

## LETTERS

### Integrated Molecular Orbital Method with Harmonic Cap for Molecular Forces and Its Application to Geometry Optimization and the Calculation of Vibrational Frequencies

José C. Corchado and Donald G. Truhlar\*

*Department of Chemistry and Supercomputer Institute, University of Minnesota,  
Minneapolis, Minnesota 55455-0431*

*Received: October 31, 1997*

When applied to large systems, the integrated-levels strategy has been demonstrated to be a good approximation for the calculation of energy differences with an accuracy comparable to the accuracy that would be obtained by a single-level calculation at the higher level with only a part of its computational expense. In the present paper we introduce a modification to the original integrated method, namely, an additive harmonic term that allows a very simple and general extension of the integrated methods to the calculation of forces (gradients), optimization of geometries, and calculation of harmonic vibrational frequencies. The new method, called integrated molecular orbital method with harmonic cap (IMOHC), has been successfully tested by calculating the C–H bond distances, vibrational frequencies, and bond energies in the ethane molecule. The method should serve as an efficient method for creating dual-level methods of the QM:MM and IMOMO variety, and the forces can be used to drive classical (Car–Parrinello) or semiclassical (variational transition-state theory with semiclassical tunneling) direct dynamics calculations.

#### Introduction

In the quest for practical yet accurate quantum mechanical (QM) methods of electronic structure calculations for larger and larger systems, dual-level and multiple-level methods are arising as the logical consequence of a large number of independent lines of development. Recent examples include double-slash geometry optimization,<sup>1</sup> SEC,<sup>2</sup> SAC,<sup>3</sup> G2,<sup>4</sup> VTST-IC,<sup>5</sup> QM:MM,<sup>6</sup> “integrated” methods (IMOMM,<sup>7</sup> IMOMO,<sup>8</sup> CCSS,<sup>9</sup> and ONIOM<sup>10</sup>), IRCMax,<sup>11</sup> dual-level Shepard interpolation,<sup>12</sup> and the combination of ab initio gas-phase energies with semi-empirical solvation energies.<sup>13</sup> The QM:MM<sup>6</sup> and integrated methods<sup>7–10</sup> share a common goal and a common challenge. The goal is to use a higher-level method for a subsystem of a large system, while treating the rest of the system at a lower level. The challenge is devising a method to “link” the large system to the small system. Although there has been considerable experimentation and debate, no method has emerged as totally satisfactory, especially when geometry optimization is involved.

In the present Letter we present a new approach to geometry optimization in integrated calculations, which we believe can have widespread utility for QM:MM<sup>6,7</sup> and dual-level-QM<sup>8–10</sup> methods. We call the method integrated molecular orbital method with harmonic cap (IMOHC) because the essential element is a harmonic cap on the subsystem. Following Morokuma and co-workers,<sup>8,10</sup> the acronym MO stands for both correlated and uncorrelated QM methods, which are both based on molecular orbitals. The IMOHC method will be illustrated for both geometry optimization and the calculation of vibrational frequencies.

#### Theory

In an integrated calculation, the energy of a large system is approximated by

$$E_{\text{ES}}(\text{I}) = E_{\text{CSS}}(\text{HL}) - E_{\text{CSS}}(\text{LL}) + E_{\text{ES}}(\text{LL}) \quad (1)$$

where I denotes integrated, HL indicates the higher level of

calculation, LL is the lower level, ES is the entire (large) system, and CSS is the capped subsystem. As an example, we can consider the ethane molecule, CH<sub>3</sub>CH<sub>3</sub>, as the entire system and the methane molecule, CH<sub>4</sub>, as the capped system obtained by substituting an H atom for a CH<sub>3</sub> group. The energy will be given by the previously defined combination of HL and LL energies calculated for a given fixed geometrical conformation of the CH<sub>4</sub> molecule and an LL energy for a given fixed geometry for the ethane molecule.<sup>7-10</sup>

