistic *ab initio* total-energy calculations within the local-density theory.[7] These pseudopotentials are generated in such a way that for a reference configuration of the free atom, the pseudoeigenvalues match the all-electron eigenvalues for the valence states, and the corresponding wave functions match exactly (not just to within a multiplicative constant) outside some core radius (norm conservation[1]). These criteria ensure the transferability of the pseudopotential to other chemical configurations.

While there are many ways to construct such potentials,[1–6] I concentrate on those which meet an additional criterion, that of being optimally smooth. That is, the pseudopotential should have as few high Fourier components as possible. The widespread use of the momentum-space representation[8] for total-energy calculations provides an immediate motivation for this criterion. In this approach, the potentials and wave functions are always truncated, in practice, outside some reciprocal-space sphere. The q-space cutoffs have to be extended until convergence is obtained, at the expense of rapidly increasing computer time. While the wave-function cutoff is most crucial (it determines the size of the matrix to be diagonalized), the inclusion of high Fourier components in the wave functions is expected to be less critical if the potential is smoother. Even if the plane-wave approach is not used, smooth potentials may be desirable for a variety of reasons. For example, in localized-orbital approaches it is common to fit the potential to a sum of Gaussians with adjustable decay constants.[9] This is much easier to do accurately with few decays if the potentials are smooth.

Many of the previously proposed norm-conserving pseudopotentials[4–6] are of the "hard-core" type having a repulsive r^{-2} divergence at the origin. These do not meet the criterion of smoothness; the corresponding decay in reciprocal space is as q^{-1}, which is much too slow. Among the most frequently used norm-conserving pseudopotentials of the soft-core type are those of Hamann, Schlüter and Chiang[1] (HSC) and Kerker.[2] The latter are generated from a pseudo-wave-function having a discontinuity in its third derivative at the origin and at the cutoff radius; this in turn gives rise to slope discontinuities in the potential at the same locations, and a still-slow q^{-2} decay of the potential in reciprocal space. I prefer instead to use the HSC scheme as a starting point, as the potentials generated in this scheme are free of such kinks.

The HSC method begins with a reference self-consistent full-core atomic calculation, using a local exchange-correlation energy functional scheme. The screened all-electron potential $V(r)$, valence eigenvalues ϵ_l, and valence wave functions $u_l(r)=r\psi_l(r)$ are obtained. For each l, a cutoff radius r_{c_l} is chosen beyond which the screened pseudopotential $V_l^{\rm ps}(r)$ is required to converge to $V(r)$. The generation of the pseudopotential proceeds as follows. First, the full potential $V(r)$ is subjected to a cutoff in the core region:

$$V_l^{(1)}(r)=[1-f_1(r/r_{c_l})]V(r) \quad . \tag{1}$$

Second, the constant c_l is chosen so that the nodeless solution $w_l^{(2)}(r)$ of the potential

$$V_l^{(2)}(r)=V_l^{(1)}(r)+c_l f_2(r/r_{c_l}) \tag{2}$$

has the correct energy ϵ_l. Third, the constants γ_l and δ_l are chosen so that the pseudo-wave-function

$$w_l^{\rm ps}(r)=\gamma_l[w_l^{(2)}(r)+\delta_l r^{l+1}f_3(r/r_{c_l})] \tag{3}$$

is normalized to unity and $w_l^{\rm ps}(r)\rightarrow u_l(r)$ for $r>r_{c_l}$ (norm conservation). Finally, the Schrödinger equation is inverted to obtain $V_l^{\rm ps}(r)$. After this has been done for each l, the pseudo-charge-density is formed from the $w_l^{\rm ps}$ wave functions, and the bare potentials $V_l^{\rm ion,ps}(r)$ are obtained by subtracting the corresponding screening potential from the $V_l^{\rm ps}(r)$.

In the HSC method, the functions f_1, f_2, and f_3 are chosen to be identical and of the form

$$f(x)=g_a(x)\equiv\exp(-x^a) \tag{4}$$

and $a=4$ [Ref. 1(a)] or 3.5 [Ref. 1(b)]. As will be seen, these three functions need not be identical; in fact, each is constrained in a different way. By optimizing each step of the process (1)–(3) separately, using a different value of a or a different functional form entirely, the potentials can be tailored to have the desired smoothness properties.

