

Geometry optimization in quantum Monte Carlo with solution mapping: application to formaldehyde

C. A. Schuetz, M. Frenklach

Department of Mechanical Engineering, University of California at Berkeley, Berkeley, CA 94720-1740, and Environmental Energy Technologies Division, Lawrence Berkeley National Laboratory

A. C. Kollias, and W. A. Lester, Jr.

Kenneth S. Pitzer Center for Theoretical Chemistry, Department of Chemistry, University of California, Berkeley, CA 94720-1460, and Chemical Sciences Division, Lawrence Berkeley National Laboratory

The solution mapping methodology is presented as a method of geometry optimization of molecules in the quantum Monte Carlo method. Applied to formaldehyde as a test system, this approach is found to yield optimized bond lengths and bond angle in the diffusion Monte Carlo method that lie within experimental error. The variational Monte Carlo optimized geometry also lies within experimental error, with the exception of the CH bond length, which is slightly underestimated. Additionally, the resulting quadratic representation of the potential energy surface in the region of the minimum is used to calculate three of the force constants and harmonic frequencies.

I. INTRODUCTION

The quantum Monte Carlo (QMC) method is a stochastic many-body approach to solving the Schrödinger equation.¹⁻⁴ Although several techniques are used to provide optimized geometries in QMC⁵⁻⁹, there is presently no general method to address this need. In ab initio basis set methods, analytical gradients of the energy with respect to nuclear positions are readily available, and are routinely used to determine critical points on the potential energy surface (PES).¹⁰⁻¹³ Although QMC methods yield molecular energies of high accuracy, they generally do not provide gradients of the energy with respect to the nuclear positions. Therefore, QMC energies are traditionally calculated at PES critical points that have been taken from other methods. To date, many QMC calculations have been carried out that validate this approach; see for example Refs. 14,15. However, it would be beneficial to determine both the energies and locations of PES critical points, entirely within the QMC method.

Difficulties arise in the direct evaluation of QMC derivatives due to terms involving the reciprocal of electron-nuclear distances, i.e., $\left(1/r_{i\alpha}^{\rightarrow}\right)^3$ found in the first derivative, and $\left(1/r_{i\alpha}^{\rightarrow}\right)^5$ found in the second derivative of the potential. During sampling of the force, these terms in analytic derivative expressions cause calculated variances to grow rapidly.⁸ Caffarel *et al.* introduced a zero variance theorem for determining the Hellmann-Feynman forces.^{16,17} In this approach, terms that lead to large variance are canceled, making it possible to evaluate the average force with a relatively small variance. Recently Casalegno *et al.*¹⁸ calculated forces by introducing correction terms to the Hellmann-Feynman force as proposed by Pulay.¹¹ To date, such calculations have been carried out on several small diatomic systems and have led to good agreement with experimental bond lengths. Tanaka⁷ has used the Car-Parrinello¹⁹ method to optimize the geometric and wave function parameters of the water molecule. By introducing a

Lagrangian, the Euler-Lagrange equations generate dynamics for the geometric and wave function parameters. The problem of geometry optimization is reformulated to solving classical equations of motion. Forces are obtained using correlated sampling with fixed iterative derivatives.²⁰ At present, water is the largest molecular system for which a VMC geometry optimization has been published.⁷

Another reason the techniques of geometry optimization, as carried out in ab initio basis set quantum chemistry and density functional theory (DFT), are not directly applicable to QMC is the cost of optimizing trial wave function parameters. In QMC, the trial wave function is often written as the product of an independent-particle function and a correlation function that depends explicitly on interparticle distances. Ideally, the coefficients of both of these wave function factors would be optimized for a given geometry. However, it is common practice to hold the independent-particle function fixed, and optimize only the correlation function to conserve computational effort. In principle, one can determine the force on each atom and pursue a steepest descent to the optimal geometry, but the Slater determinant and correlation function must be re-optimized at each geometry. For small systems one can readily construct an independent-particle function (single- or multi-determinant) for parameter optimization at each geometry with the QMC method.²¹ However, as system size grows, such an approach becomes impractical because of the computer time required. A geometry optimization procedure along these lines for diffusion Monte Carlo (DMC) has recently been applied to diatomic systems by Fillipi *et al.*^{5,6}

In this study, we examine the alternative of computing an approximate QMC PES by employing the numerical approach of solution mapping. In this approach, the minimum energy and optimal geometry can be determined solely based on QMC energies. Our example molecule

is formaldehyde, which is optimized using only single point energy calculations from single determinant, effective core potential variational Monte Carlo (VMC) and DMC methods. Formaldehyde was chosen as a test case, because there is good experimental data on the geometry of the system and its size presents a nontrivial example of the methods proposed here.

II. SOLUTION MAPPING

The solution mapping (SM) methodology has evolved from the needs of reactive flow simulations of combustion processes,^{22,23} where the large size of reaction models has driven the development of economic numerical strategies. In the SM approach an approximation is sought not to the mathematical equations that define the model, but to the solution of these equations.²⁴ The approximation is developed through the statistical techniques of response surface design,²⁵⁻²⁷ by performing a relatively small number of computer experiments with the original model (QMC in the present case) and fitting the numerical results with simple functions such as polynomials. A statistical surrogate obtained in this manner is then used in subsequent numerical calculations (potential energy minimization in the present case), replacing the need for repeated solutions of the original equations.

