

# Direct inversion in the iterative subspace-induced acceleration of the ridge method for finding transition states

Irina V. Ionova and Emily A. Carter

*Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90095-1569*

(Received 22 May 1995; accepted 30 June 1995)

We extend the application of the direct inversion in the iterative subspace (DIIS) technique to the ridge method for finding transition states (TS). The latter is not a quasi-Newton-type algorithm, which is the only class of geometry optimization methods that has been combined with DIIS. With this new combination, we obtain a factor of two speedup due to DIIS, similar to the DIIS-related speedup achieved in other methods including quasi-Newton geometry optimization and self-consistent field iterations. We also demonstrate that DIIS is useful even in cases where optimization is started far from the quadratic region of the TS, provided that only one previous iteration is used in the DIIS expansion. We compare the performance of the new ridge-DIIS method to that of the TS algorithm utilized in GAUSSIAN 92. We find that the computational cost of the former is similar (when both methods converge) to that of the latter. The examples considered in the paper include a novel TS found for an isomerization of a cluster of six Na atoms. Locating such a TS poses a known problem for second-derivatives-based algorithms that fail on very flat potential energy surfaces. Thus, the gradient-based ridge-DIIS method is the only TS search method that is robust, does not need second derivatives and/or an initial guess for the TS geometry, and whose performance matches or exceeds that of a second-derivatives-based algorithm. © 1995 American Institute of Physics.

## I. INTRODUCTION

The process of searching for a transition state (TS) on a potential energy surface (PES) involves two stages. First, one has to identify the quadratic region of the TS of interest and, second, to refine the TS geometry. The latter can be dealt with rather efficiently by variable-metric second order algorithms.<sup>1-3</sup> These algorithms are well suited for the direct inversion in the iterative subspace (DIIS) procedure<sup>4</sup> that further improves their performance.<sup>5-7</sup>

It is the stage of approaching the TS region that is a source of problems in TS search methods. Sometimes, “chemical intuition” alone can supply a good guess for a TS geometry or identify a reaction coordinate. In the latter case, the TS region can be determined by reoptimizing the remaining degrees of freedom for fixed values of the reaction coordinate.<sup>3,7</sup> Unfortunately, it is often unclear what the TS structure should look like or what the reaction coordinate is; also, there is always a danger of introducing a bias by relying on such guesses.

On the other hand, algorithms that are designed to find a TS between known reactants and products either require substantial computational effort or are not sufficiently robust. The former methods essentially construct the entire reaction path,<sup>8-11</sup> while the latter are based on the assumption that the reaction path is quadratic.<sup>12,13</sup> The ridge method<sup>14</sup> recently proposed by us is positioned somewhere in the middle between the two sets of approaches mentioned above, both in terms of computational cost and reliability. It works by searching for the minimum on the ridge separating reactants’ and products’ wells on the PES. Unlike ascent along the reaction path that requires either a good approximation to the Hessian matrix (the matrix of second derivatives of energy with respect to molecular coordinates) or an

$(n - 1)$ -dimensional minimization (where  $n$  is the number of molecular degrees of freedom) at each step, the descent on the ridge can be performed by doing merely a one-dimensional maximization at each step. This algorithm is also capable of following curved ridges that correspond to curved reaction paths, which distinguishes it from less expensive but more failure-prone synchronous transit type methods.<sup>12,13</sup>

The major disadvantage of the ridge method is its inefficiency near the TS. The obvious solution would be a switch to fast second order methods once the quadratic region of a TS is reached. This, however, involves the introduction of a criterion to trigger such a switch, and necessitates obtaining the Hessian matrix. (Note that updating schemes<sup>15</sup> used to build an approximate Hessian in variable-metric methods make sense only in the quadratic region of the TS, and, therefore, no reasonable Hessian can be constructed *on the way to the TS region*).