Even within the standard HSC procedure, the smoothness can be improved somewhat by the proper choice of r_{c_l} and a. First, a larger r_{c_l} gives a smoother potential, although at the expense of reduced transferability if r_{c_l} becomes too large. Some gains can be made in this way; the pseudopotentials of Ref. 1(b), for example, appear to have an unnecessarily small r_{c_s} for certain portions of the Periodic Table. Second, the choice $a = 3.5$ [Ref. 1(b)] appears to give the smoothest potentials, for reasons which will emerge shortly.

Consider now each of the three steps of the HSC procedure separately, starting with Eq. (1). The leading term in the $r\to 0$ expansion of the cutoff function is $1-f_1(r/r_{c_l})\approx(r/r_{c_l})^a$ in the HSC scheme. Since $V(r)$ diverges as r^{-1} at small r, $a>2$ is needed to avoid a slope discontinuity in $V_l^{(1)}$ at $r=0$; $a=3$ is a practical minimum. Even $a=3$ is unacceptably small for all but the lightest atoms, however, because of the core shell structure. For example, for $a=3$, $V_l^{(1)}(r)$ has an unnecessarily large second derivative $-4Z$ at $r=0$ (Z is the full atomic number). The effects of this shell structure can be clearly seen in the Fourier transform $V_l(q)$ for heavy atoms at $a=3$, but are greatly reduced at $a=3.5$.

An alternative cutoff procedure is simply to replace $V(r)$ by a polynomial inside a cutoff radius $\tilde{r}_{c_l}$. I use

$$V_l^{(1)}(r)=b_0+b_2r^2+b_4r^4 \tag{5}$$

(this is analytic at the origin) and choose the parameters to match the value and first two derivatives at $\tilde{r}_{c_l}$. A scaling $\tilde{r}_{c_l}=1.5r_{c_l}$ has been used so that for a given r_{c_l} the potentials generated using Eqs. (1) and (4) or Eq. (5) look similar in their convergence beyond r_{c_l} and have similar transferability. The choice (5) completely removes the core shell structure from $V_l^{(1)}$, at the expense of introducing a third-derivative discontinuity in $V_l^{(1)}(r)$ and consequently a q^{-4} tail in $V_l^{(1)}(q)$. In a series of tests, this tail has been found to be negligible, and Eq. (5) is found to provide a slight improvement over Eqs. (1) and (4) with $a=3.5$. Equation (5) has thus been adopted here.

In the second step, Eq. (2), the choice of the function f_2 has no particular constraints. It should approach unity at least quadratically at $r=0$ (a finite slope at $r=0$ is to be avoided); $f_2(x)=g_{2.0}(x)$ is thus acceptable. It should also decay to zero beyond r_{c_l} "fast enough." To systematize this latter consideration, the somewhat arbitrary constraint

$$f(1.5)=10^{-2} \tag{6}$$

is imposed upon all functions to be tested. This ensures that they will all have roughly the same transferability, insofar as this depends on the rapidity of the cutoff. To satisfy Eq. (6), Eq. (4) is replaced by