In the present study, we apply SM to develop a functional dependence of the potential energy of formaldehyde, computed with QMC, on geometric molecular properties. We assume the optimal geometry to be planar and symmetric about the C-O bond, and hence the PES of formaldehyde to be defined by three internal coordinates: the C-O bond length, the C-H bond length, and the OCH angle. Conventionally, for the field of mathematical statistics, the internal coordinates are expressed in a dimensionless variable through the transformation

$$x_i = \frac{S_i - S_{i,\text{center}}}{\Delta S_i} \quad (1)$$

where s_i is the i th internal coordinate, $s_{i,\text{center}}$ the center about which s_i is varied, and Δs_i the span of s_i . The centers were set at the experimental values of $r(\text{C}=\text{O}) = 1.21 \text{ \AA}$, $r(\text{C}-\text{H}) = 1.12 \text{ \AA}$, and $\angle\text{OCH} = 121^\circ$.²⁸ The spans were 0.05 \AA in the bond lengths and 3° in OCH angle, or 5 times the experimental error in the bond lengths and 3 times the experimental error in the OCH angle. Our choice of Δs values is a compromise between accuracy of the approximation to the PES, which increases as variable spans decrease, and mapping the largest possible area.

The computer experiments are performed at preselected combinations of the factorial variables, and the set of these computer experiments is called a factorial design. Factorial designs originate from rigorous analysis of variance, with the objective of minimizing the number of computer experiments that must be performed in order to gain the desired information; for further discussion see Refs. 25-27.

A 2^3 central composite orthogonal design,²⁹ shown in Table I, was employed in the present study. This notation implies that the design is geared toward producing a quadratic approximation to a function of 3 independent variables. As can be seen in Table I, the design consists of a total of 15 computer experiments. The first 8 of these are performed at the corner points of the chosen region, and the 9th point is at the center of the region. The remaining 6 computer experiments are at “star” points, which are defined so that the columns of the design matrix are orthogonal.

The results of VMC and DMC single-point energy calculations for each of the computer experiments in the design matrix, are presented in Table I. These QMC energies were fit to the following second-order polynomial of the factorial variables,

$$V(x_1, x_2, x_3) = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_{11} (x_1^2 - \overline{x_1^2}) + b_{22} (x_2^2 - \overline{x_2^2}) + b_{33} (x_3^2 - \overline{x_3^2}) + b_{12} x_1 x_2 + b_{13} x_1 x_3 + b_{23} x_2 x_3, \quad (2)$$

where the square terms are re-centered by their respective averages, $\overline{x_i^2} \equiv \frac{1}{15} \sum_{u=1}^{15} x_{u,i}^2$, to attain column orthogonality of the extended design matrix yielding

$$\mathbf{X} = [\mathbf{x}_0 \quad \mathbf{x}_1 \quad \mathbf{x}_2 \quad \mathbf{x}_3 \quad \mathbf{x}_{11} \quad \mathbf{x}_{22} \quad \mathbf{x}_{33} \quad \mathbf{x}_{12} \quad \mathbf{x}_{13} \quad \mathbf{x}_{23}]. \quad (3)$$

In Eq. (3), the \mathbf{x}_i and \mathbf{x}_{ij} are orthogonal column vectors, with $\mathbf{x}_0 \equiv \mathbf{1}$, $\mathbf{x}_{ii} = \mathbf{x}_i^2 - \overline{\mathbf{x}_i^2}$, and $\mathbf{x}_{ij} = \mathbf{x}_i \mathbf{x}_j$.

The \mathbf{X} matrix rows are the values that the column vectors \mathbf{x} assume in the computer experiments prescribed by the factorial design. Under these conditions the matrix $(\mathbf{X}^T \mathbf{X})^{-1}$ is diagonal, and the coefficients in Eq. (2) are determined by

$$b_k = \frac{\sum_{u=1}^{15} X_{u,k} V_u}{\sum_{u=1}^{15} X_{u,k}^2}, \quad (4)$$

where $X_{u,k}$ is the (u,k) -th element of matrix \mathbf{X} or the value of the k th variable, \mathbf{x}_k , in Eq. (3) realized in the u th computer experiment. The subscript k in this notation spans all the indices of the column vectors in Eq. (3). With Eq. (2) established, the minimum energy and optimized geometry, for each of the QMC methods, was determined by numerical optimization of the respective polynomials constrained to the chosen domain, namely, $[-1, +1]$ of the factorial variables.

The energy found in each QMC calculation has an uncertainty, which results in an uncertainty in the statistical surrogate to the PES, and all properties that are derived from it. The uncertainty of a point on the response surface can be evaluated by calculating the variance of Eq. (2). By virtue of the orthogonality of the columns of \mathbf{X} , and the consequent independence of the estimates of all b 's in Eq. (2), the variance of the potential energy estimate at a given point, (x_1, x_2, x_3) , is

$$\begin{aligned}
\sigma^2 \{V(x_1, x_2, x_3)\} &= \sigma^2 \{b_0\} + x_1^2 \sigma^2 \{b_1\} + \dots + x_3^2 \sigma^2 \{b_3\} \\
&+ \left(x_1^2 - \overline{x_1^2}\right)^2 \sigma^2 \{b_{11}\} + \dots + \left(x_3^2 - \overline{x_3^2}\right)^2 \sigma^2 \{b_{33}\} \\
&+ x_1^2 x_2^2 \sigma^2 \{b_{12}\} + \dots + x_2^2 x_3^2 \sigma^2 \{b_{23}\},
\end{aligned} \tag{5}$$

where the variance of the coefficients is given by

$$\sigma^2 \{b_k\} = \frac{\sum_{u=1}^{15} x_{u,k}^2 \sigma^2 \{V_u\}}{\left(\sum_{u=1}^{15} x_{u,k}^2\right)^2}, \tag{6}$$

and $\sigma^2 \{V_u\}$ is the variance of the u th QMC energy calculation. Inspection of Eq. (5), shows that the uncertainty in the response surface is small for points near the center of the design, $\mathbf{s} = \mathbf{s}_{\text{center}}$, where the values of x_i are close to zero, and grows larger at the edges of the fitted region, where the values of $|x_i|$ approach unity.