The optimization of the geometry is more complicated, since one substituent of the primary carbon atom is a hydrogen atom in the capped system but a carbon atom in the entire system. Call this substituent X. The partial derivative of the energy with respect to a Cartesian coordinate  $x$  is given by

$$\frac{\partial E_{\text{ES}}(\text{I})}{\partial x} = \frac{\partial E_{\text{CSS}}(\text{HL})}{\partial x} - \frac{\partial E_{\text{CSS}}(\text{LL})}{\partial x} + \frac{\partial E_{\text{ES}}(\text{LL})}{\partial x} \quad (2)$$

The gradient whose components are given by eq 2 will not be zero at any physical geometry of the system, since the distance C–X will tend to the bond length of a C–C bond in one system and to the bond length of a C–H bond in the other system. Therefore, we do not require the C–X distance to be the same in the ES and CSS. Instead we increase the number of independent variables (and hence the length of the gradient vector) from 24 to 27. Geometry optimization is now carried out in a space of 27 Cartesian coordinates, 24 corresponding to the 3-Cartesian coordinates of the 8 atoms in ethane, and 3 describing the position of the extra H that completes the methane. The last term in eqs 1 and 2 is independent of the latter 3 coordinates, and the 2 first terms in eqs 1 and 2 are independent of the 12 coordinates of the secondary CH<sub>3</sub> group, but the left-hand side of eq 2 is a legitimate analytical function of all 27 variables, which are taken as independent variables with no constraints between them. The total elimination of relationships between the large and small systems is a key attractive feature of the new method, and it greatly simplifies the so-called “link atom” problem. If we were to apply the method in this starkly simple form, though, it would fail for reasons described next.

The individual terms of the integrated method may be grouped in two intuitive ways:

$$E_{\text{ES}}(\text{I}) = E_{\text{CSS}}(\text{HL}) + [E_{\text{ES}}(\text{LL}) - E_{\text{CSS}}(\text{LL})] \quad (3)$$

or

$$E_{\text{ES}}(\text{I}) = E_{\text{ES}}(\text{LL}) + [E_{\text{CSS}}(\text{HL}) - E_{\text{CSS}}(\text{LL})] \quad (4)$$

Equation 3 considers the integrated calculation as a lower-level inclusion of global substituent effects on a higher-level calculation for a subsystem. Equation 4 considers the integrated calculation as the inclusion of higher-level local effects on a lower-level calculation for the entire system. It is the latter viewpoint that provides the key to the new method. In eq 4 the first term is totally independent of the location of the coordinates of the capping H atom. However, since the second term involves a difference of energies, a minimization according to the scheme described in the previous paragraph will locate the capping H at the geometry that minimizes the energy difference of the two levels rather than at a physical bonding geometry. For example, if both levels are realistic quantum mechanical electronic structure levels, the geometry that minimizes the difference will probably involve dissociating the capping atom. But this is very unsatisfactory; the whole idea

of the capping atom is that it should mimic the bonding properties of the larger substituent it displaces. (The C–H bond in the primary CH<sub>3</sub> group should involve sp<sup>3</sup> carbon, not sp<sup>2</sup> carbon.) The problem having been diagnosed, a viable strategy becomes clear. We write

$$E_{\text{ES}}(\text{I}) = E_{\text{CSS}}(\text{HL}) - E_{\text{CSS}}(\text{LL}) + E_{\text{ES}}(\text{LL}) + \frac{1}{2}k(R_{\text{C-H}} - R_{\text{eq}})^2 \quad (5)$$

and

$$\frac{\partial E_{\text{ES}}(\text{I})}{\partial x} = \frac{\partial E_{\text{CSS}}(\text{HL})}{\partial x} - \frac{\partial E_{\text{CSS}}(\text{LL})}{\partial x} + \frac{\partial E_{\text{ES}}(\text{LL})}{\partial x} + \frac{1}{2}k \frac{\partial [(R_{\text{C-H}} - R_{\text{eq}})^2]}{\partial x} \quad (6)$$

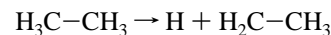
where we have included a harmonic term to maintain this atom at a physical value of the bond distance in the capped subsystem.