$$g_a(x)\equiv 100^{-(x/1.5)^a} \tag{4'}$$

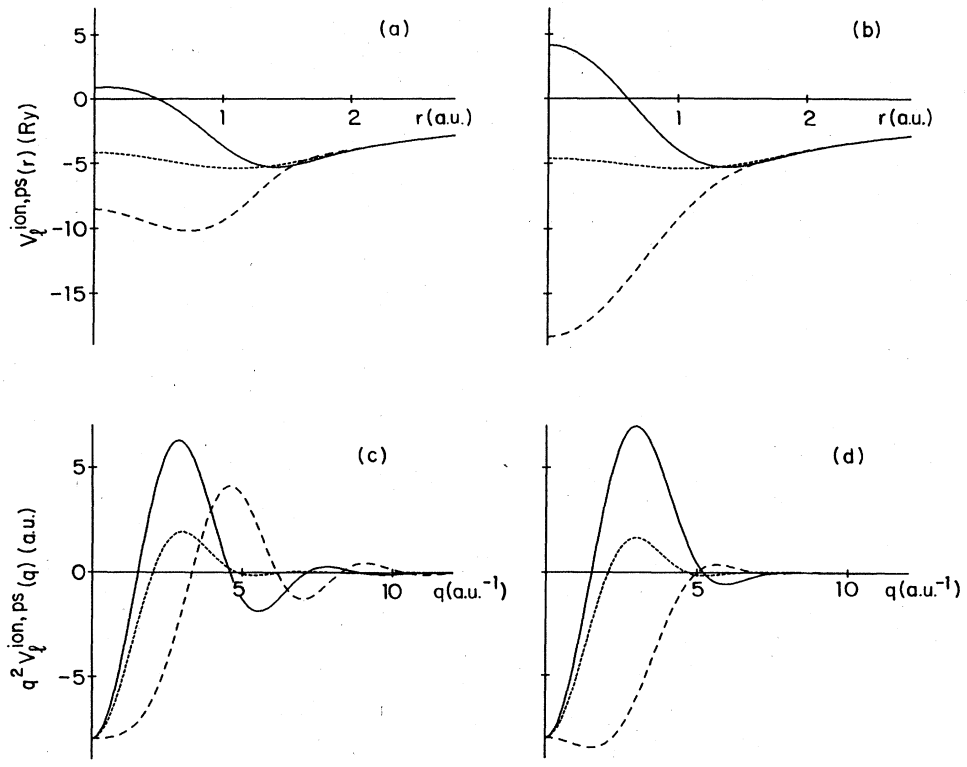

FIG. 1. Silicon ionic pseudopotentials in real space for (a) HSC method and (b) present method; Fourier transforms are given for (c) HSC method and (d) present method. Solid curve, s potential; dotted curve, p potential; dashed curve, d potential.

henceforth. This is really just an a-dependent redefinition of the cutoff radius; Eqs. (4) and (4′) are almost identical for $3.5 \lesssim a \lesssim 4$.

If $f_2(x)=g_a(x)$ with $a=3.5$ or 4 as in the standard HSC scheme, the function f_2 decays rapidly beyond r_{c_l}, but has an unnecessary plateaulike behavior $g_a(x)\approx 1$ for $x \lesssim 0.5$, and consequently the transition from $g=1$ to $g=0$ is rather sharp in the vicinity of the cutoff radius. This introduces high Fourier components. If instead $f_2(x)=g_a(x)$ with $a=2$ is used (a Gaussian), the plateau is avoided, but the decay is rather slow. Equations (6) and (4′) force the decay to be rapid at the expense of reducing the effective cutoff radius and making the function unnecessarily narrowly peaked near the origin. Again this introduces high Fourier components. A compromise functional form can be used which avoids both of these problems. Consider the family of functions

$$h_b(x)\equiv \exp[-\sinh^2(bx)/\sinh^2(b)] \tag{7}$$

again parametrized by a constant b. A Taylor-series expansion of the exponent shows that for small x this function follows a Gaussian (no plateau), but for $x \gtrsim 1$ the exponent picks up quartic and higher even powers of x, and the function therefore decays faster than a Gaussian in real space. To satisfy Eq. (6), $h_b(x)$ is redefined as

$$h_b(x)\equiv 100^{-\sinh^2(bx/1.5)/\sinh^2(b)} \ . \tag{7′}$$

At $b=0$, $h_b(x)$ is just a Gaussian; the larger is b, the sharper is the cutoff. For very large b, the function approaches a step function, again introducing high Fourier components. The parameter value $b=1.0$ has been found to be a good compromise; while $g_{3.5}$ and $h_{1.0}$ have very similar real-space decays beyond the cutoff, the Fourier transform of $h_{1.0}$ drops off much faster. The choice $f_2(x)=h_{1.0}(x)$ has been adopted here.