III. QMC METHODOLOGY

The DMC method is a stochastic approach for obtaining time independent solutions to the Schrödinger equation by solving the time-dependent Schrödinger equation in imaginary time^{1-4,9,30-34}. Fermion antisymmetry is introduced through the fixed-node approximation (FNA), which imposes the nodes of an approximate function Ψ_T onto the unknown exact function Φ . The FNA has been found to provide accurate results for ground and excited states with trial functions constructed using various ab initio basis set methods. The introduction of information about the system from another source is the essence of importance sampling,³⁵ which improves convergence to the state of interest. In the DMC method the propagation walker distribution is achieved with the short-time approximation, which provides an analytical approximation to the Green's function.³⁶

In the present study Ψ_T is written as a product of a single determinant D and a correlation function $U(r_{i\alpha}, r_{ij})$,

$$\Psi_T = D \exp(U(r_{i\alpha}, r_{ij})) \quad (7)$$

where $U(r_{i\alpha}, r_{ij})$ depends explicitly on interparticle coordinates. The orbitals of the single determinant in this study are Hartree-Fock (HF) and natural orbitals (NOs) obtained from HF and MCSCF calculations using the GAMESS quantum chemistry package.³⁷

The form chosen for $U(r_{i\alpha}, r_{ij})$ is the 9-parameter function adapted by Schmidt and Moskowitz³⁸ from a function introduced by Boys and Handy (SMBH).³⁹ This function contains two- and three-body terms in the form of electron-electron, electron-nucleus, and electron-other-nucleus distances. The SMBH correlation function contains first-order Jastrow terms that enable satisfying electron-electron and electron-nuclear cusp conditions.⁴⁰ The present DMC implementation includes a tenth function to satisfy the electron-nucleus cusp condition. The optimization of correlation function parameters is accomplished through fixed sample optimization using the absolute deviation (AD) functional^{41,42} that minimizes the energy of Ψ_T and is given by,

$$AD = \frac{1}{N} \sum_{i=1}^N |E_T - E_{L_i}| \quad (8)$$

Here N is the number of walkers, E_{L_i} is the local energy of the i th configuration, and E_T is reference energy chosen to minimize fluctuations.

The two trial functions were constructed from HF and single determinant MCSCF calculations. The C and O 1s atomic cores were replaced by Steven-Basch-Krauss (SBK)⁴³ and soft effective core potentials (ECP).^{44,45} The NOs for the MCSCF trial function were derived

from a small MCSCF calculation that included 30 virtual and all occupied valence orbitals in the active space. Due to computational constraints the soft ECPs were implemented for the HF trial wave functions while SBK ECPs were utilized for the MCSCF trial function.

The uncontracted Partridge basis set⁴⁶ was used with soft ECPs⁴⁵ for C and O and the cc-pVTZ basis set was introduced for H. The Partridge basis set was modified by the addition of a single d polarization basis function for O and C. At each geometry, the d basis function was variationally optimized at the HF level of theory. Optimization of the d basis functions on H resulted in negligible improvement in the HF energy (<0.1 kcal/mol). The final optimized exponents are presented in Table II. For the MCSCF calculations the cc-pVQZ basis was used with the SBK ECPs.

IV. GEOMETRY OPTIMIZATION

QMC energy calculations were performed at the 15 points of the factorial design using DMC and VMC. With each method, two different trial wave functions were employed: HF with soft-ECPs, and MCSCF with SBK ECPs. For each combination of QMC method and trial wave function, a statistical surrogate for the PES was obtained. The QMC energies are reported in Table I, and the coefficients of the quadratic fits are listed in Table III. The accuracy of the statistical surrogate can be assessed by the root mean squared (RMS) error, and the maximum deviation of the fitted surface from the QMC energies. The accuracy of the statistical surrogate is generally good, with RMS errors of 0.51 kcal/mol or less, as shown in Table IV. However, there are points of significant error, such as the statistical surrogate, the PES, for the VMC energies obtained with the MCSCF trial function. In this case there is a point where the statistical surrogate and the calculated QMC energy differ by 1.37 kcal/mol. This error is most likely due to a combination of the non-quadratic shape of the true PES, and the uncertainties in

the QMC energies. However, in each case, it can be seen that the one standard deviation (1σ) uncertainty in the statistical surrogate is significantly lower than that of the individual QMC energy calculations.

The statistical surrogates were used to find the minimum total energies and optimal geometries of the molecule. With the HF trial wave function, the optimized DMC geometry is within the uncertainty of the experimental values, as shown in Table V. For DMC with the MCSCF trial wave function, the optimized C-O bond length and OCH angle lie within the experimental uncertainties, yet the C-H bond length is slightly shorter than the experimental value. For both trial wave functions tested, the VMC optimal geometry lies within experimental uncertainty, with the exception of the C-H bond length. It is interesting to note that the C-H bond length is predicted to be shorter than the experimental value for three of the four methods and trial wave function combinations used. Additionally, the C-H bond length was the only geometric parameter that was predicted to lie outside experimental uncertainty, which suggests a systematic difficulty in predicting this property. It is also worth noting that the VMC geometries are only slightly less accurate than the DMC results and that a VMC energy calculation requires roughly one tenth of the computational expense of the corresponding DMC calculation.