The calculation of the constant  $k$  and the equilibrium distance  $R_{\text{eq}}$  is carried out in such a way that does not involve an excessive computational cost. This is accomplished by three single-point energy calculations at the high level. The first one is a single-point calculation based on the geometry of the equilibrium conformation of methane as predicted by the LL. The other two calculations are based on geometries obtained by slightly elongating and contracting the capping C–H distance in the capped subsystem. The three single-point energies are fitted to a parabola as a function of this C–H distance, providing us with an approximate HL force constant ( $k$ ) and equilibrium distance ( $R_{\text{eq}}$ ) for this bond.

In the general case one would also add a harmonic term for bending and torsional motion; such terms were not needed for the present example.

### Example

To test the ability of this method to predict the geometry, frequencies, and relative energies with a quality comparable to that which we would obtain with a complete HL calculation, we optimized the geometry and calculated the frequencies and the energies for two systems, ethane and ethyl radical. We also calculated the dissociation energy for breaking a C–H bond in ethane



To test the sensitivity of the results to the value of  $k$ , we calculated it in two ways. In the first way we used the reactant CSS, H<sub>3</sub>C–H, and in the second we used the product CSS, H<sub>2</sub>C–H. These values of  $k$  are denoted  $k_1$  and  $k_2$ , respectively. Each value of  $k$  has its corresponding value of  $R_{\text{eq}}$ . The calculation of the  $k$  and  $R_{\text{eq}}$  parameters was made by distorting the optimized LL predicted geometry by 0.001 Å in each direction (one C–H bond only as explained above) and fitting to a parabola.<sup>14</sup>

To provide a challenging test case, the levels of calculation were chosen so that the C–H bond distances are very different at the two levels; thus, we will be able to check if our scheme provides us with a geometry noticeably closer to the HL geometry than would be provided by an LL geometry optimization. Thus, the HF/3-21G level<sup>1</sup> was chosen as LL, while the MP2/6-31G level<sup>1</sup> was selected as HL.

In Table 1 we compare the predicted C–H bond distances in ethane at both levels of calculation, as well as the optimized

**TABLE 1: C–H Bond Distance (in Å) in Ethane As Predicted by Single-Level Methods and IMOHC Calculations**

level	$R_{C-H}(H-CH_3)$	$R_{C-H}(H-CH_2CH_3)$
HF/3-21G	1.0830	1.0841
MP2/6-31G	1.0959	1.0988
MP2/6-31G:HF/3-21G ( $k_1$ )		1.0973
MP2/6-31G:HF/3-21G ( $k_2$ )		1.0972

**TABLE 2: C–H Bond Distance (in Å) in Ethyl As Predicted by Single-Level Methods and by the IMOHC Method**

level	$R_{C-H}(H-CH_2)$	$R_{C-H}(H-CH_2CH_3)$
HF/3-21G	1.0717	1.0734
MP2/6-31G	1.0830	1.0867
MP2/6-31G:HF/3-21G ( $k_1$ )		1.0854
MP2/6-31G:HF/3-21G ( $k_2$ )		1.0854

**TABLE 3: Frequencies (in  $cm^{-1}$ ) and ZPE (in kcal/mol) for Ethane As Predicted by Single-Level Methods and IMOHC Calculations**

HF/3-21G	MP2/6-31G	MP2/6-31G: HF/3-21G ( $k_1$ )	MP2/6-31G: HF/3-21G ( $k_2$ )
$\omega_m$			
3268	3155	3257	3257
3268	3155	3257	3257
3241	3132	3199	3199
3241	3132	3160	3160
3201	3064	3160	3160
3197	3063	3083	3083
1678	1580	1677	1677
1678	1580	1677	1677
1677	1575	1576	1576
1677	1575	1569	1569
1580	1496	1569	1569
1572	1493	1531	1534
1352	1290	1330	1330
1352	1290	1330	1330
1005	1022	1085	1091
921	863	893	893
921	863	893	893
313	327	314	313
ZPE			
50.24	48.11	49.40	49.42

IMOHC bond distance. The same information appears in Table 2 for the ethyl radical.