In the third step, the pseudo-wave-function is augmented as

$$w_l^{\rm ps}(r)=\gamma_l[w_l^{(2)}(r)+\delta_l y_l(r)] \ , \tag{8}$$

where

$$y_l(r)=r^{l+1}f_3(r/r_{c_l}) \ . \tag{9}$$

In the HSC scheme, $f_3=g_{3.5}$ or $g_{4.0}$. The change in the potential from the inversion of the Schrödinger equation is

$$V_l^{\rm ps}=V_l^{(2)}+\frac{\delta_l\gamma_l y_l}{w_l^{\rm ps}}\left[\epsilon_l-V_l^{(2)}-\frac{l(l+1)}{r^2}+\frac{y_l''}{y_l}\right] . \tag{10}$$

The second derivative in the last term increases the sensitivity of the potential to kinks in f_3. For example, a third-derivative discontinuity in f_3 gives rise to a first-derivative discontinuity in $V_l^{\rm ps}$. Thus the choice $f_3=g_a$ with $a=3$ introduces a finite slope in $V_l^{\rm ps}$ as $r\to 0$, which is undesirable. A Gaussian $g_{2.0}$ is free of these problems, but has the slow-decay problem discussed for step (2). The choices $g_{3.5}$ and $g_{4.0}$ are also possible, but again the plateau behavior is unnecessary. Once more the choice $f_3(x)=h_{1.0}(x)$ is found to be a good compromise.

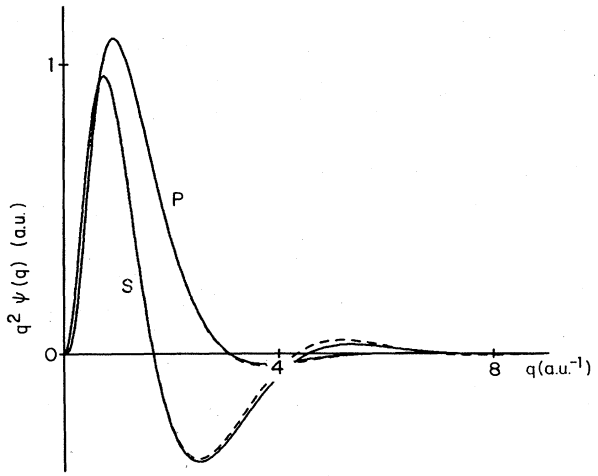

FIG. 2. Fourier transform of s and p pseudo-wave-functions for Si. Dashed line, HSC method; solid line, present method.

Actually, the use of Eqs. (8) and (9) is sometimes problematic for cases in which the pseudopotential is repulsive in the core region. For example, the s potential for Si is repulsive, so that the wave function $\psi_s^{(2)}(r)=w_s^{(2)}(r)/r$ is depressed near the origin. If it happens that the norm-conservation condition requires $\delta_l<0$ in Eq. (8), the subtraction of the piece $\delta_l y_l(r)$ may shift $\psi_s^{\rm ps}(r)=\gamma_l[\psi_s^{(2)}(r)+\delta_l f_3(r/r_{c_l})]$ so that $\psi_s^{\rm ps}(0)\ll\psi_s^{(2)}(0)$. Upon inverting the Schrödinger equation, the potential $V_s^{\rm ps}$ acquires a large repulsive peak at the origin, which introduces high Fourier components. A convenient way to avoid this problem is to *multiply* ψ by $1+\delta_l f_3$, instead of adding $\delta_l f_3$ to ψ. That is, take

$$y_l(r)=w_l^{(2)}(r)f_3(r/r_{c_l}) \ . \tag{9′}$$

The use of Eq. (9′) has been adopted here, with $f_3=h_{1.0}$ as in step (2). The Schrödinger equation is inverted as before.[10]

Here, the new pseudopotential generation scheme is compared with the standard HSC scheme ($a=3.5$) for the

TABLE I. Values ($w_{\max}$) and positions ($r_{\max}$) of the radial wave-function maxima for two sets of pseudopotentials, compared with the corresponding all-electron results, for Si. Atomic units are used.

	$r_{\max}(s)$	$w_{\max}(s)$	$r_{\max}(p)$	$w_{\max}(p)$	$r_{\max}(d)$	$w_{\max}(d)$
All electron	1.730	0.7670	1.995	0.6815	3.199	0.4872
HSC[a]	1.740	0.7656	1.997	0.6813	3.199	0.4872
Present	1.741	0.7650	2.000	0.6809	3.199	0.4872

[a]Reference 1.