The global minimum of the PES defines the optimal geometry, and thus the uncertainty in energy leads to an uncertainty in optimal geometry. Having a PES region parameterized with a simple algebraic expression, Eq. (2), enables us to examine the nature of the uncertainties in a detailed manner. By solving Eq. (2) for the internal coordinates, with $V = V_{\min} + \sigma$, where V_{\min} is the minimum of Eq. (2) and σ is taken from Eq. (5), the results can be visualized either as an energy isosurface in all three dimensions, or as a two-dimensional cross-section by setting one of the geometric parameters to its optimized value.

Figure 1 displays such results for DMC geometries obtained with the HF trial function. The ranges of the computed $1-\sigma$ region are quite similar to the reported experimental uncertainties, roughly ± 0.01 Å for the bond length, and $\pm 1^\circ$ for the OCH angle. The ranges of the computed $1-\sigma$ region for the VMC PES with the MCSCF trial wave function, shown in Fig. 2, are somewhat larger; roughly ± 0.018 and ± 0.03 Å for the C-O and C-H bonds, respectively, and $\pm 1.25^\circ$ for the OCH angle. In the case of the MCSCF VMC surface, the computed region of uncertainty extends beyond the fitted area. Still, the optimal geometry lies well within the fitted region and the uncertainties can be determined by considering the symmetry of the quadratic surface. The uncertainty ranges for each of the method and trial wave function combinations are reported along with the optimized geometric parameters in Table V.

V. FORCE CONSTANTS AND FREQUENCIES

Using the statistical surrogate to the PES, it is also possible to calculate some of the force constants and vibrational frequencies. Only three internal coordinates were used to map the PES, thus the force constants and vibrational frequencies can only be determined for the: C-O stretch, C-H symmetric stretch, and the C-H bending modes. The force constants and normal mode frequencies were determined using Eq. (2) and the method of Wilson et al.⁴⁷ The results are reported in Table VI, in the symmetry coordinates given in Ref. 48. Inspection of these results indicates that the errors in the force constants and the vibrational frequencies are sizable. (it is pertinent to mention that such error levels are not uncommon for ab initio methods.) The relatively large errors in the present method originate from the double differentiation of Eq. (2), and improving the quality of fit, through a smaller fitting domain or using other designs, can reduce them.

The propagation of the variance in the QMC energy calculations contributes to the errors

in frequencies. The latter can be determined from the coefficients of the second order terms in Eq. (5). The uncertainty in the harmonic frequencies is more difficult to characterize. We estimated the uncertainty approximately by calculating the change in frequency associated with a change of σ in the force constants, or

$$\sigma_\nu \approx \frac{\nu(F + \sigma_F) - \nu(F - \sigma_F)}{2} \quad (9)$$

The calculated uncertainties in the force constants and frequencies are shown in Table VI. In most cases, they are systematically larger than the QMC variance. This suggests the existence of systematic errors, and the most likely source is the accuracy of the quadratic approximation of the PES, Eq. (2), as discussed above.

VI. CONCLUSIONS

This study demonstrates that QMC, aided by SM, can be used to reliably perform geometry optimizations of polyatomic systems. A statistical surrogate for the PES was determined for a relatively large region, and with DMC and a HF trial function, the optimal geometry was found to lie within experimental error. For the DMC with MCSCF trial function and the VMC optimal geometries, only the predictions of C-H bond length lie outside the experimental uncertainty. Additionally, the total energy, force constants, and vibrational frequencies were determined. The accuracy of the force constants and vibrational frequencies were relatively low and indicate that further work is needed to improve these properties in the present approach: one could, for example, improve the accuracy of the frequencies by treating a smaller region of the PES, by employing a better functional form for the statistical surrogate, by using an improved factorial design, and by reducing the variance of the QMC calculations.

In general, the solution mapping methodology provides an analytical framework for

conducting a series of numerical experiments and developing accurate numerical models. The approach has two properties that make it particularly well suited for use with QMC. First, a minimum number of calculations are required to develop the approximate PES, which is of paramount importance because of the high computational cost of QMC calculations. Second, the resulting surface generally has a variance that is lower than the variance of the individual computer experiments. Consequently, the approach results in energy predictions that are more accurate than those directly obtained from the QMC calculations, and moderates the computational expense of geometry optimization so that polyatomic systems can be effectively studied.

ACKNOWLEDGEMENTS

This work was supported by the Director, Office of Science, Office of Basic Energy Sciences, Chemical Sciences Division of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098. The calculations were carried out at the National Energy Research Supercomputer Center (NERSC).

