Efficient Molecular Dynamics Simulations of Multiple Radical Center Systems Based on the Fragment Molecular Orbital Method

Hiroya Nakata
Tokyo Institute of Technology

Michael Schmidt
Iowa State University, mike@si.fi.ameslab.gov

Dmitri G. Fedorov
National Institute of Advanced Industrial Science and Technology

Kazuo Kitaura
Kobe University

Shinichiro Nakamura
RIKEN

Follow this and additional works at: http://lib.dr.iastate.edu/chem_pubs

Part of the Chemistry Commons

The complete bibliographic information for this item can be found at http://lib.dr.iastate.edu/chem_pubs/590. For information on how to cite this item, please visit http://lib.dr.iastate.edu/howtocite.html.
Authors
Hiroya Nakata, Michael Schmidt, Dmitri G. Fedorov, Kazuo Kitaura, Shinichiro Nakamura, and Mark S. Gordon
Efficient Molecular Dynamics Simulations of Multiple Radical Center Systems Based on the Fragment Molecular Orbital Method

Hiroya Nakata,†‡§ Michael W. Schmidt,∥ Dmitri G. Fedorov,*†−‡ Kazuo Kitaura,‡ Shinichiro Nakamura,‡ and Mark S. Gordon*†∥

†Department of Biomolecular Engineering, Tokyo Institute of Technology, 4259 Nagatsutacho, Midori-ku, Yokohama, Kanagawa 226-8501, Japan
‡Research Cluster for Innovation, Nakamura Lab, RIKEN, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan
§Kojimachi Business Center Building, Japan Society for the Promotion of Science, 5-3-1 Kojimachi, Chiyoda-ku, Tokyo 102-0083, Japan
∥Department of Chemistry and Ames Laboratory, Iowa State University, Ames, Iowa 50011, United States
⊥NRI, National Institute of Advanced Industrial Science and Technology (AIST), 1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan
#Graduate School of System Informatics, Kobe University, 1-1 Rokkodai-cho, Nada-ku, Kobe 657-8501, Japan

Supporting Information

ABSTRACT: The fully analytic energy gradient has been developed and implemented for the restricted open-shell Hartree–Fock (ROHF) method based on the fragment molecular orbital (FMO) theory for systems that have multiple open-shell molecules. The accuracy of the analytic ROHF energy gradient is compared with the corresponding numerical gradient, illustrating the accuracy of the analytic gradient. The ROHF analytic gradient is used to perform molecular dynamics simulations of an unusual open-shell system, liquid oxygen, and mixtures of oxygen and nitrogen. These molecular dynamics simulations provide some insight about how triplet oxygen molecules interact with each other. Timings reveal that the method can calculate the energy gradient for a system containing 4000 atoms in only 6 h. Therefore, it is concluded that the FMO-ROHF method will be useful for investigating systems with multiple open shells.

1. INTRODUCTION

Molecular dynamics (MD) simulations, especially using classical molecular mechanics (MM) force fields, and geometry optimizations have become indispensable tools in chemistry, physics, biology, and material science. The applications of MM range from drug design to transition metal catalysis. Especially in the biological field, classical MD using, for example, the GROMACS,1 CHARMM,2 AMBER,3 Tinker,4 and AMOEBA5 force fields have been developed and successfully applied to many interesting phenomena.

Chemical reactions, in which bonds are broken and formed, generally must be described with quantum mechanics. Several groups have developed and implemented quantum mechanics MD methods, often called ab initio MD (AIMD), for this purpose.6–18 To reproduce experimental properties accurately, AIMD simulations over a long time period are necessary. Because the systems of interest usually contain many atoms and electrons, AIMD simulations are computationally challenging. To overcome this computational bottleneck, several methods have been developed to reduce the computational effort per time step in AIMD simulations. These methods include quick step,7 GPU-accelerated AIMD,19 the effective fragment potential method,20 and fragmentation methods21 such as the fragment molecular orbital (FMO) method.13 Fragmentation methods have been applied to investigate reactions in solvents.9–17,22 However, many fewer MD simulations for open shell systems have been reported than for closed-shell systems.23 In particular, there are few reports of AIMD simulations for systems with many open shell solvent molecules, such as simulations of liquid oxygen.24 Such simulations would be useful, for example, to apply quantum chemistry to electronics and spintronics.25,26 Such AIMD simulations require the availability of an exact analytic gradient for the chosen electronic structure method.

