PHYSICAL REVIEW B **71**, 121105(R) (2005)

Optimized nonorthogonal localized orbitals for linear scaling quantum Monte Carlo calculations

Fernando A. Reboredo* and Andrew J. Williamson†
Lawrence Livermore National Laboratory, Livermore, California 94550, USA
(Received 3 January 2005; published 23 March 2005)

We derive an automatic procedure for generating a set of highly localized, nonorthogonal orbitals for linear scaling quantum Monte Carlo (QMC) calculations. We demonstrate the advantage of these orbitals for calculating the total energy of both semiconducting and metallic systems by studying bulk silicon and the homogeneous electron gas. For silicon, the improved localization of these orbitals reduces the computational time by a factor of 5 and the memory by a factor of 6 compared to localized, orthogonal orbitals. For jellium at typical metallic densities, we demonstrate that the total energy is converged to 3 meV per electron for orbitals truncated within spheres with radii $7r_s$, opening the possibility of linear scaling QMC calculations for realistic metallic systems.

DOI: 10.1103/PhysRevB.71.121105 PACS number(s): 71.15.Dx, 71.15.Nc

In recent years, one of the most promising developments in the field of electronic structure calculations has been the development of algorithms whose cost grows as the first power of the system size. These linear scaling algorithms have been developed for several electronic structure techniques, including tight binding,[1] density functional theory[1–6] (DFT), coupled cluster,[7] and quantum Monte Carlo[8,9] (QMC). In all these approaches, extended Bloch orbitals, ψ_{nk}, are transformed into localized Wannier-like orbitals. The speedup provided by the transformation to localized orbitals depends on the extent to which the orbitals can be localized and subsequently truncated. Therefore, improved methods for constructing localized orbitals have attracted intense attention in recent years.[10–12]

In QMC calculations, truncated, localized orbitals can be used to introduce sparsity into the Slater determinant part of the trial wave function. As the calculation of the orbitals used to construct this determinant is the dominant cost of QMC calculations, this transformation yields a near linear scaling QMC algorithm. In our approach to linear scaling QMC calculations,[8] the Slater determinant was constructed from a set of orthonormal Wannier functions. This choice of orbitals produces a near linear scaling algorithm, which has successfully been applied to calculations of the total energies and optical gaps of a variety of semiconductor systems.[13,14] However, this method is only applicable to systems where the Wannier functions decay rapidly (exponentially), i.e., it works well for semiconductors and insulators, but it is not applicable to metallic systems where orthonormal Wannier functions decay polynomially.[11]

In this paper we derive and demonstrate the use of a *nonorthogonal* transformation of the Bloch orbitals that overcomes this limitation. This transformation is based on algorithms developed for linear scaling DFT calculations[1–6] and is designed to minimize a cost function associated with the total number of orbital evaluations required in a linear scaling QMC calculation. For representative semiconductor systems, the orbitals obtained from this nonorthogonal transformation are significantly more localized and smoother than orthogonal Wannier functions, and can typically be truncated in one sixth of the volume of the equivalent orthogonal function without sacrificing accuracy. This produces an algorithm approximately five times faster than previous linear scaling QMC calculations[8] and requiring one sixth the memory. In addition, we demonstrate that while orthogonal Wannier functions for metallic systems cannot be truncated within a practical volume, nonorthogonal orbitals constructed via our procedure can be truncated within a practical cutoff radius.

Our QMC calculations use a linear scaling[8] version of the CASINO (Ref. 15) code with a standard Slater-Jastrow trial wave function, $\Psi_T(\mathbf{R})$.[16] The Slater determinants are constructed from a set of truncated, localized linearly independent orbitals $D_{ij}=\phi_i(\mathbf{r}_j)\Theta_i(\mathbf{r}_j)$, where ϕ are the nonorthogonal orbitals and Θ are the truncation functions. In principle, one can optimize the shape of the truncation functions, however, for the systems studied here, we find that spherical step functions are a simple and stable solution where,

$$\begin{aligned}\Theta^i(\mathbf{r}) &= 1, \quad |\mathbf{r}-\mathbf{R_i}| < R_i^{cut} \\ &= 0, \quad |\mathbf{r}-\mathbf{R_i}| > R_i^{cut}. \end{aligned} \tag{1}$$

The truncation functions Θ^i are defined by two parameters, the cutoff radii R_i^{cut} and the centers $\mathbf{R_i}$. These parameters are optimized iteratively using a procedure designed to minimize the computational cost of the QMC calculation. The nonorthogonal orbitals, ϕ_i, associated with each Θ^i are obtained during the iterative process.