Literature Cited

- ¹A. Luchow and J. B. Anderson, in *Ann. Rev. Phys. Chem.*, edited by H. L. Strauss, G. T. Babcock, and S. R. Leone (Annual Reviews, Palo Alto, 2000), Vol. 51, pp. 501.
- ²L. Mitas and D. M. Ceperley, in *Adv. Chem. Phys.*, edited by I. Prigogine and S. A. Rice (Wiley, New York, 1996), Vol. 93, pp. 1.
- ³J. B. Anderson, in *Rev. Comput. Chem.*, edited by K. B. Lipkowitz and D. B. Boyd (Wiley, New York, 1999), Vol. 13, pp. 133.
- ⁴W. M. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, *Rev. Mod. Phys.* **73**, 33 (2001).
- ⁵C. Filippi and C. J. Umrigar, *Phys. Rev. B* **61**, R16291 (2000).
- ⁶C. Filippi and C. J. Umrigar, in *Recent Advances in Quantum Monte Carlo Methods*, edited by W. A. Lester, Jr., S. M. Rothstein, and S. Tanaka (World Scientific, New Jersey, 2002), pp. 12.
- ⁷S. Tanaka, *J. Chem. Phys.* **100**, 7416 (1994).
- ⁸B. L. Hammond, W. A. Lester, Jr., and P. J. Reynolds, *Monte Carlo Methods in Ab Initio Quantum Chemistry*. (World Scientific, Singapore, 1994).
- ⁹P. J. Reynolds, R. N. Barnett, B. L. Hammond, R. M. Grimes, and W. A. Lester, Jr., *Int. J. Quantum Chem.* **29**, 589 (1986).
- ¹⁰Y. Yamaguchi, Y. Osamura, J. D. Goddard, and H. F. Schaefer, III, *A New Dimension to Quantum Chemistry: Analytic Derivative Methods in Ab Initio Molecular Electronic Structure*

Theory. (Oxford University Press, New York, 1994).

¹¹P. Pulay, *Mol. Phys.* **17**, 197 (1969).

¹²P. Pulay, in *Applications of electronic structure theory*, edited by H. F. Schaefer, III (Plenum Press, New York, 1977), pp. 153.

¹³H. B. Schlegel, in *Computational Organic Chemistry: Proceedings of the NATA Advanced Study Institute held at Menton France, June 29-July 13, 1980*, edited by I. G. Csizmadia and R. Daudel (Dordrecht, Boston, 1981).

¹⁴J. C. Grossman, L. Mitas, and K. Raghavachari, *Phys. Rev. Lett.* **75**, 3870 (1995).

¹⁵Grossman J.C., W. A. Lester, Jr., and S. G. Louie, *J. Am. Chem. Soc.* **112**, 705 (2000).

¹⁶R. Assaraf and M. Caffarel, *Phys. Rev. Lett.* **83**, 4682 (1999).

¹⁷R. Assaraf and M. Caffarel, *J. Chem. Phys.* **113**, 4028 (2000).

¹⁸M. Casalegno, M. Mella, and A. M. Rappe, *J. Chem. Phys.* **118**, 7193 (2003).

¹⁹R. Car and M. Parrinello, *Phys. Rev. Lett.* **55**, 2471 (1985).

²⁰Z. Sun, B. L. Hammond, and W. A. Lester, Jr., *J. Chem. Phys.* **97**, 7585 (1992).

²¹C. J. Umrigar, *Int. J. Quantum Chem. Suppl.* **23**, 217 (1989).

²²M. Frenklach, in *Computational Fluid and Solid Mechanics*, edited by K. J. Bathe (New York, Elsevier, 2001), pp. 1177.

²³M. Frenklach, H. Wang, and M. J. Rabinowitz, *Prog. Energy Combust. Sci.* **18**, 47 (1992).

- ²⁴M. Frenklach, in *Combustion Chemistry*, edited by W. C. Gardiner, Jr. (Springer-Verlag, New York, 1984), pp. 423.
- ²⁵G. E. P. Box and N. Draper, *Empirical Model-Building and Response Surfaces*. (Wiley, New York, 1987).
- ²⁶G. E. P. Box, W. G. Hunter, and J. S. Hunter, *Statistics for Experimenters. An Introduction to Design, Data Analysis, and Model Building*. (Wiley, New York, 1978).
- ²⁷R. H. Myers and D. C. Montgomery, *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. (Wiley, New York, 2002).
- ²⁸M. W. Chase, Jr., C. A. Davies, J. R. Downey, Jr., D. J. Frurip, R. A. McDonald, and A. N. Syverud, *JANAF Thermochemical Tables*, 3rd ed. (American Institute of Physics for the National Bureau of Standards, New York, 1985).
- ²⁹V. V. Nalimov and N. A. Chernova, *Statistical Methods of Design of Extremal Experiments (in Russian)*. (Nauka, Moscow, 1965).
- ³⁰J. B. Anderson, *J. Chem. Phys.* **65**, 4121 (1976).
- ³¹J. B. Anderson, *Int. J. Quantum Chem.* **15**, 109 (1979).
- ³²P. J. Reynolds, D. M. Ceperley, B. J. Alder, and W. A. Lester, Jr., *J. Chem. Phys.* **77**, 5593 (1982).
- ³³P. J. Reynolds, M. Dupuis, and W. A. Lester, Jr., *J. Chem. Phys.* **82**, 1983 (1985).
- ³⁴R. M. Grimes, B. L. Hammond, P. J. Reynolds, and W. A. Lester, Jr., *J. Chem. Phys.* **85**, 4749

(1986).

³⁵M. H. Kalos, D. Levesque, and L. Verlet, *Phys. Rev. A* **9**, 2178 (1974).

³⁶B. L. Hammond, P. J. Reynolds, and W. A. Lester, Jr., *J. Chem. Phys.* **87**, 1130 (1987).

³⁷M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. J. Su, T. L. Windus, M. Dupuis, and J. A. Montgomery, Jr., *J. Comp. Chem.* **14**, 1347 (1993).

³⁸K. E. Schmidt and J. W. Moskowitz, *J. Chem. Phys.* **93**, 4172 (1990).

³⁹S. F. Boys and N. C. Handy, *Proc. R. Soc. London, Ser. A* **310**, 43 (1969).

⁴⁰R. Jastrow, *Phys. Rev.* **98**, 1479 (1955).

⁴¹A. Aspuru-Guzik, O. Couronne, I. Ovcharenko, and W. A. Lester, Jr., *Abstr. Pap. Am. Chem. Soc.* **221** (2001).