For systems with open-shell molecules, one can use either restricted open shell Hartree–Fock (ROHF)27–33 or unrestricted Hartree–Fock (UHF).34 The present work presents the exact analytic FMO-ROHF energy gradient for multiple
open-shell systems. The coupled perturbed Hartree–Fock (CPHF) equations have been developed and implemented to apply to any number of ROHF fragments. Because the response theory for the ROHF method has been thoroughly discussed by Schaefer et al., only the modifications that are required for the FMO method are discussed here.

The FMO method is based on a fragmentation approach in which a system is divided into small fragments. The energy of each fragment (also called a monomer) is then calculated independently of the others, but in the presence of a bath (i.e., embedding) electrostatic potential (ESP) that represents all of the other fragments. The monomer calculations are done self-consistently, as one would do in an ordinary Hartree–Fock calculation. Once the monomer calculations have been completed, one can perform a set of (single) dimer (fragment pair) calculations. This level of FMO theory is called FMO2. One can also include explicit calculations on all sets of three monomers (trimers) to obtain the FMO3 method. Several approximations can be introduced to improve the efficiency of the FMO method. The most commonly used approximations are the electrostatic dimer (ES-DIM) approximation for far separated dimers and the point charge approximation for the electrostatic potential (ESP-PC). With these approximations, one attains nearly linear scaling with the system size. The FMO method can be interfaced with most levels of electronic structure theory and has been applied to a variety of chemical and biological systems. DNA, ionic liquids, organic pigments, boron nitride nanorings, and silicon nanowires and nanodots has also been developed for the FMO method. One can replace the ES-DIM and ESP-PC approximations with the more sophisticated effective fragment potential (EFP) method to produce the integrated e fragment potential (EFP) method. This study presents the derivation of the fully analytic energy gradient for the FMO-ROHF method and extends its applicability to any number of open-shell fragments. The accuracy of the FMO-ROHF method for multiple open shell systems is evaluated by comparison with fully ab initio calculations. FMO NVE MD simulations were performed on pure oxygen and N2–O2 mixtures. The results are compared to experimental radial distribution functions. Finally, the computation time requirement is evaluated using mixed molecular clusters of oxygen and nitrogen.

2. MATHEMATICAL FORMULATION

2.1. Overview of FMO-ROHF Analytic Gradient. The FMO2 energy may be written as

$$E = \sum_{i<\text{RHF}} E_i + \sum_{i>j<\text{RHF}} [E_{ij} - E_i - E_j + \text{Tr}(\Delta D^i y^{ij})] + \sum_{k<\text{ROHF}} E_k + \sum_{k<l<\text{ROHF}} [E_{kl} - E_k - E_l + \text{Tr}(\Delta D^k y^{kl})] + \sum_{i<\text{RHF}} \sum_{j<\text{ROHF}} [E_{ij} - E_i - E_j + \text{Tr}(\Delta D^i y^{ij})]$$

(1)

In eq 1 N_{RHF} and N_{ROHF} are the numbers of RHF or ROHF fragments, respectively. I, J denote RHF fragments, and K, L denote ROHF fragments. $E_i$ is the internal energy of monomers (X = I) or dimers (X = IJ). $y^{ij}$ is the ESP used for dimer $IJ$. $\Delta D^i$ is the density transfer matrix, $\Delta D^i = D^i - (D^i \otimes D^i)$, where $D^i$ is the density matrix of fragment X. Though the monomer energies are iterated to self-consistency with respect to the ESP, the dimer energies are not (ESP is fixed). Therefore, the FMO2 method is not fully variational, so response terms must be included to obtain a fully analytic gradient.

The analytic gradient of the energy with respect to a nuclear coordinate $a$ for open-shell systems can be written as a linear combination of terms for both RHF and ROHF fragments. The derivative of the total energy (for simplicity, without making the ESP approximation), is

$$\frac{\partial E}{\partial a} = \sum_{i<\text{RHF}} \frac{\partial E_i}{\partial a} + \sum_{i>j<\text{RHF}} \frac{\partial}{\partial a} [E_{ij} - E_i - E_j + \text{Tr}(\Delta D^i y^{ij})] + \sum_{k<\text{ROHF}} \frac{\partial E_k}{\partial a} + \sum_{k<l<\text{ROHF}} \frac{\partial}{\partial a} [E_{kl} - E_k - E_l + \text{Tr}(\Delta D^k y^{kl})] + \sum_{i<\text{RHF}} \sum_{j<\text{ROHF}} \frac{\partial}{\partial a} [E_{ij} - E_i - E_j + \text{Tr}(\Delta D^i y^{ij})]$$

