Correlated sampling in quantum Monte Carlo: A route to forces

Claudia Filippi

Department of Physics, National University of Ireland, Cork, Ireland

C. J. Umrigar

Cornell Theory Center and Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, New York 14853

(Received 20 April 2000)

In order to find the equilibrium geometries of molecules and solids and to perform *ab initio* molecular dynamics, it is necessary to calculate the forces on the nuclei. We present a correlated sampling method to efficiently calculate numerical forces and potential energy surfaces in diffusion Monte Carlo. This method employs a coordinate transformation, earlier used in variational Monte Carlo, to greatly reduce the statistical error. Results are presented for first-row diatomic molecules.

Over the past decade, quantum Monte Carlo (QMC) methods[1–3] have been used to calculate the electronic properties of a variety of atoms, clusters and solids, and have provided the most accurate benchmark calculations of structural energies of systems with large numbers of electrons. However, a major difficulty of QMC methods has been the determination of equilibrium geometries and potential energy surfaces. Hence, most QMC calculations have been performed on geometries obtained with either density functional theory (DFT) or conventional quantum chemistry methods. The computation of forces on nuclei has been a stumbling block that has limited a more widespread use of QMC methods.

DFT methods or standard quantum chemistry techniques use the Hellman-Feynman theorem to compute the forces on nuclei.[4] Unfortunately, this is not practical within QMC for three reasons. First, the wave functions used in QMC are usually not obtained by minimizing the energy. Therefore, if the Hellman-Feynman theorem were employed in variational Monte Carlo (VMC), the forces would have a systematic error. Second, in fixed-node diffusion Monte Carlo (DMC), the Hellman-Feynman force has an error due to the discontinuity in the derivative of the fixed-node wave function at nodes.[5] Finally, in both VMC and DMC, the statistical errors would be too large, since the fluctuations of the potential energy are much larger than those of the total energy.

Alternatively, one could simply compute energy differences to obtain either forces (for an infinitesimal displacement of the ions) or the full potential energy surface of the system. However, while quantum chemistry methods can rely on having an approximately constant and smoothly varying error in the energy, a major disadvantage of QMC methods is that, in addition to systematic errors, one has statistical errors which make the determination of energy differences or smooth potential energy surfaces very computationally expensive. Even though it is not possible to entirely eliminate the statistical errors, it is possible, by using correlated sampling,[6] to make the statistical errors in the relative energies of different geometries much smaller than the errors in the separate energies and to make them vanish in the limit that the two geometries become identical. In the past, the correlated sampling technique has been used within VMC,[7,8]

but there have been very few attempts[9] to extend the approach to DMC, and these were approximate and/or inefficient and were tested only on H_2, H_3^+, and LiH.

In this paper, we present a DMC correlated sampling technique to efficiently compute accurate forces and potential energy surfaces. The DMC bond lengths of first-row diatomic molecules computed with this algorithm are found to be in better agreement with experimental values than are the VMC, Hartree-Fock (HF), local density approximation (LDA), and generalized gradient approximation (GGA) values.

Correlated sampling in variational Monte Carlo. Correlated sampling enables us to compute, from a single reference Monte Carlo walk, the relative energies of different geometries, a reference and one or more secondary geometries, with nuclear coordinates $\mathbf{R}_\alpha$ and $\mathbf{R}_\alpha^s$, Hamiltonians $\mathcal{H}$ and $\mathcal{H}_s$, and wave functions ψ and ψ_s, respectively. Unbiased expectation values are obtained by reweighting the configurations sampled from ψ^2,

$$
\begin{aligned}
E_s - E &= \frac{\langle \psi_s | \mathcal{H}_s | \psi_s \rangle}{\langle \psi_s | \psi_s \rangle} - \frac{\langle \psi | \mathcal{H} | \psi \rangle}{\langle \psi | \psi \rangle} \\
&= \frac{1}{N_{\text{conf}}} \sum_{i=1}^{N_{\text{conf}}} \left\{ \frac{\mathcal{H}_s \psi_s(\mathbf{R}_i)}{\psi_s(\mathbf{R}_i)} W_i - \frac{\mathcal{H}\psi(\mathbf{R}_i)}{\psi(\mathbf{R}_i)} \right\},
\end{aligned}
\tag{1}
$$