⁴²D. Bressanini, G. Morosi, and M. Mella, *J. Chem. Phys.* **116**, 5345 (2002).

⁴³W. J. Stevens, H. Basch, and M. Krauss, *J. Chem. Phys.* **81**, 6026 (1984).

⁴⁴C. W. Greeff and W. A. Lester, Jr., *J. Chem. Phys.* **109**, 1607 (1998).

⁴⁵I. Ovcharenko, A. Aspuru-Guzik, and W. A. Lester, Jr., *J. Chem. Phys.* **114**, 7790 (2001).

⁴⁶H. Partridge, *J. Chem. Phys.* **90**, 1043 (1989).

⁴⁷E. B. Wilson, J. C. Decius, and P. C. Cross, *Molecular Vibrations: The Theory of Infrared and Raman Vibrational Spectra*. (Dover, New York, 1980).

⁴⁸J. L. Duncan and P. D. Mallinson, Chem. Phys. Lett. **23**, 597 (1973).

TABLE I. A 2^3 central composite orthogonal design of computer experiments ($\alpha = 1.215$) and the calculated energies. The factorial variables x_1 , x_2 , and x_3 represent the C-O and C-H bond lengths and the OCH angle, respectively.

Computer experiment	Factorial variables			Energies (a.u.)			
	x_1	x_2	x_3	VMC/HF ^a	VMC/MCSCF ^b	DMC/HF ^a	DMC/MCSCF ^b
1	1	1	1	-22.771902	-22.781235	-22.831317	-22.838882
2	1	1	-1	-22.771458	-22.782648	-22.831180	-22.841812
3	1	-1	1	-22.778289	-22.789659	-22.836486	-22.843913
4	1	-1	-1	-22.780129	-22.794528	-22.837767	-22.847728
5	-1	1	1	-22.779199	-22.790564	-22.837084	-22.843830
6	-1	1	-1	-22.777801	-22.786542	-22.834080	-22.842110
7	-1	-1	1	-22.782454	-22.795032	-22.838092	-22.842738
8	-1	-1	-1	-22.781878	-22.792426	-22.836544	-22.846043
9	0	0	0	-22.785688	-22.799110	-22.843844	-22.850806
10	α	0	0	-22.777967	-22.789146	-22.837378	-22.843221
11	$-\alpha$	0	0	-22.782583	-22.793107	-22.839235	-22.846519
12	0	α	0	-22.777506	-22.790659	-22.839188	-22.845898
13	0	$-\alpha$	0	-22.784607	-22.795886	-22.840857	-22.851588
14	0	0	α	-22.785676	-22.791812	-22.842597	-22.849443
15	0	0	$-\alpha$	-22.784442	-22.793120	-22.841076	-22.846355

^aTrial function with HF orbitals. The C and O 1s cores were replaced by a soft ECP^{44,45} and the remaining electrons were described by the Partridge basis set;⁴⁶ H was described by the cc-pVTZ basis set.

^bFrom reference 45; a single reference trial function using natural orbitals (NOs) obtained from a MCSCF calculation with double excitations of all valence electrons into 30 virtual orbitals. The C and O 1s cores were replaced by SBK ECPs⁴³ and the remaining electrons were described using the cc-pVQZ basis set.

TABLE II. Hartree-Fock optimized d-polarization basis functions for C and O atoms.

Computer experiment	Atoms ^a	
	C	O
1	0.856687	0.954891
2	0.848945	0.956350
3	0.886927	0.958964
4	0.872974	0.959773
5	0.945166	1.056393
6	0.936912	1.057879
7	0.970295	1.062937
8	0.957129	1.063311
9	0.905953	1.007276
10	0.858770	0.948128
11	0.961989	1.074231
12	0.891856	1.004575
13	0.921809	1.010570
14	0.911303	1.006296
15	0.899330	1.006703

^aThe uncontracted Partridge basis set⁴⁶ with the soft ECP for C and O,^{44,45} was augmented with optimized d-polarization functions.

TABLE III. The coefficients of the quadratic approximation to the PES^a in hartrees.

Coefficients ^b	VMC		DMC	
	HF ^c	MCSCF ^d	HF ^c	MCSCF ^d
b_0	-22.7860	-22.7970	-22.8443	-22.8508
b_1	2.2972	1.9452	1.0322	0.5837
b_2	2.8318	3.3784	1.5753	1.8900
b_3	-0.1897	0.1135	-0.4799	0.4178
b_{11}	3.9480	3.5144	4.1593	4.0387
b_{12}	0.9657	1.2440	1.0355	1.0133
b_{13}	0.4212	1.6137	0.7120	0.6450
b_{22}	3.4190	2.0617	2.9976	1.4169
b_{23}	-0.3882	-0.6090	-0.3592	-0.7388
b_{33}	0.7095	2.6077	1.7697	1.9883

^aThe energies are fit to a function of the form given in Eq. (2).

^bCoefficients b_1 through b_{33} are multiplied by 1×10^3

^cSee footnote a of Table II for description of trial function.

^dSee footnote b of Table II for description of trial function.

TABLE IV. The accuracy of the QMC energy calculations and the error in fitting the QMC energies to the statistical surrogate, in kcal/mol.

Method	Average σ_{QMC}	RMS fitting error	Maximum deviation in fitted points	σ of statistical surrogate
VMC				
HF ^a	0.36	0.164	0.37	0.23
MCSCF ^b	0.42	0.489	1.37	0.25
DMC				
HF ^a	0.23	0.26	0.77	0.15
MCSCF ^b	0.37	0.51	1.29	0.22

^aSee footnote a of Table II for description of trial function.