(2)

The main difference between the FMO-ROHF and FMO-UHF methods is found in the orbital relaxation terms. In the FMO-ROHF method the total energy gradient is separated into terms that include the response terms $U^{\alpha} r^\alpha$ and the other terms as follows:

$$\frac{\partial E}{\partial a} = \frac{\partial E'}{\partial a} + \frac{\partial E"}{\partial a}$$

(3)

where $(\partial E'/\partial a)$ is the analytic energy gradient without the response terms and $(\partial E"/\partial a)$ come from the response term $U^{\alpha} r^\alpha$. The gradient without the response terms mentioned below is $(\partial E'/\partial a)$.

$$r^\alpha + U^{\alpha} = 4 \sum_{i<\text{RHF}} \sum_{j<j} \sum_{i<j} l_{i}^{\alpha} l_{j}^{\alpha} + 2 \sum_{k<\text{ROHF}} \sum_{l<l} \sum_{k<l} l_{k}^{\alpha} l_{l}^{\alpha} + 4 \sum_{k<\text{ROHF}} \sum_{l<l} \sum_{k<l} l_{k}^{\alpha} l_{l}^{\alpha}$$

(4)

where MOs $r$ and $i$ run over all of the independent pairs (see below), $\alpha$ refers to a virtual orbital, $\alpha$ refers to a singly occupied molecular orbital, $\delta$ refers to a doublt occupied molecular orbital, and $\nu$ denotes all occupied molecular orbitals ($\alpha$ and $\delta$). In eq 4, I runs over closed shell fragments and K runs over open shell fragments. Without the point charge ESP approximation (ESP-PC), the Lagrangian is defined as $Y = I$ or $K$

$$l^{\alpha}_{ri} = -V^{\alpha}_{ri} + \sum_{j \neq i} \sum_{\mu, \nu} \Delta l^{\alpha}_{\mu \nu} (\mu, \nu | ri)$$

(5)

In the Lagrangian, $l_{ri}$ there are three contributions (see ref 64 and 85 for details on the formalism of the Lagrangian with the ESP approximation ESP-PC). The first contribution comes from the ESP term in monomer Y. The second and third terms come from the dimer ESP, so there are summations over all of the fragment pairs except those treated approximately. $(\mu, \nu | ri)$ is
a two-electron Coulomb integral. The Greek indices denote the atomic orbital basis. \( R_{ij} \) is the \( i-j \) interfragment distance. If the distance between two monomers is greater than a threshold \( R_{ij,\text{DIM}} \), then the dimer energy is approximated as the electrostatic interaction between the two monomers. \( V^{\text{el}}_{ij} \) is the electrostatic potential in the MO basis. \( V^{\text{el}}_{ii} \) and \( V^{\text{el}}_{ij} \) are defined as

\[
V^{\text{el}}_{ij} = \sum_{\mu \neq \nu} \left( \frac{Z_{\mu}}{|r_{\mu} - R_{ij}|} \right) + \sum_{\mu, \nu} D^{j}_{\mu} (\mu \nu i j)
\]

\( I \neq j \) (6)

\[
V^{\text{el}}_{ii} = \sum_{K \neq I, J} \sum_{\mu, \nu \in K} \left( \frac{Z_{\mu}}{|r_{\mu} - R_{ij}|} \right) + \sum_{K \neq I, J} \sum_{\mu, \nu \in K} D^{K}_{\mu} (\mu \nu i j)
\]

\( ri \in I \) (7)

The electron and nuclear positions are \( r \) and \( R_{ij} \), respectively, \( A \) labels atoms, and \( Z_{\mu} \) is the atomic nuclear charge.

The analytic energy gradient for FMO-ROHF requires solving for the unknown response terms \( U \) running over all three types of independent pairs: from virtual to doubly occupied, from singly occupied to doubly occupied, and from virtual to singly occupied molecular orbitals. In the following subsections, the derivation of the CPHF equations is discussed to obtain the unknown response terms \( U \); this is the main new development in the present study.