While the HF/3-21G calculation predicts a bond length too short by 0.0147 and 0.0133 Å as compared to the MP2/6-31G calculation, the IMOHC method predicts a geometry that deviates from the MP2/6-31G calculation by only 0.0015–0.0016 Å for the ethane molecule and 0.0013 Å for the ethyl radical; i.e., it reduces the error by a factor of 10. The choice between  $k_1$  or  $k_2$  has a negligible effect on the results. Notice that if we take the higher-level calculation on the entire system as our standard of accuracy and measure the errors in units of  $10^{-4}$  Å, we have errors of 133–147 in the lower-level calculations on the entire system and 29–37 in higher-level calculations on the model system (capped subsystem), but only 13–16 in the integrated calculations. Thus, combining the HL-CSS and LL-ES results by the strategy presented here gives better results than using either component-type calculation separately.

Once we had the geometry of both systems optimized, we calculated the  $27 \times 27$  Hessian matrix and performed a normal-mode analysis using the  $24 \times 24$  submatrix. The frequencies and zero-point energy (ZPE) of ethane at the three levels of calculation appear in Table 3, while Table 4 lists the results for the ethyl radical. These tables show that the IMOHC methods

**TABLE 4: Frequencies (in  $cm^{-1}$ ) and ZPE (in kcal/mol) for Ethyl As Predicted by Single-Level Methods and IMOHC Calculations**

HF/3-21G	MP2/6-31G	MP2/6-31G: HF/3-21G ( $k_1$ )	MP2/6-31G: HF/3-21G ( $k_2$ )
$\omega_m$			
3396	3310	3329	3330
3292	3200	3252	3251
3252	3142	3221	3221
3220	3100	3220	3220
3153	3023	3153	3153
1665	1570	1664	1664
1660	1558	1660	1660
1597	1546	1576	1577
1574	1493	1567	1569
1319	1263	1310	1310
1110	1073	1136	1141
1046	1050	1104	1105
894	855	878	878
439	479	487	487
129	147	129	129
ZPE			
39.67	38.34	39.58	39.59

**TABLE 5: Dissociation Energy (kcal/mol) of a C–H Bond**

molecule	level	$\Delta E^a$	$\Delta H_0^b$
CH <sub>4</sub>	MP2/6-31G	99.66	89.90
C <sub>2</sub> H <sub>6</sub>	HF/3-21G	84.15	73.58
C <sub>2</sub> H <sub>6</sub>	MP2/6-31G	97.02	87.25
C <sub>2</sub> H <sub>6</sub>	MP2/6-31G:HF/3-21G ( $k_1$ ) <sup>b</sup>	97.09	87.27
C <sub>2</sub> H <sub>6</sub>	MP2/6-31G:HF/3-21G ( $k_2$ ) <sup>b</sup>	97.08	87.25
C <sub>2</sub> H <sub>6</sub>	MP2/6-31G:HF/3-21G ( $k_1, k_2$ ) <sup>c</sup>	97.22	87.41

<sup>a</sup>  $\Delta E$  is the Born–Oppenheimer classical energy of dissociation; i.e., it excludes zero-point energy;  $\Delta H_0^b$  is the standard-state enthalpy of dissociation at 0 K, including zero-point energy. <sup>b</sup> Calculated using  $k_1$  or  $k_2$  for both ethane and ethyl. <sup>c</sup> Calculated using  $k_1$  for the ethane molecule and  $k_2$  for the ethyl radical.