The other way around the inefficiency of the ridge method near convergence is to implement DIIS. In the present paper, we show that this approach indeed results in substantial speedup of the original ridge method without sacrificing any of its attractive features. Ridge-DIIS is still robust, it still does not require a Hessian matrix or an initial guess for a TS geometry. We demonstrate that the computational cost of this method is comparable to that of GAUSSIAN 92,<sup>16</sup> which does not possess the advantages mentioned above.

## II. METHOD

Before we describe the implementation of DIIS in the ridge method, let us briefly sketch the original ridge method (the detailed description is given in Ref. 14). At every itera-

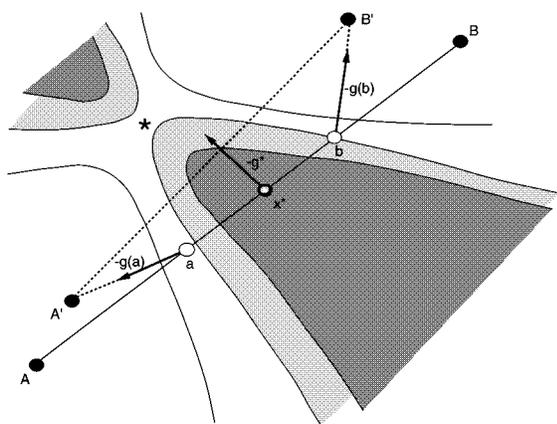


FIG. 1. Schematic illustration of the original ridge method. A few contour lines indicate the PES, where the high energy areas along the ridge separating the reactant  $A$  from the product  $B$  are shaded. The next line to search for a maximum,  $(A', B')$ , is obtained from the line  $(A, B)$  that has a maximum at  $x^*$ .  $g(\cdot)$  denotes a gradient at a corresponding point, and the symbol  $\star$  indicates the TS.

tion  $n$ , we start from two molecular structures  $A_n$  and  $B_n$  that “straddle” the ridge on the PES, so that there is a maximum of energy on the line  $(A_n, B_n)$  at some point  $x_n^*$  (see Fig. 1). ( $A_1$  corresponds to reactants,  $B_1$  to products; the case of possible multiple maxima along  $(A_1, B_1)$  is addressed in Ref. 14). Then the next line to be searched for the energy maximum is constructed as follows. First, we obtain the points  $a_n = (x_n^* - s \cdot p_n)$  and  $b_n = (x_n^* + s \cdot p_n)$ , where  $p_n = (B_n - A_n) / \|B_n - A_n\|$  and  $s$  is a side step size. These points  $a_n, b_n$  also straddle the ridge but  $(a_n, b_n)$  is centered around the top of the ridge. Then we step downhill from  $a_n$  and  $b_n$  to obtain new  $A_{n+1} = a_n + p_n^a$  and  $B_{n+1} = b_n + p_n^b$ , respectively, where the downhill directions  $p_n^a, p_n^b$  is calculated based on the relevant gradients  $g(\cdot)$  either as  $p_n^a = p_n^b = -d \cdot g(x_n^*)$  or  $p_n^a = -d \cdot g(a_n)$ ,  $p_n^b = -d \cdot g(b_n)$ , and the “downhill factor”  $d$  is an adjustable parameter. The latter choice of  $p_n^a, p_n^b$  is twice as computationally expensive, but it allows adjustment of the direction of  $(A_{n+1}, B_{n+1})$  so as to accommodate twisting of the ridge, which is indicated by marked deviation of the newly found  $x_{n+1}^*$  from the middle of the corresponding  $(A_{n+1}, B_{n+1})$ .

It is proven in Ref. 14 that, unless  $g(x_n^*)$  is zero, there are values of  $s$  and  $d$  such that (i) there is a maximum of energy on the new line  $(A_{n+1}, B_{n+1})$ , and (ii) this maximum is less than that for the previous line. To obtain these values of  $s$  and  $d$ , we first try the ones from the previous iteration and use some guesses for the first iteration. If the new line has an energy maximum that is lower than the previous one, we continue the process described above from the points  $A_{n+1}$  and  $B_{n+1}$  unless the maximum component of  $g(x_{n+1}^*)$  is less than a certain criterion, in which case the process is terminated. We increase the downhill factor  $d$  (by a factor of 1.3) if the energy maximum is attained in the middle ( $\pm 10\%$ ) of  $(A_n, B_n)$ . If the new line  $(A_{n+1}, B_{n+1})$  does not have a maximum, or if the new maximum is higher than the previous one, we decrease downhill factor (by half) and try again. When the downhill step size  $d \cdot \|g(x_n^*)\|$  be-

comes too small compared to the side step size  $s$ , we decrease  $s$  (by half) as well.