^bSee footnote b of Table II for description of trial function.

TABLE V. The minimum energies and optimized geometries found with the VMC and DMC methods. The computed uncertainties, from the ranges of the isosurface $V_{\min} + \sigma$, are listed in parentheses.

Method	Minimum energy (a.u.)	Optimal geometry		
		$r(\text{C}=\text{O})$ (Å)	$r(\text{C}-\text{H})$ (Å)	$\angle\text{COH}$ (°)
VMC				
HF ^a	-22.7868295	1.197(15)	1.101(15)	121.3(2.25)
MCSCF ^b	-22.7984708	1.204(18)	1.08(3)	120.8(1.25)
DMC				
HF ^a	-22.8445736	1.20(1)	1.11(1)	121.4(1.0)
MCSCF ^b	-22.8515749	1.212(15)	1.08(15)	120.2(1.13)
Experiment ^c	—	1.21(1)	1.12(1)	121(1)

^aSee footnote a of Table II for description of trial function.

^bSee footnote b of Table II for description of trial function.

^cReference 28.

TABLE VI. The force constants, F , and vibrational frequencies, ν , for the three normal modes represented by the statistical surrogate. Units of force constants are mdyn \AA^{-1} for stretching, mdyn \AA for the bending modes, and cm^{-1} for frequencies. The uncertainties are given in parentheses.

Method	C-O stretch		C-H stretch		C-H bending	
	F	ν	F	ν	F	ν
VMC						
HF	13.8(1.9)	1820(194)	5.96(70)	3311(43)	0.38(21)	1275(88)
MCSCF	12.3(2.2)	1639(59)	3.59(82)	2759(85)	1.38(25)	2327(173)
DMC						
HF	14.5(1.2)	1785(56)	5.23(46)	3109(60)	0.94(14)	2087(55)
MCSCF	14.1(2.0)	1732(106)	2.47(74)	2446(75)	1.05(22)	1975(146)
Experiment ^c	12.903(62)	1764(17)	4.963(34)	2944(100)	0.570(4)	1563(15)

^aSee footnote a of Table II for description of trial function.

^bSee footnote b of Table II for description of trial function.

^cReference 48.

Figure Captions

FIG. 1. Visualizations of uncertainties in optimal geometry using Eq. (2) for the DMC calculations with HF trial wave function: (a) The $V_{\min} + \sigma$ isosurface; (b) $\angle\text{OCH}$ held to the optimal value; (c) $r(\text{C-H})$ held to the optimal value; and (d) $r(\text{C=O})$ held to the optimal value. The white ellipsoid in b, c, and d is the $V_{\min} + \sigma$ region. Potential energy is reported in a.u., bond lengths in Å, and bond angle in degrees.

FIG. 2. Visualizations of uncertainties in optimal geometry using Eq. (2) for the VMC calculations with MCSCF trial wave function: (a) The $V_{\min} + \sigma$ isosurface; (b) $\angle\text{OCH}$ held to the optimal value; (c) $r(\text{C-H})$ held to the optimal value; and (d) $r(\text{C=O})$ held to the optimal value. The white ellipsoid in b, c, and d is the $V_{\min} + \sigma$ region. Potential energy is reported in a.u., bond lengths in Å, and bond angle in degrees.