where the weights of the N_{conf} MC configurations are

$$
W_i = \frac{N_{\text{conf}} | \psi_s(\mathbf{R}_i)/\psi(\mathbf{R}_i)|^2}{\sum_{i=1}^{N_{\text{conf}}} |\psi_s(\mathbf{R}_i)/\psi(\mathbf{R}_i)|^2},
\tag{2}
$$

and $\mathbf{R} \equiv (\mathbf{r}_1, \ldots, \mathbf{r}_N)$. The statistical error in $E_s - E$ goes to zero linearly as the secondary geometry approaches the reference geometry, thereby permitting the evaluation of numerical forces with a finite statistical error, that can be reduced by increasing the computer time.

Space-warp coordinate transformation. The electronic coordinates sampled from the reference wave function squared, ψ^2, will not be optimal for computing the energy E_s corresponding to the nuclear coordinates $\mathbf{R}_\alpha^s$, since the electron
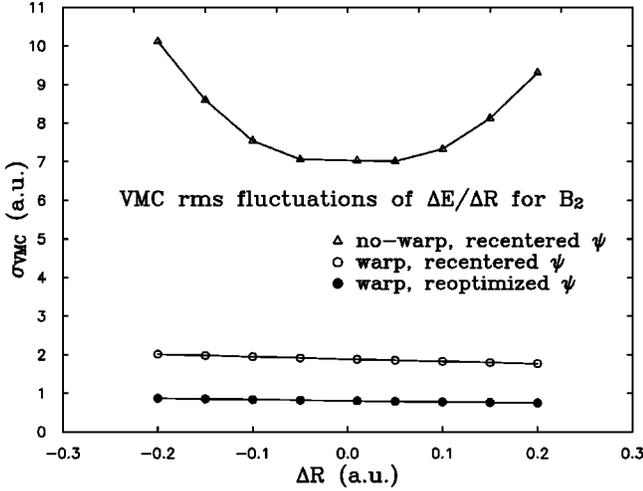
FIG. 1. VMC fluctuations (σ_{VMC}) of the relative energy of the primary and secondary geometries divided by the bond stretch for B_2. If correlated sampling were not used, σ_{VMC} would diverge at $\Delta R=0$. The smallest σ_{VMC} is achieved by using warping along with reoptimized secondary wave functions.

density will be peaked at $\mathbf{R}_\alpha$ rather than at $\mathbf{R}_\alpha^s$. This problem can be solved by mapping the electron coordinates so that the electrons close to a given nucleus move almost rigidly with that nucleus:[8]

$$\mathbf{r}_i^s = \mathbf{r}_i + \sum_{\alpha=1}^{N_{atoms}} (\mathbf{r}_\alpha^s - \mathbf{r}_\alpha) \; \omega_\alpha(\mathbf{r}_i), \qquad (3)$$

where

$$\omega_\alpha(\mathbf{r}_i) = \frac{F(|\mathbf{r}_i - \mathbf{r}_\alpha|)}{\sum_{\beta=1}^{N_{atoms}} F(|\mathbf{r}_i - \mathbf{r}_\beta|)}; \quad \sum_{\alpha=1}^{N_{atoms}} \omega_\alpha(\mathbf{r}_i) = 1. \qquad (4)$$

(We use Latin indices for electronic coordinates and Greek indices for nuclear coordinates.) $F(r)$ is any sufficiently rapidly decaying function, e.g., $r^{-\kappa}$, $e^{-\kappa r}$, or $e^{\kappa/r}$. The reduction in statistical error is large (see Fig. 1) and almost independent of the choice for $F(r)$. In this paper, we use $F(r) = r^{-\kappa}$ and $\kappa=4$.

The equation for $E_s - E$ [Eq. (1)] is now

$$E_s - E = \frac{1}{N_{conf}} \sum_{i=1}^{N_{conf}} \left(\frac{\mathcal{H}_s \psi_s(\mathbf{R}_i^s)}{\psi_s(\mathbf{R}_i^s)} W_i - \frac{\mathcal{H}\psi(\mathbf{R}_i)}{\psi(\mathbf{R}_i)} \right), \qquad (5)$$

where

$$W_i = \frac{N_{conf} |\psi_s(\mathbf{R}_i^s)/\psi(\mathbf{R}_i)|^2 J(\mathbf{R}_i)}{\sum_{j=1}^{N_{conf}} |\psi_s(\mathbf{R}_j^s)/\psi(\mathbf{R}_j)|^2 J(\mathbf{R}_j)}, \qquad (6)$$

and $J(\mathbf{R})$ is the Jacobian for the transformation [Eq. (3)].