The approach outlined above is robust, but it is unnecessarily slow near convergence. Implementation of DIIS should remedy this situation, but such an implementation has to be customized to accommodate for the ridge method’s specific features. The presently available combinations of DIIS with geometry optimization (GDIIS)<sup>5–7</sup> involve iterative updating of only one point  $x_n$  that describes the molecular geometry at the  $n$ th iteration. This updating is performed according to the quasi-Newton scheme

$$x_{n+1} = x_n - H_n g_n, \quad (1)$$

where  $g_n$  is the gradient at  $x_n$  and  $H_n$  is an approximate inverse Hessian matrix which is also updated according to some known rule.<sup>15</sup>

For process (1), the DIIS procedure consists of substituting  $x_n$  and  $g_n$  by  $\tilde{x}_n$  and  $\tilde{g}_n$ , respectively, where

$$\tilde{x}_n = \sum_{i=i_0}^n c_i x_i, \quad (2)$$

$$\tilde{g}_n = \sum_{i=i_0}^n c_i g_i, \quad (3)$$

and the coefficients  $c_i$  are obtained from the conditions

$$\left\| \sum_{i=i_0}^n c_i H_n g_i \right\| \rightarrow \min, \quad (4)$$

$$\sum_{i=i_0}^n c_i = 1. \quad (5)$$

Within the quadratic approximation, this scheme corresponds to finding the optimal structure  $\tilde{x}_n$  in the subspace spanned by  $(n - i_0 + 1)$  previous iterates  $x_i$ .

As far as the ridge method is concerned, no single point is sufficient for the advancement of iterations. In order to obtain the line to search at iteration  $(n + 1)$ , one has to know the two points  $a_n$  and  $b_n$  that sandwich the ridge and are centered around its top point  $x_n^*$ , or it is sufficient to know  $x_n^*$  and the direction  $p_n$  across the ridge. We implement DIIS in the ridge method as follows.

(1) We form linear combinations

$$\tilde{x}_n^* = \sum_{i=i_0}^n c_i x_i^*, \quad (6)$$

$$\tilde{g}_n^* = \sum_{i=i_0}^n c_i g(x_i^*), \quad (7)$$

where the coefficients  $c_i$  are obtained from condition (5) and

$$\left\| \sum_{i=i_0}^n c_i g(x_i^*) \right\| \rightarrow \min. \quad (8)$$

We include in Eq. (6) only those  $x_i^*$ ,  $i = i_0, n$ , that were obtained for the same direction  $p_n$ . In this case,  $p_n$  remains the direction along which there is an energy maximum at  $\tilde{x}_n^*$  as indicated by the condition  $p_n^T \tilde{g}_n^* = 0$  (see Eq. (7)). The direction  $p_n$  changes every time the  $x_n^*$  found is too close to

either end point  $A_n$  or  $B_n$  [we use the criterion  $\|x_n^* - \text{End Point}_n\| < (0.25)\|B_n - A_n\|$ ]. Our extensive experience with the original ridge method shows that such a situation happens rather infrequently (in at most 5% of line searches). Therefore, when a change of the direction  $p_n$  occurs, we just discard the previous  $x_i^*$  that enter Eq. (6) and perform an iteration of the original ridge method as described earlier.