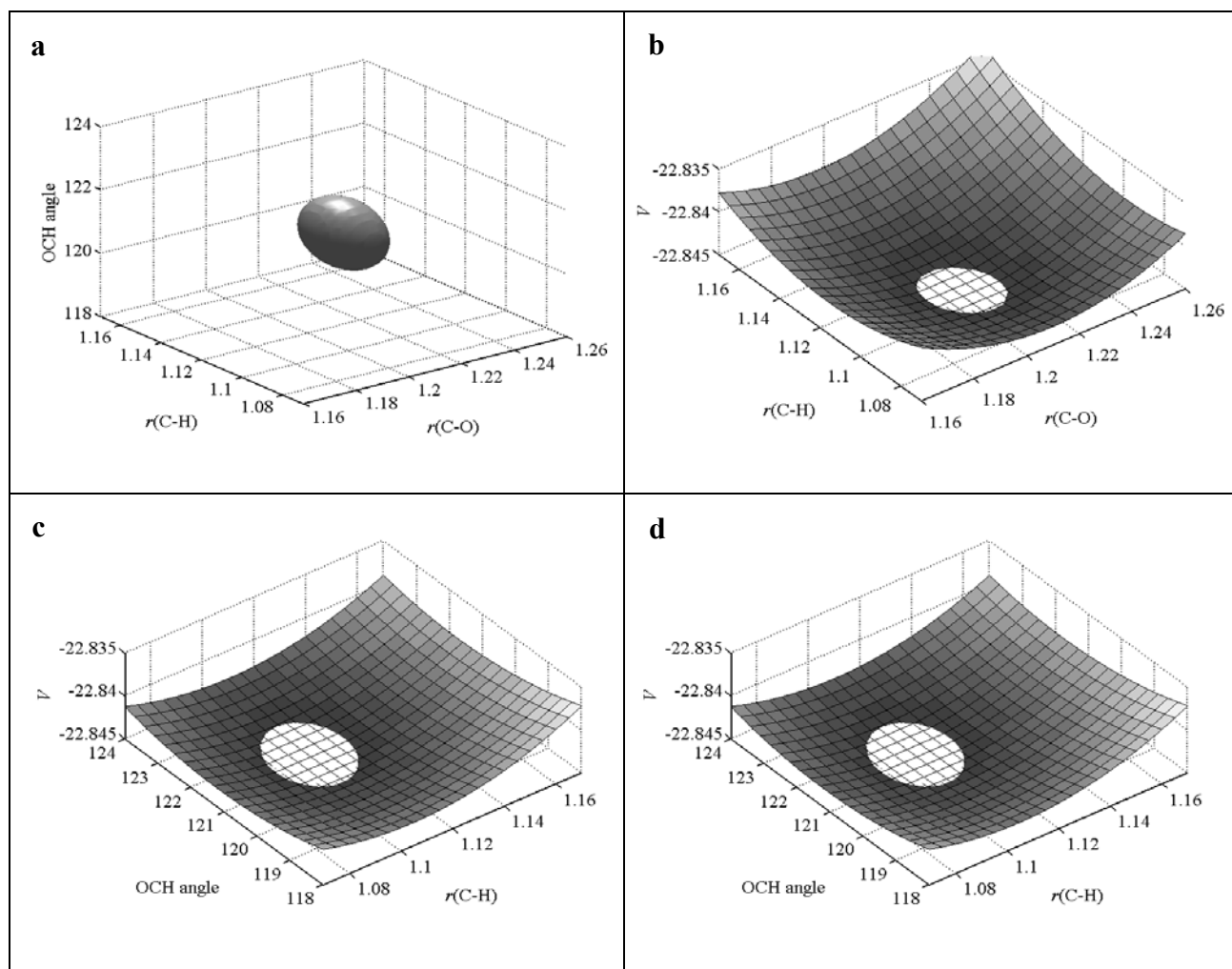


FIG. 1.

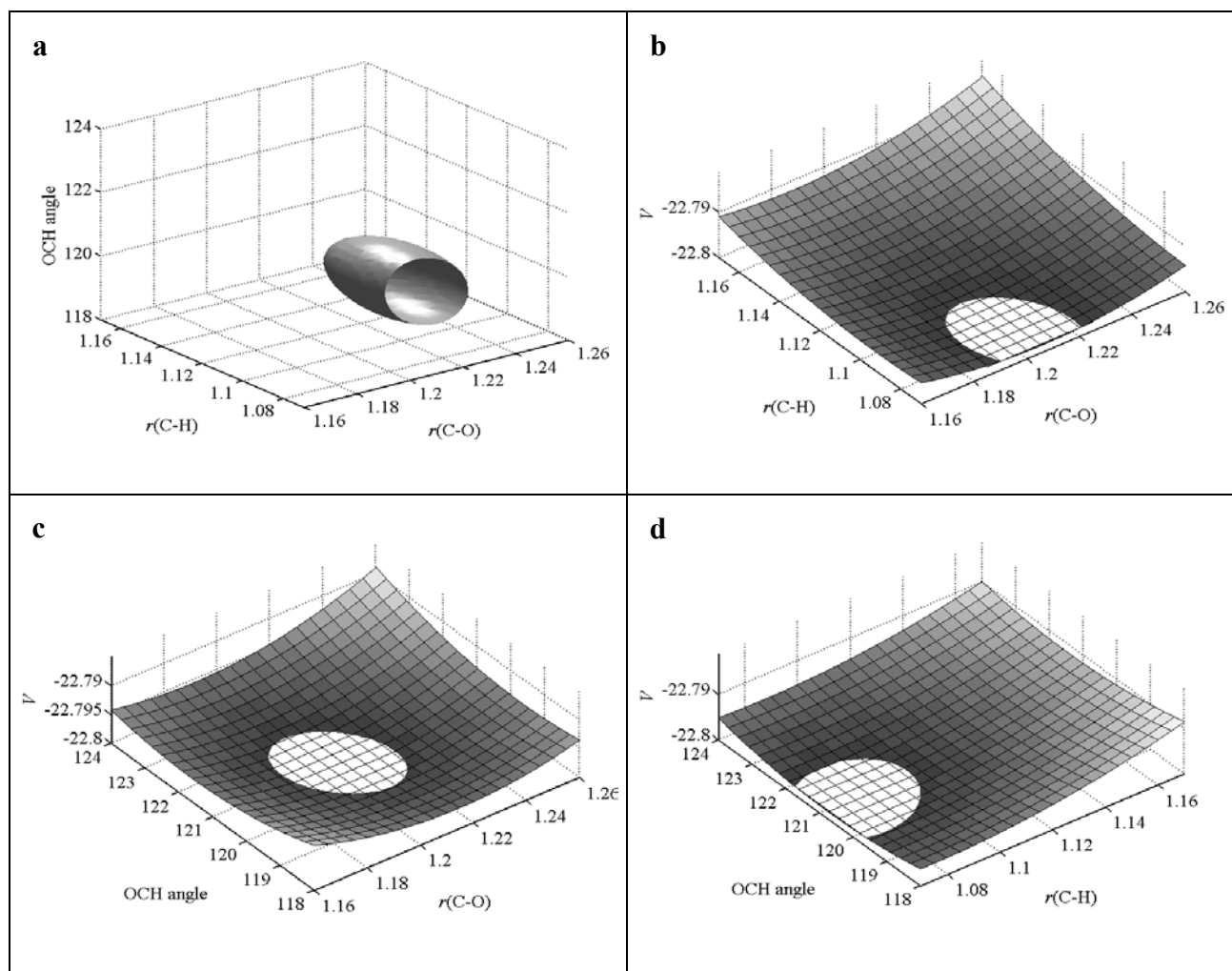


FIG. 2.