give dramatic improvement in the frequencies of the subsystem. The frequencies of the primary subsystem have the accuracy of the higher level, while the substituent that is treated only at lower level retains the lower-level accuracy with no degradation by the cap. As an example, we can examine the set of four bending frequencies of ethane with values between 1678 and 1677  $cm^{-1}$  at the HF/3-21G level and between 1580 and 1575  $cm^{-1}$  at the MP2/6-31G level. The IMOHC method yields two frequencies at 1677  $cm^{-1}$  and two frequencies at 1569  $cm^{-1}$ , values that are very close to the HF/3-21G and MP2/6-31G frequencies, respectively. Similarly, the lowest frequency in ethane, which is the frequency of the torsion that occurs only in the full system, is almost unchanged from the lower-level entire system result by the IMOHC calculation.

Finally, in Table 5 we list the results of the calculations of bond energies with and without ZPE contribution. All calculations in this table are based on geometries and frequencies calculated by the method described above. The agreement between the MP2/6-31G calculation and the IMOHC calculation is excellent whether using the same constant  $k$  for both ethane and ethyl or using the constants  $k$  and  $R_{eq}$  calculated specifically for each system ( $k_1$  for ethane and  $k_2$  for ethyl), although the former choice seems to give ever so slightly better results.

One could probably use even more approximate or standard values for  $k$  and  $R_{eq}$ , but we leave that for future work.

## Conclusion

We have presented and tested an extension to the original integrated dual-level methods that appears to be a promising

tool for calculations of gradients, Hessians, and optimized geometries of large molecules. The extension is simple and does not introduce any significant extra computational cost into the original IMOMM, IMOMO, and CCSS methods. It is based on a harmonic restoring potential for the capping atom in the capped subsystem. The computational cost for an integrated geometry optimization of the entire system is essentially the same as for a higher-level geometry optimization on the capped subsystem and a lower-level geometry optimization on the entire system, plus three single-point energy calculations for the capped subsystem at the higher level. Furthermore, since it seems unnecessary to calculate a set of  $k$  and  $R_{\text{eq}}$  constants for each case, these latter three calculations need not be repeated for all cases, and they represent a negligible increase in the cost of the calculation.

When applied to the calculation of geometries, frequencies, and energies of the ethane molecule and ethyl radical, the method gives rise to improved geometries and frequencies for the critical part of the molecules under study, as well as C–H bond energies very similar those obtained by a higher-level calculation in the whole system, but with a much lower cost. The computational advantages of this method are expected to be more noticeable as the size of the system increases, allowing accurate dual-level calculations in systems for which similarly accurate single-level calculations are not affordable at present or in the near future.

A special advantage of eqs 5 and 6 is that they allow for a very modular approach in which totally separate calls to HL and LL subprograms (or electronic structure<sup>15</sup> or molecular mechanics<sup>16,17</sup> packages) may be made and the results combined ex post facto by eqs 5 and 6 at each step of the geometry optimization. The gradients of eq 6 can also be formed at every step of a dynamics calculation and used (since the force is minus the gradient) for dual-level direct dynamics calculation by the Car–Parrinello method<sup>18</sup> or for direct dynamics calculations including quantum effects by semiclassical methods.<sup>19,20</sup> Dual-level direct dynamics calculations in which the bond-breaking and bond-making subsystems are treated at a higher level than the rest of the system will allow for much greater accuracy than if the entire system must be treated at the higher level.

**Acknowledgment.** We are very grateful to Professor E. L. Coitiño for many helpful discussions. J.C.C. acknowledges the Spanish Ministerio de Educación y Cultura and the Fulbright Commission for a postdoctoral scholarship. The work on molecular structure is supported in part by NSF Grant No. CHE94-23927, and the work on GAUSSRATE is supported by part by the U.S. Department of Energy, Office of Basic Energy Sciences, Grant No. DE-FG02-86ER13579.