(2) We form the next line to search,  $(A_{n+1}, B_{n+1})$ , as

$$A_{n+1} = \tilde{x}_n^* - s \cdot p_n + d \cdot \tilde{g}_n^*, \quad (9)$$

$$B_{n+1} = \tilde{x}_n^* + s \cdot p_n + d \cdot \tilde{g}_n^*. \quad (10)$$

If the search is unsuccessful, i.e., there is no maximum on this line or this maximum is higher than the previous one, we perform the original ridge method's iteration and discard all previous  $x_i^*$  used in the DIIS expansion (6). In order to reduce the CPU time spent on unsuccessful line searches, we do not even attempt searching the line  $(A_{n+1}, B_{n+1})$  obtained according to Eqs. (9) and (10) if the structure  $\tilde{x}_n^*$  differs significantly from the  $x_n^*$ . (The criterion we use is that the difference between corresponding bond lengths of the two structures must not exceed 0.3 Å). This approach also ensures that the initial guess for a wave function is reasonable in cases when such a guess is the last converged wave function.

The ridge-DIIS method described above can be interpreted as finding the point  $\tilde{x}_n^*$  on the current linear approximation to the ridge such that the gradient at  $\tilde{x}_n^*$  is minimal. This is slightly different from the GDIIS concept which is to find a point  $x_n$  in the given linear subspace such that it is the closest to the ultimate solution  $x$ . The former does not require the Hessian matrix, which is especially useful when the TS search is not started in the quadratic region of the TS, because in this case, commonly used updating schemes<sup>15</sup> are not applicable.

The results of testing the ridge-DIIS method described above against the original ridge method and the TS search algorithm from the popular commercial package GAUSSIAN 92<sup>16</sup> are presented in the following section.

### III. RESULTS AND DISCUSSION

In all the examples considered in this section, we are interested mostly in the performance of the methods used for a TS search. Therefore, we employ small basis sets (typically 3-21G<sup>17-19</sup>) and use Hartree-Fock (HF) wave functions to find ground state energies. In order to make the comparison between the ridge method and the GAUSSIAN 92 algorithm meaningful, we use  $x_1^*$  found by the ridge method as a starting structure for GAUSSIAN 92.

We judge the methods' performance based on the number of energy, gradient, and Hessian matrix evaluations required to find the same TS according to the same criterion for convergence. In order to put these numbers on a common footing, we calculate the "effective number of energy evaluations,"  $N_{\text{eff}}$ , according to the formula  $N_{\text{eff}} = N_e + 2N_g + 3N_{\text{at}}N_H$ , where  $N_e$ ,  $N_g$ , and  $N_H$  are the numbers of energy, gradient, and analytical Hessian matrix

evaluations, respectively, and  $N_{\text{at}}$  is the number of atoms. Obviously, the implicit assumption here is that at the HF level, the CPU time per gradient should be roughly twice the CPU time to compute an energy. Also, since the analytical evaluation of second derivatives involves construction of  $3N_{\text{at}}$  Fock matrices,<sup>20</sup> we estimate the CPU time per analytical Hessian matrix accordingly. It should be noted that for correlated wave functions calculation of the analytical Hessian is even more expensive. We do not compare the CPU times corresponding to different methods directly because these times are affected not only by the optimization technique, but also by the efficiency of the energy and gradient evaluations, which is a separate issue.

In order to have a fair comparison, we also use the same choice of coordinates (Cartesian) in both methods. It seems logical to expect both methods to benefit from "natural internal coordinates"<sup>7,21,22</sup> roughly similarly, so that the relative performance of these methods should remain the same.

Results of the TS searches described below are presented in Table I. The first example involves a search for a TS for the isomerization process



considered at the level HF/3-21G. It is worth noting that, although it is rather clear what the TS structure should look like in this case, its determination via searching the line between reactant and product involves dealing with at least two maxima. The reason for this is the presence of too short a distance between either Si atom and the "traveling" H atom during the linear transformation of  $\text{H}_2\text{Si}=\text{SiH}_2$  into  $\text{H}_3\text{Si}-\text{SiH}$ . This situation was successfully handled by the Detour algorithm that is a part of the ridge method (see Ref. 14 for details).