Correlated sampling in diffusion Monte Carlo. In DMC,[10] the primary walk is generated according to a stochastic implementation of the integral equation:

$$f(\mathbf{R}', t+\tau) = \int d\mathbf{R} \, G(\mathbf{R}', \mathbf{R}, \tau) f(\mathbf{R}, t), \qquad (7)$$

where the importance-sampled Green's function $G(\mathbf{R}', \mathbf{R}, \tau) = \psi(\mathbf{R}') \langle \mathbf{R}' | \exp\{-\mathcal{H}\tau\} | \mathbf{R} \rangle / \psi(\mathbf{R})$, $f = \phi\psi$, ϕ is the ground state wave function and ψ the trial wave function. For small values of τ (short-time approximation), $G(\mathbf{R}', \mathbf{R}, \tau)$ is given by the product of three factors, drift, diffusion and growth/decay:

$$G(\mathbf{R}', \mathbf{R}, \tau) \approx \frac{1}{(2\pi\tau)^{3N/2}} e^{-[(\mathbf{R}' - \mathbf{R} - \mathbf{V}(\mathbf{R})\tau)^2/2\tau]} e^{S(\mathbf{R}', \mathbf{R}, \tau)}, \qquad (8)$$

where $\mathbf{V} = \nabla \psi(\mathbf{R})/\psi(\mathbf{R})$ and $S(\mathbf{R}', \mathbf{R}, \tau) = (2E_T - E_L(\mathbf{R}') - E_L(\mathbf{R}))\tau/2$ with $E_L = \mathcal{H}\psi(\mathbf{R})/\psi(\mathbf{R})$. A set of primary walkers characterized by the pairs $(\mathbf{R}_i, w_i)$ is a random realization of the distribution f. Each walker executes a branching random walk: a walker originally at $\mathbf{R}$ drifts to $\mathbf{R} + \mathbf{V}(\mathbf{R})\tau$ and then diffuses to $\mathbf{R}'$ according to the Gaussian term in Eq. (8). To ensure that when ψ is the ground state wave function, ψ^2 is sampled exactly despite the short-time approximation in the Green's function, the move is accepted with probability

$$p = \min\left\{ 1, \frac{|\psi(\mathbf{R}')|^2 \; T(\mathbf{R}, \mathbf{R}', \tau)}{|\psi(\mathbf{R})|^2 \; T(\mathbf{R}', \mathbf{R}, \tau)} \right\}, \qquad (9)$$

as prescribed by the detailed balance condition. We denote by T the drift-diffusion part of the Green's function G. Finally, the weight of the walker is multiplied by $\exp[S(\mathbf{R}', \mathbf{R}, \tau)]$. In practice, we employ the improved version of G presented in Ref. 11.

Given a primary walk generated according to Eq. 8, the secondary walk is specified by the space-warp transformation [Eq. (3)]. Two complications, absent in VMC, arise for correlated sampling in DMC. First of all, the dynamics of the secondary walker should have been governed by an importance sampled Green's function constructed from the secondary wave function ψ_s, $G_s(\mathbf{R}^{s'}, \mathbf{R}^s, \tau)$, and the move should have been accepted with probability p_s obtained by substituting in Eq. (9) ψ and T with ψ_s and T_s, respectively. However, the secondary-geometry move was effectively proposed according to the drift-diffusion Green's function $T(\mathbf{R}', \mathbf{R}, \tau)/J(\mathbf{R}')$ and accepted with probability p defined in Eq. (9). To correct for the wrong dynamics, we should multiply the weights of the secondary walkers by

$$r \frac{G_s(\mathbf{R}^{s'}, \mathbf{R}^s, \tau)}{T(\mathbf{R}', \mathbf{R}, \tau)/J(\mathbf{R}')}, \qquad (10)$$

where $r = p_s/p$ if the move is accepted and $r = (1-p_s)/(1-p)$ if the move is rejected. However, these products fluctuate wildly (r can be anywhere between zero and infinity). Therefore, it is not practical to follow this route to perform correlated sampling unless bounds can be placed on the ratios while at the same time ensuring that unbiased results are obtained in the $\tau \to 0$ limit.