The computational cost of a typical QMC calculation is proportional to the number of orbital evaluations required to construct the Slater determinant for each configuration of electron coordinates, $\mathbf{R}=(\mathbf{r}_1,\mathbf{r}_2\ldots\mathbf{r}_N)$. The cost is therefore the product of the probability, $|\Psi_T(\mathbf{R})|^2$, of sampling a given configuration, $\mathbf{R}$, and the cost of evaluating each of the nonzero elements in the Slater determinant produced by that configuration. For each element, $\phi_i(\mathbf{r}_j)$, if $\mathbf{r}_j$ falls within the truncation function, Θ_i, this adds 1 to the cost, i.e.,

$$\text{Cost} = \int d\mathbf{R}|\Psi_T(\mathbf{R})|^2 \sum_{ij}^{N} \Theta_i(\mathbf{r}_j). \tag{2}$$

By integrating out all but one electron coordinates, Eq. (2) can be expressed in terms of the density $\rho(\mathbf{r})$ as

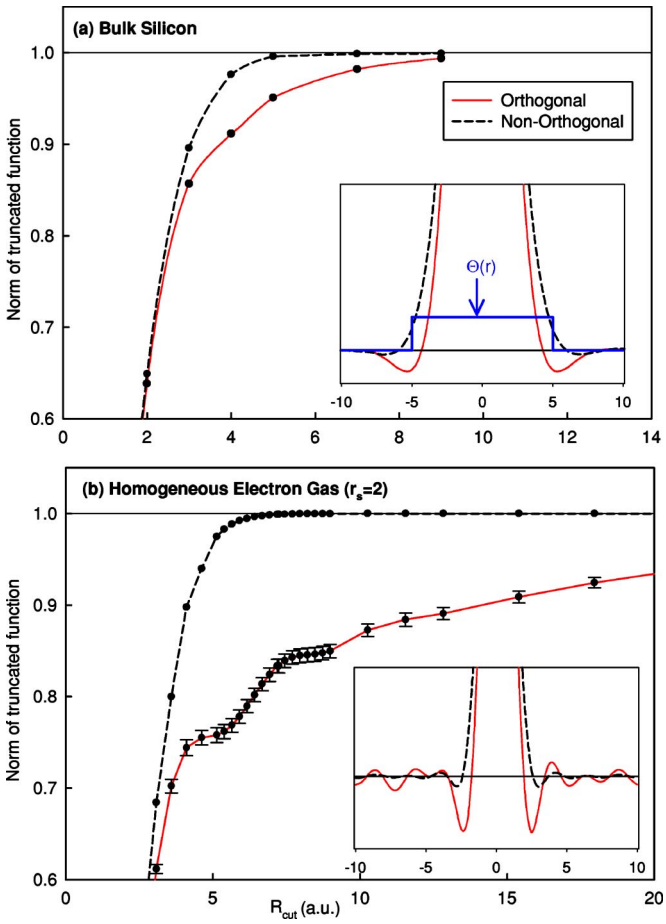

FIG. 1. (Color online) Comparison of the norm of orthogonal (red, solid lines) and nonorthogonal (black, dashed lines) localized orbitals in (a) bulk silicon and (b) a HEG at the same r_s. The error bars show the spread in norm. The insets compare the orthogonal and nonorthogonal localized orbitals. The truncation potential is labeled Θ.

$$\text{Cost} = \sum_i \int d\mathbf{r}\rho(\mathbf{r})\Theta_i(\mathbf{r}). \tag{3}$$

We find a satisfactory minimum of Eq. (3) by starting from an initial choice of Θ_i and iteratively updating first the cutoff radii R_i^{cut} and then the centers $\mathbf{R}_i$.