As shown in Table I, the speedup resulting from employing DIIS in the ridge method is a factor of two. However, GAUSSIAN 92 is still 70% faster, which is not a surprise since the case under study is well suited for Hessian-based optimization techniques. (By this we mean that the PES corresponding to reaction (11) does not have a sharp barrier, and that the potential wells corresponding to reactant and product are reasonably deep). The reason we present these data is to show that even for cases where the Hessian-based algorithm is preferable, the computational cost of the ridge-DIIS method is not as high as it may seem from the method description. When one recalls that there is no need for the Hessian matrix in this method, the 70% increase in  $N_{\text{eff}}$  does not look that bad.

The next example involves the seemingly simple exchange reaction



that we treat at the ROHF level with the correlation-consistent basis set due to Dunning.<sup>23</sup> The corresponding PES has a rather sharp and curved ridge (see Fig. 2), which has several implications. First, the improvement of the ridge method due to DIIS is only a factor of 1.5 (based on the corresponding  $N_{\text{eff}}$ ), as follows from Table I. Second, the number of energy evaluations per line search is rather large (~7 versus a typical number of 4-5), indicating that the en-

TABLE I. Comparative performance of the ridge method (with and without DIIS) and the TS search algorithm of GAUSSIAN 92.

Case	Method	Iterations	$N_e^b$	$N_g^c$	$N_{\text{eff}}^d$
Si <sub>2</sub> H <sub>4</sub>	original ridge	51	226	53	332
	ridge-DIIS( $N_{\text{store}}=1$ ) <sup>a</sup>	23	120	26	172
	ridge-DIIS( $N_{\text{store}}=2$ )	28	146	24	194
	GAUSSIAN 92, anal. Hessian	27	28	28	102
F+H <sub>2</sub>	original ridge	33	230	36	302
	ridge-DIIS( $N_{\text{store}}=1$ )	24	150	24	198
	ridge-DIIS( $N_{\text{store}}=2$ )	24	150	24	198
	GAUSSIAN 92, num. Hessian	70	74	74	222
Na <sub>6</sub>	original ridge	64	351	68	487
	ridge-DIIS( $N_{\text{store}}=1$ )	37	219	35	289
	ridge-DIIS( $N_{\text{store}}=2$ )	42	238	41	320
	GAUSSIAN 92, num. Hessian	failed	-	-	-

<sup>a</sup> $N_{\text{store}}$ =number of previous iterations used in DIIS.

<sup>b</sup> $N_e$ =number of energy evaluations.

<sup>c</sup> $N_g$ =number of gradient evaluations.

<sup>d</sup> $N_{\text{eff}}$ =number of effective energy evaluations;  $N_{\text{eff}}=N_e+2N_g+3N_{\text{at}}N_H$ , where  $N_H$  is the number of analytical Hessian evaluations and  $N_{\text{at}}$  is the number of atoms.

ergy across the ridge is poorly approximated by a parabola even in a relatively small PES region (this is also obvious from Fig. 2). Third, even though the Hessian-based algorithm used in GAUSSIAN 92 starts within a PES region where the Hessian matrix has only one negative eigenvalue, it experiences a rather bumpy ride, failing to converge in 100 iterations. Only after turning off the default updating of the trust radius did it converge in 70 iterations. Note that the initial Hessian matrix in this case was calculated numerically, since there are no analytical second derivatives available for ROHF in GAUSSIAN 92. Finally, the efficiency of the ridge-DIIS method in this case is slightly better than that of GAUSSIAN 92, as one can see from Table I.

The last example illustrates a qualitative difference between ridge-DIIS and the Hessian-based algorithm used in Gaussian 92. We consider a transition between two minima on a PES obtained at the HF level with the basis set and effective core potential due to Melius and Goddard<sup>24</sup> for a Na<sub>6</sub> cluster (see Fig. 3). Since these minima are very shallow, which is typical for PESs of metal clusters, a tight con-

vergence criterion is needed (we found  $10^{-4}$  a.u. for the maximum component of the gradient to be sufficient).