An additional complication is the common practice in fixed-node DMC to reject moves that cross nodes. If primary and secondary walkers were to be treated on the same footing (p_s set to zero when the secondary walker crosses its own nodes), the weights of the secondary walkers would all become zero in a sufficiently long run. Even though this

problem can be easily overcome since it is legitimate to do fixed-node DMC allowing walkers to cross nodes,[11] reweighting as in Eq. (10) remains impractical due to the large fluctuations.

In this paper, we propose an alternative correlated sampling algorithm which is approximate but very accurate. Given the successful implementation of correlated sampling within VMC and the large gain in efficiency obtained in DMC when including an accept/reject step, we wish to devise a scheme that differs as little as possible from VMC, reduces to the DMC algorithm with accept/reject for the primary geometry, and yields results very close to the DMC value for the secondary geometry.

Observe that, in the absence of the growth/decay factor and presence of the accept/reject step, we would be sampling ψ^2 for the primary walk, and ψ_s^2 for the secondary walk by reweighting the averages with the ratio of wave functions [Eq. (6)]. By multiplying the weights of the primary and secondary walkers by the corresponding growth/decay factors, we recover the fixed-node solution for the primary walk, but we do so only approximately for the secondary walk since the moves were not proposed with the right dynamics. To partially correct for this, we introduce a secondary time step as discussed below.

We summarize our algorithm as follows: (1) We generate secondary walks from the reference walk according to the space-warp transformation. (2) In the averages, we retain the ratios of the secondary and primary wave functions as in VMC [Eqs. (5) and (6)]. (3) The secondary weights are the primary ones multiplied by the product of the factors $\exp[S_s(\mathbf{R}^{s'},\mathbf{R}^s,\tau_s) - S(\mathbf{R}',\mathbf{R},\tau)]$ for the last N_{proj} generations. (N_{proj} is chosen large enough to project out the secondary ground state, but small enough to avoid a considerable increase in the fluctuations.) In the exponential factors, we introduced τ_s because the secondary moves are effectively proposed with a different time step, τ_s, in the drift-diffusion term of Eq. (8). A sensible definition of τ_s is $\tau_s = \tau \langle \Delta R_s^2 \rangle / \langle \Delta R^2 \rangle$, where ΔR is the displacement resulting from diffusion, and ΔR_s is the displacement needed to take the secondary walker from its drifted position to the position specified by the space-warp transformation. τ_s is computed over the first equilibration blocks of the DMC run.

In the limit of vanishing displacement, the difference of primary and secondary energies and its statistical error vanish linearly, so the force and its error are well behaved.

Secondary geometry wave functions. We considered three choices for secondary geometry wave functions:

(1) The secondary wave functions have the same parameters $\{\mathbf{p}\}$ as the primary one but the coordinates are relative to the new nuclear positions: $\psi_s(\mathbf{R}_i,\mathbf{R}_\alpha^s) = \psi(\mathbf{R}_i,\mathbf{R}_\alpha^s,\mathbf{p}_s)$ with $\mathbf{p}_s = \mathbf{p}$, possibly with the minimal changes required to impose the cusp conditions.

(2) The secondary geometry wave functions at warped electron positions are related to the primary ones at the original positions, $\psi_s(\mathbf{R}_i^s,\mathbf{R}_\alpha^s) = \psi(\mathbf{R}_i,\mathbf{R}_\alpha,\mathbf{p})/\sqrt{J(\mathbf{R}_i)}$. This wave function depends on the transformation (it was used in Ref. 9(b) with a different transformation) and has the advantage that the weights W_i in [Eq. (2)] are unity. Surprisingly, it gives larger fluctuations of the energy differences than choice (1).

(3) $\psi_s(\mathbf{R}_i,\mathbf{R}_\alpha^s) = \psi(\mathbf{R}_i,\mathbf{R}_\alpha^s,\mathbf{p}_s)$ with reoptimized parameters $\mathbf{p}_s$. This choice gives the smallest fluctuation of the energy differences and the best potential energy surface.

We calculate all molecules with choice (1) but also demonstrate the superiority of choice (3) for B_2.