(i) *Generating optimal nonorthogonal orbitals and cutoff radii.* Each truncation function, Θ_i, can be considered as a potential acting on the Hilbert space of Bloch orbitals. In the inset to Fig. 1(a) this potential is shown with a blue line. If one constructs the matrix elements of the Bloch orbitals with this potential, $\Theta^i_{jk} = \langle \phi_j^{Bloch} | \Theta^i(\mathbf{r}) | \phi_k^{Bloch} \rangle$, then the eigenstate ϕ_i of Θ^i with the largest eigenvalue is the most localized state within the truncation region. This is the orbital with the maximum truncated norm, X, defined as

$$X = \int d\mathbf{r}|\phi_i(\mathbf{r})|^2\Theta_i(\mathbf{r}). \tag{4}$$

Increasing R_i^{cut} increases the above value of X, reducing the resulting truncation error in the QMC calculation, but also increases the computational cost in Eq. (3). Therefore, we adjust the cutoff radius R_i^{cut} to achieve a target norm, e.g., $X=0.999$. Repeating this diagonalization procedure for each truncation function Θ^i generates an associated set of nonorthogonal orbitals $\{\phi\}$. This procedure for generating a set of nonorthogonal orbitals associated with a set of truncation regions is similar to those adopted in linear scaling density functional calculations[1,3,4,6] and recently in a QMC calculation of MgO.[9] Next, we extend this procedure to automatically optimize the centers of the truncation functions for systems where they cannot be guessed *a priori*.

(ii) *Updating the truncation centers.* The cost function in Eq. (3) can be rewritten as

$$\text{Cost} = NX + \sum_i \int d\mathbf{r}\left[\rho(\mathbf{r}) - |\phi_i(\mathbf{r})|^2\right]\Theta_i(\mathbf{r}), \tag{5}$$

where N is the number of orbitals and X is defined in Eq. (4). The first term in Eq. (5), NX, cannot be reduced without losing accuracy. Therefore the only way to reduce the computational cost is to minimize the second term in Eq. (5) by placing the truncation centers where $\rho(\mathbf{r}) - |\phi_i(\mathbf{r})|^2$ is minimum. Since $\rho(\mathbf{r}) \geqslant |\phi_i(\mathbf{r})|^2$, this is minimized in regions where ϕ_i is most localized and therefore closest to ρ. Therefore for the next iteration, we move the truncation centers towards the center of mass of the $|\phi_i(\mathbf{r})|^2$ for the current iteration. To ensure linear independence, we orthogonalize the set $\{\phi\}$ with a polar decomposition before calculating this center of mass.

This updated set of truncation functions, Θ^i, with new centers, $\mathbf{R}_i$, is then used to generate a new set of nonorthogonal orbitals using the procedure in (i) above and the process is repeated. Starting from a random choice of centers, 10–15 iterations are typically required to find a minimum of Eq. (3) and to converge the centers. If one uses a good starting set of centers, such as the centers of maximally localized Wannier functions,[10] the Θ_i converge in one or two iterations.

To analyze the properties of these nonorthogonal orbitals we first compare their localization properties with an equivalent set of orthogonal orbitals. We then examine the convergence of the total energy in quantum Monte Carlo calculations using these orbitals. Comparisons are made for the prototypical semiconductor and metal systems, silicon, and the homogeneous electron gas.

In Fig. 1 the norm contained within a spherical truncation region of the orthogonal and nonorthogonal localized orbitals are compared as a function of R_{cut}. Figure 1(a) compares orbitals constructed for bulk silicon. The input states were obtained from a 64-atom local-density approximation (LDA) calculation,[17] using a norm-conserving Hamann pseudopotential and a 35-Ry cutoff. The nonorthogonal orbitals were constructed using the iterative procedure described above, where the desired norm, X, was varied from 0.5 to 0.999 99 to construct the plot. The orthogonal states were obtained by performing a final polar decomposition orthogonalization step. Due to symmetry all the nonorthogonal orbitals are equivalent. Within a given radius, the nonorthogonal orbitals contain significantly more charge; e.g., 99.9% of the norm is contained within a sphere of radius 5.5 a.u. compared to

orthogonal orbitals which require an 11-a.u. sphere to capture the same charge. The origin of this dramatically improved localization is shown in the inset to Fig. 1(a), which shows a line plot through the center of the $R_{cut}=5$ orbitals. While the nonorthogonal states converge smoothly to zero with minimal oscillations, the orthogonal orbitals oscillate around zero for >5 a.u. after initially crossing zero to maintain orthogonality between states. While the amplitude of these oscillations is small compared to the central peak, the r^2 prefactor leaves a significant amount of charge in these oscillations.