### 2.2. First-Order Perturbed Hartree–Fock Equation for FMO-ROHF

For FMO analytic energy gradients, it is necessary to solve the first-order coupled perturbed Hartree–Fock equations. The FMO CPHF equation is derived from the first-order derivative of the Fock matrix on the basis of the Roothaan restricted open-shell HF theory, in the presence of the ESP potential:

\[
F^{K}_{ij} = h^{K}_{ij} + \sum_{k \in K} [2(ijkl) - (iklj)] + \sum_{k \in K} (ijkl) - \frac{3-f_{i}-f_{j}}{2} \sum_{m \in K} (imljm) + V^{K}_{ij}
\]

In eq 8, \( i,j \) are two orbitals in one of the following sets: (virtual \( i \), doubly occupied \( j \)) or (singly occupied \( i \), doubly occupied \( j \)) or (virtual \( i \), singly occupied \( j \)). These orbital pairs are called independent pairs. \( f_{i} \) and \( f_{j} \) are occupation numbers for molecular orbitals \( i \) and \( j \), respectively (i.e., for a doubly occupied orbital, \( f_{i} = 2 \)). The last term in eq 8 is identified in eq 7.

By taking \( F_{ij} = \delta_{ij} F_{p} \) and the derivative of the canonical Fock matrix (a detailed derivation is provided in the Supporting Information), one can obtain the first-order coupled perturbed Hartree–Fock equations for the FMO-ROHF method:

\[
AU^{\sigma} = B_{0}^{\sigma}
\]

In eq 9 \( A \) is the orbital Hessian matrix and \( U^{\sigma} \) is the response term, and \( B_{0}^{\sigma} \) is the derivative of the integral terms. The explicit formulations for these terms are discussed below. For convenience of solving eq 9, it is separated into a set of linear equations for individual fragment pairs. Then, the product of \( AU^{\sigma} \) can be separated into fragment pair contributions. The equations are reformulated using the self-consistent-Z-vector (SCZV) method, for each fragment with the inclusion of contributions of \( A \) elements from other fragments:

\[
\begin{align*}
AU^{\sigma}_{ij} &= \sum_{k,l \in K} A^{K}_{ij,kl} U^{\sigma}_{kl} + \sum_{N} \sum_{l \in L} A^{L}_{ij,kl} U^{\sigma}_{kl} \\
N \end{align*}
\]

(10)

\( N \) is the total number of fragments (i.e., \( N = N_{\text{HF}} + N_{\text{ROHF}} \)). The supermatrix \( A \) is built from blocks for fragment pairs such as K, L. The diagonal contribution for fragment \( K \) is \( A^{K}_{i,k} \), whereas the off-diagonal contribution for fragments \( K \) and \( L \) is \( A^{K}_{i,k} \). For an independent orbital pair \( k \) and \( l \), the off-diagonal part of the orbital Hessian is

\[
A^{K}_{i,k} = -4 ijkl \quad (K \neq L)
\]

(11)

for virtual \( i \) to doubly occupied \( j \) molecular orbital, and

\[
A^{K}_{i,k} = -2 ijkl \quad (K \neq L)
\]

(12)

for virtual \( i \) to singly occupied \( j \) molecular orbitals or singly occupied \( i \) to doubly occupied \( j \) molecular orbitals. Rather than giving an explicit definition of the diagonal block \( A^{K}_{i,k} \), we give its contraction with \( U \), as needed in eq 10. For fragment \( K \) and \( i \in \text{vir} \) and \( j \in \text{d.o.} \), the contraction of \( A \) with \( U \) is given by

\[
\sum_{i \in \text{vir} \text{ or } j \in \text{d.o.}} \sum_{k \in K} \sum_{l \in K} U^{K}_{ij,kl} A^{K}_{ij,kl} + \sum_{i \in \text{d.o.}} \sum_{k \in K} \sum_{l \in K} U^{K}_{ij,kl} A^{K}_{ij,kl}
\]

Likewise, the product \( AU^{\sigma} \) within the fragment \( K \) for \( i \in \text{s.o.} \) and \( j \in \text{d.o.} \), is given by

\[
\sum_{i \in \text{vir}} \sum_{k \in K} \sum_{l \in K} U^{K}_{ij,kl} A^{K}_{ij,kl} + \sum_{i \in \text{d.o.}} \sum_{k \in K} \sum_{l \in K} U^{K}_{ij,kl} A^{K}_{ij,kl}
\]