Results and conclusions. The algorithms presented in the previous sections are tested on first-row homonuclear dimers. The primary wave functions[3] were optimized close to the experimental bond length by the variance minimization method.[2] The potential energy curves were obtained with correlated sampling from ten geometries, using the warp transformation and recentered secondary geometry wave functions [choice (1) above]. Values of $N_{\text{proj}}\tau$ of 5–10 H^{-1} were sufficient to project out the secondary wave functions.

To ascertain the efficiency of our method, we performed two additional calculations for B_2; in the first, we omitted the warp transformation, whereas in the second we employed reoptimized, rather than recentered, secondary wave functions. In Fig. 1, we present the VMC root-mean-square fluctuations (σ_{VMC}) of the relative energy of primary and secondary geometries divided by the atomic displacement, $\Delta E/\Delta R$, for B_2. Introducing the warp transformation yields a reduction of about a factor of 3.5–5 in σ_{VMC}, which corresponds to a factor of 12–25 saving in computer time. Moreover, σ_{VMC} is only slightly dependent on the secondary geometry used. As expected, a further reduction in σ_{VMC} is obtained when the space-warp transformation is used in combination with reoptimized, rather than recentered, secondary geometry wave functions. The space-warp transformation was found to be of even greater help for heavier molecules, e.g., for F_2 the reduction in the fluctuations was at least a factor of 3.5–10. In fact, in the absence of space-warp transformation for F_2, it is even difficult to reliably estimate the statistical error of secondary geometries that differ considerably from the primary one.

To test the accuracy of our DMC correlated sampling algorithm, we performed DMC runs for H_2 and B_2 for three different primary geometries, (a) the equilibrium geometry, (b) a geometry stretched by 0.2, and (c) by -0.2 a.u. The runs (a), (b), and (c) should give identical potential energy curves if the algorithm were exact. In Fig. 2, we show results for B_2 that reveal the high accuracy of our DMC algorithm: the three DMC curves are very close and clearly distinguishable from the VMC results. These results are confirmed by the calculations for H_2 where, despite the use of an intentionally poor wave function, the three curves gave the equilibrium bond lengths (a) 1.4014(2), (b) 1.4014(2), and (c) 1.4015(2) a.u. The true equilibrium bond length, from a careful fit to the results of Ref. 12, is 1.4011 a.u.

To test the improvement resulting from employing $\tau_s \neq \tau$, we performed, for H_2, DMC correlated sampling with $\tau_s = \tau$. Since $\tau_s > \tau$ for $\Delta R > 0$ and $\tau_s < \tau$ for $\Delta R < 0$, we expect this potential energy curve to yield an equilibrium bond length that is too short. The equilibrium bond length is indeed 1.4003(2) a.u., which is 4 standard deviations from the true bond length, whereas that obtained with our $\tau_s \neq \tau$ algorithm, 1.4014(2) a.u., is 1.5 standard deviation from the true bond-length.

Having ascertained the accuracy and efficiency of our algorithm, we computed the bond lengths of all first-row
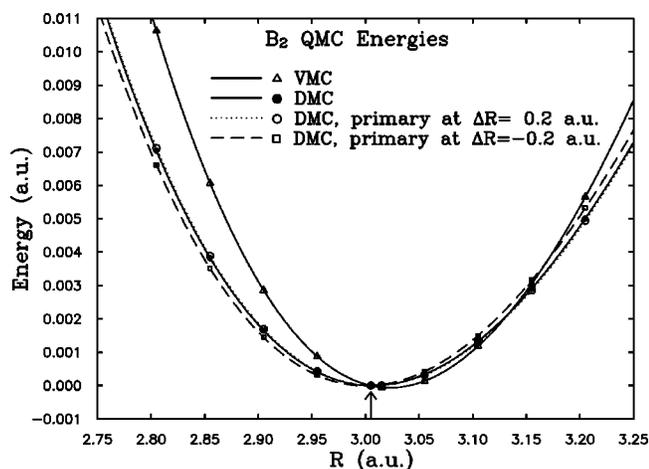
FIG. 2. Potential energy curve for B_2 in VMC and DMC. The three DMC curves are obtained with three different primary geometries (equilibrium, stretched by 0.2 and -0.2 a.u.) and using re-centered wave functions. All curves are shifted with the energy at the equilibrium distance (arrow) defined as the zero. Atomic units are used.