Comparing the orthogonal orbitals shown in Fig. 1(a) with maximally localized (MLW) orbitals constructed according to Ref. 10, which essentially finds the localized eigenstates of the $e^{i2\pi r/L}$ operator, we find the centers of our nonorthogonal and orthogonal states are identical to the MLW function centers due to symmetry. Additionally, the shape and norm convergence of our orthogonal states is almost identical to the MLW functions. It therefore appears that the shape of *orthogonal* localized orbitals is relatively insensitive to the choice of operator used to localize the states.

Figure 1(b) shows orthogonal and nonorthogonal orbitals constructed for the homogeneous electron gas (HEG) with $r_s=2$ (same as silicon). The input states were the lowest 1935 plane waves in a 50-a.u. cubic box. The norm of the orthogonal orbitals slowly approaches 1.0 as the radius is increased, as would be expected given the slow polynomial decay of orthogonal orbitals in metallic systems.[11] In contrast, the nonorthogonal orbitals rapidly approach 1. For example, a nonorthogonal orbital optimized for $R_{cut}=7$ a.u. contains 99.9% of the norm within R_{cut}, while even the largest sphere inscribed within the supercell (25 a.u. radius) contains only 94% of the norm of the orthogonal orbitals. Note, the nonorthogonal orbitals are still less localized than those in silicon, where 99.9% of the norm is contained within a sphere of radius 5.5 a.u. compared to the 7 a.u. required for jellium. As in silicon, the inset plot shows pronounced, long-range oscillations in the orthogonal orbitals and a much smoother convergence of the nonorthogonal orbitals with minimal oscillation.

Figure 2 compares the truncated norm of orthogonal and nonorthogonal localized orbitals for bulk silicon and the HEG, where the truncated norm is defined as 1 minus the norm contained within a sphere of radius R. As expected from Fig. 1, Fig. 2 shows that the optimal nonorthogonal orbitals (black dashed lines) converge to 1 more rapidly than the equivalent orthogonal orbitals (red solid lines). Figure 2 also illustrates that to obtain maximum localization within a given volume, the nonorthogonal orbitals must be reoptimized for each value of R_{cut}. The blue dotted line in Fig. 2(b) shows the truncated norm of a nonorthogonal orbital optimized to be maximally localized in a Θ function with $R_{cut}=10$. For $R_{cut}=10$ the optimized and fixed nonorthogonal orbitals are identical. For all other values of R_{cut}, the orbital optimized with $R_{cut}=10$ (blue, dotted) is no longer the optimal orbital for that choice of Θ function. For $R>10$, this suboptimal orbital decays polynomially, similar to orthogonal orbitals.

Our previous work with orthogonal, truncated, localized

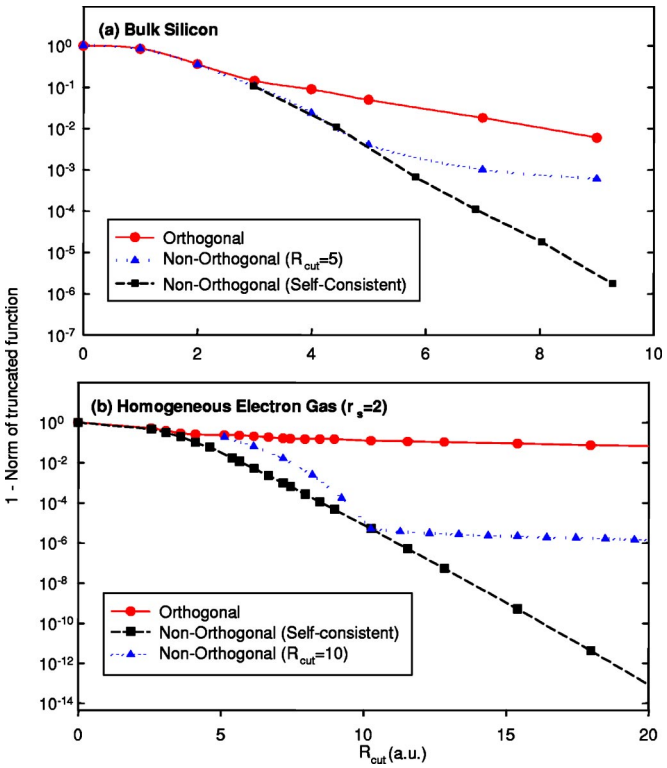

FIG. 2. (Color online) Comparison of the truncated norm of orthogonal (red, solid lines, circles), self-consistent, nonorthogonal (black, dashed lines, squares), and fixed radius, nonorthogonal (blue, dotted lines, triangles) localized orbitals in (a) bulk silicon and (b) HEG ($r_s=2$).