From Table I, we see that DIIS again results in a speedup of the original ridge method by a factor of two. By contrast, the Hessian-based algorithm of GAUSSIAN 92 failed to locate a TS. Thus, we see that the ridge method is indeed robust and is able to handle flat PESs that present a major challenge for Hessian-based techniques.

The data in Table I also demonstrate that by using information from only one previous iteration we recover practically all the speedup that one can get from DIIS. This is due to the fact that large values of  $N_{\text{store}}$  are counterproductive at early iterations where steps taken by the algorithm are large and the region of the PES containing all  $x_i^*$ ,  $i=i_0, \dots, n$  cannot be represented well by a quadratic approximation. On the other hand, within a quadratic region of a TS, large  $N_{\text{store}}$  should be better, and it should somewhat compensate for the lag at early iterations. Also, the actual number of previous iterations used in Eqs. (6)-(8) will be relatively low on a

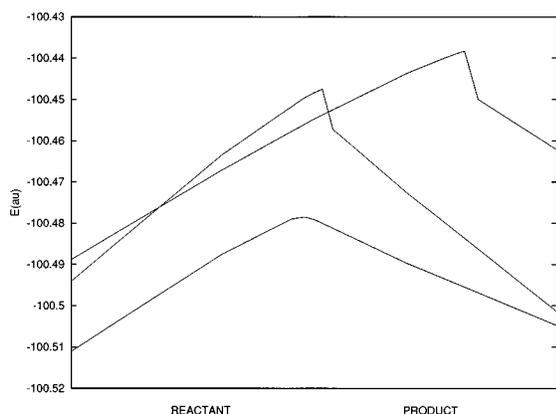
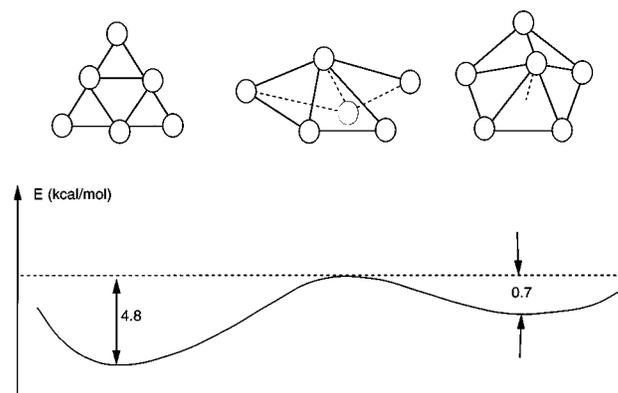


FIG. 2. Energy across the ridge on the first three iterations for reaction (12).

FIG. 3. Two minima (the planar, near-triangular structure and the pentagonal pyramid) and the butterfly-like TS for the Na<sub>6</sub> cluster.

nonquadratic PES due to the discarding of all previous  $x_i^*$  whenever there is a marked change in the direction of the ridge or an unsuccessful line search.

#### IV. CONCLUSION

We have applied the convergence accelerating procedure DIIS<sup>4</sup> to the ridge method<sup>14</sup> for finding transition states. The resulting algorithm, ridge-DIIS, is a factor of two faster than the original ridge method, while all the advantages of the original method are preserved. ridge-DIIS does not require a Hessian matrix or an initial guess for a TS geometry, and it can efficiently handle such challenging problems as finding a TS on a very flat or twisted PES.

Considerably more robust than Hessian-based second order algorithms, ridge-DIIS is not as computationally expensive as other reliable techniques for TS searches.<sup>8-11</sup> In fact, its computational cost is either comparable to that of GAUSSIAN 92 on PESs with sharp ridges or it is slower at most by a factor of two on PESs that are well suited for Hessian-based techniques. However, this additional cost is justified by the ability of ridge-DIIS to find those TSs that are not amenable to Hessian-based methods.