TABLE II. Excitation energies ΔE for excited configurations of the Si atom. For each case, the energies are referenced to the total energy of the atom in its ground state, s^2p^2. All energies in Ry.

	ΔE (s_1p_3)	ΔE (s_2p_1)	ΔE $(s_2p_{0.5}d_{0.5})$	ΔE (s_1p_1)
All electron	0.4927	0.5789	0.8764	2.3460
HSC[a]	0.4932	0.5799	0.8777	2.3458
Present	0.4933	0.5799	0.8777	2.3457

[a]Reference 1.

case of Si, a typical semiconductor. Similar results have been obtained for other atoms.

Figures 1(a) and 1(b) show the pseudopotentials of the HSC and new schemes, respectively, in real space. In both cases the same configuration $s^2p^{0.5}d^{0.5}$ and cutoffs $r_{c_s}=1.17$, $r_{c_p}=1.35$, and $r_{c_d}=1.17$, were used. The HSC potentials appear "softer" in that the potentials are weaker at the origin. However, "softer" is not necessarily "smoother," as can be seen in Figs. 1(c) and 1(d). The decay in reciprocal space is significantly faster in the new scheme, especially for the s and d components. The "softness" of the HSC potential is gained at the expense of extra oscillations; note, for example, that $V_d(r)$ is nonmonotonic. It is these real-space oscillations which are largely responsible for the higher Fourier components in the HSC potentials.

In Fig. 2 the Fourier transforms of the silicon neutral-atom pseudo-wave-functions are shown for the standard HSC and new methods. A slight improvement is evident. The form of the wave function appears to be strongly constrained by the conditions of matching beyond r_{c_l} and norm conservation, so that significant improvement in the reciprocal-space decay is difficult to achieve.

Several tests have been carried out on the transferability of the new potentials. In Table I, the positions and values of the wave-function maxima are compared with the all-electron results, for both the HSC and new schemes. The shifts of the maxima are not significantly greater in the new method. Table II shows a comparison of the excitation energies to a variety of neutral and ionic configurations, for both schemes. Both give almost identical results, and are in excellent agreement with the all-electron case. Finally, the logarithmic derivatives of the wave functions have also been compared with the corresponding all-electron results, for both schemes, and the energy dependence was found to be in equally good agreement with the all-electron calculation. Collectively, these results show that the gains in smoothness are not accompanied by a loss of transferability.

In summary, I have presented a simple modification of the Hamann-Schlüter-Chiang pseudopotential generation scheme which is optimized to generate smooth potentials. This is accomplished by choosing the functional form used in each step of the pseudopotential generation scheme separately. The new, smoother potentials are shown to decay significantly faster in reciprocal space, with no loss of transferability. It is hoped that these potentials will contribute to the tractability and accuracy of pseudopotential calculations of electronic structure and total energy.

Support for this work was provided by the National Science Foundation, through the Harvard Materials Science Laboratory and through Grant No. DMR82-07431.

[1](a) D. R. Hamann, M. Schlüter, and C. Chiang, Phys. Rev. Lett. **43**, 1494 (1979); (b) G. B. Bachelet, D. R. Hamann, and M. Schlüter, Phys. Rev. B **26**, 4199 (1982).

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

[3]J. J. Joannopoulos, Th. Starkloff, and M. Kastner, Phys. Rev. Lett. **38**, 660 (1977).

[4]A. Redondo, W. A. Goddard III, and T. C. McGill, Phys. Rev. B **15**, 5038 (1977).

[5]A. Zunger, J. Vac. Sci. Technol. **16**, 1337 (1979).

[6]P. A. Christiansen, Y. S. Lee, and K. S. Pitzer, J. Chem. Phys. **71**, 4445 (1979).

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

[8]J. Ihm, A. Zunger, and M. L. Cohen, J. Phys. C **12**, 4409 (1979).

[9]D. Vanderbilt and S. G. Louie, J. Comput. Phys. **56**, 259 (1984).

[10]With Eq. (9′) in place of (9), the derivative y_l'' must be done numerically. A five-point finite-difference approximation is found to work well for the Herman-Skillman mesh used here.