orbitals[8] indicated that the norm of localized orbitals was a good predictor of the truncation error in a QMC calculation of the total energy. For silicon we found that a truncation region large enough to capture 99.9% of the norm was sufficient to produce a converged total energy. On this basis, the improved localization properties of optimal nonorthogonal orbitals shown in Figs. 1 and 2 suggest that these orbitals can be used to perform QMC calculations with smaller truncation radii than those used for previous orthogonal orbitals, without sacrificing accuracy. However, the truncated norm of the localized orbitals in a given representation does not predict the truncation error in the density matrix and how electron-electron correlation will be affected. This is particularly relevant for metals where the nonorthogonal orbitals can be reoptimized for each radius, producing an effective exponential decay, but the density matrix is known to decay polynomially. To fully evaluate the properties of these nonorthogonal orbitals we have performed diffusion Monte Carlo (DMC) total energy calculations of bulk silicon and the HEG to evaluate the convergence of the total energy with the truncation radii of the localized orbitals.

Figure 3(a) compares the convergence of the DMC total energy of the same 64-atom, bulk silicon system shown in Figs. 1(a) and 2(a), using orthogonal and nonorthogonal input orbitals. It shows that the DMC energy converges more rapidly using nonorthogonal orbitals. To converge the total energy to within 0.01 eV per atom using orthogonal orbitals required $R_{cut}=11$ a.u. (Ref. 8) while using nonorthogonal or-

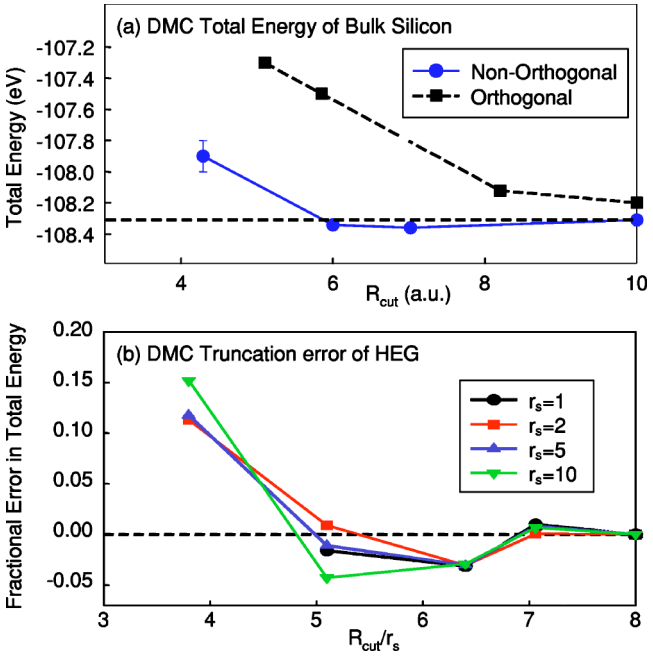

FIG. 3. (Color online) (a) Convergence of DMC total energy of bulk silicon with truncation radius for orthogonal and nonorthogonal orbitals. (b) Convergence of DMC total energy of the HEG as a function of R_{cut}/r_s.

bitals, equivalent accuracy can be obtained with $R_{cut}=6$ a.u. This results in a factor of 5 increase in speed and a factor of 6 reduction in memory.