Finally, we note that as one treats larger and larger molecules, the attractiveness of ridge-DIIS grows significantly since evaluation of the Hessian either numerically or analytically quickly becomes prohibitive. Thus, we expect ridge-DIIS to be the method of choice for at least three classes of systems: (i) PESs with sharp barriers, (ii) very flat PESs, and (iii) large molecules.

#### ACKNOWLEDGMENTS

We are grateful to the Office of Naval Research for primary support of this work. E. A. C. also acknowledges sup-

port from the Camille and Henry Dreyfus Foundation and the Alfred P. Sloan Foundation via a Dreyfus Teacher-Scholar Award and a Sloan Research Fellowship, respectively.

- <sup>1</sup>H. B. Schlegel, *J. Comp. Chem.* **3**, 214 (1982).
- <sup>2</sup>J. E. Sinclair and R. Fletcher, *J. Phys. C* **7**, 864 (1974).
- <sup>3</sup>H. B. Schlegel, *Ab Initio Methods in Quantum Chemistry—I* (Wiley, New York, 1987).
- <sup>4</sup>P. Pulay, *Chem. Phys. Lett.* **73**, 393 (1980).
- <sup>5</sup>P. Császár and P. Pulay, *J. Mol. Struct.* **114**, 31 (1984).
- <sup>6</sup>T. H. Fischer and J. Almlöf, *J. Phys. Chem.* **96**, 9768 (1992).
- <sup>7</sup>G. Fogarasi, X. Zhou, P. W. Taylor, and P. Pulay, *J. Am. Chem. Soc.* **114**, 8191 (1992).
- <sup>8</sup>C. J. Cerjan and W. H. Miller, *J. Chem. Phys.* **75**, 2800 (1981).
- <sup>9</sup>S. Fischer and M. Karplus, *Chem. Phys. Lett.* **194**, 252 (1992).
- <sup>10</sup>A. Bannerjee, N. Adams, J. Simons, and R. Shepard, *Chem. Phys. Lett.* **89**, 52 (1985).
- <sup>11</sup>J. Simons, P. Jørgensen, H. Taylor, and J. Ozment, *J. Phys. Chem.* **87**, 2745 (1983).
- <sup>12</sup>T. A. Halgren and W. N. Lipscomb, *Chem. Phys. Lett.* **49**, 225 (1977).
- <sup>13</sup>S. Bell and J. S. Crighton, *J. Chem. Phys.* **80**, 2464 (1984).
- <sup>14</sup>I. V. Ionova and E. A. Carter, *J. Chem. Phys.* **98**, 6386 (1993).
- <sup>15</sup>R. Fletcher, *Practical Methods of Optimization* (Wiley, New York, 1987).
- <sup>16</sup>M. J. Frisch *et al.*, GAUSSIAN 92, Gaussian, Inc., Pittsburgh, PA, 1992.
- <sup>17</sup>J. S. Binkley, J. A. Pople, and W. J. Hehre, *J. Am. Chem. Soc.* **102**, 939 (1980).
- <sup>18</sup>M. S. Gordon, J. S. Binkley, J. A. Pople, W. J. Pietro, and W. J. Hehre, *J. Am. Chem. Soc.* **104**, 2797 (1982).
- <sup>19</sup>W. J. Pietro, M. M. Francl, W. J. Hehre, D. J. DeFrees, J. A. Pople, and J. S. Binkley, *J. Am. Chem. Soc.* **104**, 5039 (1982).
- <sup>20</sup>M. J. Frisch, M. Head-Gordon, and J. A. Pople, *Chem. Phys.* **141**, 189 (1991).
- <sup>21</sup>P. Pulay and G. Fogarasi, *J. Chem. Phys.* **96**, 2856 (1992).
- <sup>22</sup>J. Baker, *J. Comp. Chem.* **14**, 1085 (1993).
- <sup>23</sup>T. H. Dunning, Jr., *J. Chem. Phys.* **90**, 1007 (1989).
- <sup>24</sup>C. F. Melius and W. A. Goddard III, *Phys. Rev. A* **10**, 1528 (1974).