## References and Notes

- (1) Bartlett, R. J., Ed. *Comparison of Ab Initio Quantum Chemistry with Experiment for Small Molecules*; Reidel: Dordrecht, 1985. Hehre, W.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1986. Langhoff, S. R., Ed. *Quantum Mechanical Electronic Structure Calculations with Chemical Accuracy*; Kluwer: Dordrecht, 1995.
- (2) Brown, F. B.; Truhlar, D. G. *Chem. Phys. Lett.* **1985**, *117*, 307.
- (3) Gordon, M. S.; Truhlar, D. G. *J. Am. Chem. Soc.* **1986**, *108*, 5412.
- (4) Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; Pople, J. A. *J. Chem. Phys.* **1991**, *94*, 7221.
- (5) Hu, W.-P.; Liu, Y.-P.; Truhlar, D. G. *J. Chem. Soc., Faraday Trans.* **1994**, *90*, 1715.
- (6) (a) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700. (b) Gao, J. *Rev. Comput. Chem.* **1996**, *7*, 119. (c) Thiel, W. *Adv. Chem. Phys.* **1996**, *93*, 703.
- (7) Maseras, F.; Morokuma, K. *J. Comput. Chem.* **1995**, *16*, 1170.
- (8) Humbel S.; Sieber S.; Morokuma K. *J. Chem. Phys.* **1996**, *105*, 1959.
- (9) (a) Coitiño, E. L.; Truhlar, D. G.; Morokuma, K. *Chem. Phys. Lett.* **1996**, *259*, 159. (b) Noland, M.; Coitiño, E. L.; Truhlar, D. G. *J. Phys. Chem. A* **1997**, *101*, 1193. (c) Coitiño, E. L.; Truhlar, D. G. *J. Phys. Chem. A* **1997**, *101*, 4641.
- (10) Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber S.; Morokuma, K. *J. Phys. Chem.* **1996**, *100*, 19375.
- (11) Malick, D. K.; Petersson, G. A.; Montgomery, J. A., Jr. To be published.
- (12) Nguyen, K. A.; Rossi, I.; Truhlar, D. G. *J. Chem. Phys.* **1995**, *103*, 5522.
- (13) Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1993**, *115*, 8810.
- (14) The calculated values of  $k$  and  $R_{\text{eq}}$  are (in hartrees/ $a_0^2$  and  $a_0$ , respectively):  $k_1 = 0.191\ 02$ ,  $R_{\text{eq},1} = 2.0694$ ;  $k_2 = 0.20816$ ,  $R_{\text{eq},2} = 2.0461$ .
- (15) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Ciolowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. *GAUSSIAN 94*; Gaussian Inc.: Pittsburgh, 1995.
- (16) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T.E., III; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Comput. Chem. Commun.* **1995**, *91*, 1.
- (17) Brooks, B. R.; Cruccoleri, R. E.; Olafson, B. D.; States, K. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
- (18) (a) Carr, R.; Parrinello, M. *Phys. Rev. Lett.* **1985**, *55*, 2471. (b) Parrinello, M. In *Modern Techniques in Computational Chemistry*; Clementi, E., Ed.; ESCOM: Leiden, 1991; p 833. (c) Galli, G.; Parrinello, M. In *Computer Simulations in Material Science*; Kluwer: Dordrecht, 1991; p 283.
- (19) (a) Truhlar, D. G.; Gordon M. S. *Science* **1990**, *249*, 491. (b) González-Lafont, A.; Truong, T. N.; Truhlar, D. G. *J. Phys. Chem.* **1991**, *95*, 4618. (c) Truhlar, D. G. In *The Reaction Path in Chemistry*; Heidrich, D., Ed.; Kluwer: Dordrecht, 1995; p 229. (d) Espinosa-García, J.; Corchado, J. C.; Truhlar, D. G. *J. Am. Chem. Soc.* **1997**, *119*, 9891.
- (20) In fact, the present calculations were carried out with a modified version of the modular GAUSSRATE code, <http://comp.chem.umn.edu/gaussrate>.