Figure 3(b) compares the convergence with R_{cut} of the DMC total energy of a homogeneous electron gas with $r_s=1$, 2, 5, and 10. In the HEG, the nonorthogonal orbitals for all r_s values can be obtained by scaling the $r_s=1$ orbital. The kinetic energy scales as r_s^{-2}. To enable us to plot all values of r_s on the same plot, we rescale both axes and plot the fractional DMC truncation error, defined as $\mathrm{Error}(R_{cut})=[E(R_{cut})-E_\infty]/E_\infty$ as a function of R_{cut}/r_s. After this rescaling the convergence plots for each value of r_s fall on a similar curve. Note, the negative truncation error around $R_{cut}/r_s=6$ resulting from a loss of kinetic energy, due to abrupt truncation of the orbitals. This curve shows that the total DMC energy is approximately converged for truncation radii of $7–8r_s$. These converged values are in excellent agreement with the original values from Ceperley and Alder.[18] Therefore, while the slower polynomial decay of the density matrix of metallic systems requires a larger truncation radius to converge the total energy than for semiconductors with equivalent density, the above procedure for generating nonorthogonal orbitals opens the possibility that localized orbitals for metallic systems could be truncated in a practical volume for linear scaling calculations. In addition, the above procedure for generating these nonorthogonal orbitals does not require the high symmetry of the HEG and therefore this approach could be equally applied to linear scaling DMC calculations of realistic metallic systems. For example, in a calculation for $r_s=2$, Fig. 3(b) predicts that the nonorthogonal orbitals can be truncated within $R_{cut}=7$ Å and incur a truncation error of 2 meV per electron, similar to the fixed node error. We also note that the HEG is likely to be among the most difficult metallic systems to obtain a localized-basis set for because the noninteracting wave functions and density have no internal structure to assist localization.

In conclusion, we derive a simple, automatic preprocessing procedure for generating nonorthogonal localized orbitals, with optimized shapes and centers, which minimize the total computational cost of linear scaling QMC calculations. We demonstrate the application of these orbitals in DMC calculations of silicon and the HEG. For silicon the increased localization reduces the computational time by a factor of 5 and the memory by a factor of 6 compared to orthogonal localized orbitals. In the HEG, we demonstrate that practical truncation of orbitals for linear scaling metallic calculations is possible.

The authors would like acknowledge G. Galli, R. Needs, R. Hood, and D. Prendergast for helpful discussions and comments. This work was performed under the auspices of the U.S. Department of Energy by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48.

*Electronic address: reboredo1@llnl.gov
†Electronic address: williamson10@llnl.gov

[1] G. Galli, Phys. Status Solidi B **217**, 231 (2000).
[2] X. Li, R. Nunes, and D. Vanderbilt, Phys. Rev. B **47**, 10 891 (1993).
[3] F. Mauri and G. Galli, Phys. Rev. B **50**, 4316 (1994).
[4] J. Kim, F. Mauri, and G. Galli, Phys. Rev. B **52**, 1640 (1995).
[5] E. Hernandez and M. Gillan, Phys. Rev. B **51**, 10 157 (1995).
[6] S. Liu, J. Perez-Jorda, and W. Yang, J. Chem. Phys. **112**, 1634 (2000).
[7] M. Schutz, J. Chem. Phys. **114**, 661 (2001).
[8] A. Williamson, R. Hood, and J. Grossman, Phys. Rev. Lett. **87**, 246406 (2001).
[9] D. Alfe and M. Gillan, cond-mat/0404578 (unpublished).
[10] N. Marzari and D. Vanderbilt, Phys. Rev. B **56**, 12 847 (1997).
[11] L. He and D. Vanderbilt, Phys. Rev. Lett. **86**, 5341 (2001).
[12] I. Souza, N. Marzari, and D. Vanderbilt, Phys. Rev. B **65**, 035109 (2001).
[13] A. Williamson, J. Grossman, R. Hood, A. Puzder, and G. Galli, Phys. Rev. Lett. **89**, 196803 (2002).
[14] A. Puzder, A. Williamson, F. Reboredo, and G. Galli, Phys. Rev. Lett. **91**, 157405 (2003).
[15] R. Needs, G. Rajagopal, M. Towler, P. Kent, and A. Williamson, *CASINO Code, Version 1.6 User's Manual* (University of Cambridge, Cambridge, 2003).
[16] W. Foulkes, L. Mitas, R. Needs, and G. Rajagopal, Rev. Mod. Phys. **73**, 33 (2001).
[17] F. Gygi, Computer code GP (LLNL, Lawrence, CA, 2004).
[18] D. Ceperley and B. Alder, Phys. Rev. Lett. **45**, 566 (1980).