dimers with VMC and DMC correlated sampling. In Table I, we list the errors in the bond lengths obtained from restricted Hartree-Fock (RHF),[13] LDA,[14] GGA,[15] VMC, and DMC. The RHF results show the worst agreement with experiment, with Be_2 not being bound. The DMC errors are, in all cases,

TABLE I. Experimental bond lengths (Refs. 29 and 34 of Ref. 3) of first-row dimers and theoretical errors in RHF (Ref. 13), LDA (Ref. 14), GGA (Ref. 15), VMC and DMC (in a.u.).

Molecule	Expt.	RHF	LDA	GGA	VMC	DMC
Li_2	5.051	0.270	0.069	0.057	0.101(2)	0.018(3)
Be_2	4.630		-0.109	-0.001	$-0.069(3)$	$-0.014(5)$
B_2	3.005	0.086	0.025	0.042	0.018(2)	0.002(2)
C_2	2.348	-0.007	0.006	0.023	0.006(2)	0.008(1)
N_2	2.074	-0.061	-0.006	0.011	0.012(2)	0.007(1)
O_2	2.282	-0.107	-0.012	0.044	0.028(2)	0.023(4)
F_2	2.668	-0.161	-0.053	0.040	0.021(4)	0.015(5)
rms		∞	0.054	0.036	0.049	0.014

either smaller than or comparable to those from VMC, and are smaller than LDA and GGA errors by a factor of 3.9 and 2.6, respectively.

In this paper, we presented an efficient method to compute numerical forces in DMC, a long-standing unsolved problem in QMC techniques. The method is very accurate and was tested on first-row dimers, where the DMC bond lengths were found to agree with experiment better than those from HF, LDA, GGA, and VMC.

[1] D. M. Ceperley and B. J. Alder, Phys. Rev. B **36**, 2092 (1987); S. Fahy, X. W. Wang, and S. G. Louie, *ibid.* **42**, 3503 (1990); J. C. Grossman and L. Mitas, Phys. Rev. Lett. **79**, 4353 (1997); R. Q. Hood *et al.*, *ibid.* **78**, 3350 (1997).

[2] C. J. Umrigar, K. G. Wilson, and J. W. Wilkins, Phys. Rev. Lett. **60**, 1719 (1988).

[3] C. Filippi and C. J. Umrigar, J. Chem. Phys. **105**, 213 (1996); for O_2, we used an improved wave function.

[4] H. Hellman, *Einführung in die Quanten Theorie* (Deuticke, Leipzig, 1937); R. P. Feynman, Phys. Rev. **56**, 340 (1939).

[5] K. C. Huang, R. J. Needs, and G. Rajagopal, J. Chem. Phys. **112**, 4419 (2000).

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

[7] R. E. Lowther and R. L. Coldwell, Phys. Rev. A **22**, 14 (1980).

[8] C. J. Umrigar, Int. J. Quantum Chem., Symp. **23**, 217 (1989).

[9] (a) P. J. Reynolds *et al.*, Int. J. Quantum Chem. **29**, 589 (1986); (b) C. A. Traynor and J. B. Anderson, Chem. Phys. Lett. **147**, 389 (1988); (c) J. Vrbik and S. M. Rothstein, J. Chem. Phys. **96**, 2071 (1992).

[10] J. B. Anderson, J. Chem. Phys. **63**, 1499 (1975); **65**, 4121 (1976); P.J. Reynolds *et al.*, *ibid.* **77**, 5593 (1982).

[11] C. J. Umrigar, M. P. Nightingale, and K. J. Runge, J. Chem. Phys. **99**, 2865 (1993).

[12] L. Wolniewicz, J. Chem. Phys. **99**, 1851 (1993).

[13] K. A. Peterson, R. A. Kendall, and T. H. Dunning, J. Chem. Phys. **99**, 9790 (1993).

[14] A. D. Becke, Phys. Rev. A **33**, 2786 (1986); R. M. Dickson and A. D. Becke, J. Chem. Phys. **99**, 3898 (1993).

[15] F. W. Kutzler and G. S. Painter, Phys. Rev. B **45**, 3236 (1992); they employ the Perdew-Wang '86 GGA.