(13)

Finally, for \( i \in \text{vir} \) and \( j \in \text{s.o.} \),

\[
\sum_{i \in \text{vir}} \sum_{k \in K} \sum_{l \in K} U^{K}_{ij,kl} A^{K}_{ij,kl} + \sum_{i \in \text{d.o.}} \sum_{k \in K} \sum_{l \in K} U^{K}_{ij,kl} A^{K}_{ij,kl}
\]

(14)

In eq 9 \( B_{0}^{\sigma} \) contains derivative integral terms,
where $\varepsilon^K$ is the energy of orbital $i$ in fragment $K$ and $S^K_{ij}$ is the diagonal part of the overlap matrix. The diagonal parts of the orbital Hessian in eqs 13, 14, and 15 can largely be constructed by the previously developed FMO-UHF response code,\textsuperscript{70} because $A$ corresponds to $(A' + A'')/2$ for each virtual to doubly occupied pair, to $A'$ for each singly occupied to doubly occupied pair, and to $A''$ for each virtual to singly occupied pair. Therefore, the new terms in the FMO-ROHF gradient derivation are the final four terms in eqs 13, 14, and 15 and the corresponding terms in eq 16. These new terms can be evaluated by using atomic orbital basis integrals and then transformed to MO indices for use in equations (eq 13–16).

It is necessary to solve a set of linear equations for all independent pairs coupled over the entire system. Practically, the set of CPHF equations (eq 9) is converted into a $Z$-vector equation:\textsuperscript{86}

$$AZ = L_{\text{tot}}^{\alpha}$$

(17)

where $L_{\text{tot}}^{\alpha}$ is the vector formulation of the Lagrangian, $L$, defined in eq 5. After solving eq 17 for $Z$, the gradient contribution is obtained as

$$R^\alpha + U^\alpha = [L_{\text{tot}}^{\alpha}]^{-1}B_0^\alpha = Z^\alpha B_0^\alpha$$

(18)

In eqs 17 and 18, the Langrangian $L_{\text{tot}}^{\alpha}$ (and, likewise, $Z$) is a supervector, which includes the corresponding Langrangians ($Z$ vectors) for each fragment $I$ (see eq 5). $A$ in eq 18 is a supermatrix, with blocks for pairs of fragments $I, K$ (see eqs 11 and 12). Equation 17 may be decomposed into diagonal and off diagonal parts:\textsuperscript{86}

$$\sum_{i,p} A_{ij,kl}^{I,K} Z_{ik}^{I} = L_{ij}^{I} - \sum_{k \neq l} \sum_{i,p} A_{ij,kl}^{I,K} Z_{ik}^{I}$$

(19)

Equation 19 is formulated for each independent pair $i, j$ for every fragment $I$. The product $\sum_{i,p} A_{ij,kl}^{I,K} Z_{ik}^{I}$ is calculated with eqs 13–15 by replacing $U_{kl}^{I}$ with $Z_{kl}^{I}$. One can solve the set of linear equations in eq 19 with the iterative conjugate gradient solver as is described in detail elsewhere.\textsuperscript{70,86}

### 2.3. Implementation of Multiple Open Shell Method

For FMO calculations on multiple open-shell systems, the multiplicity for each subsystem (fragment) is predefined and fixed. For fragment pair calculations (dimers), it is not necessary to define the multiplicity, as it is currently taken to be the high-spin coupling of the spins of its two monomers.

As an example, consider the system constructed with two oxygen and one nitrogen molecules. In this system, there are two different kinds of dimer calculations, oxygen--oxygen and oxygen--nitrogen. The net spin $S$ of the two oxygen molecules is taken to be 2, and the spin of the oxygen--nitrogen dimer is 1. If there is no unpaired electron in a dimer, it can be calculated by RHF with $S = 0$. The approach taken here considers the spin state of the whole system, equal to the sum of the spins of all fragments. Note that in the case when the coupling between open-shell fragments is small (as is the case in oxygen), the energy difference between the high and low spin states is very small.

To test the difference between low-spin and high-spin, gas phase calculations on isolated dimers were performed, with geometry optimizations for a number of different angular arrangements, using an MCSCF treatment of the $\pi$ and $\pi^*$ orbitals and electrons. The spin of $S = 1$ for one $O_2$ molecule may couple to a total $S = 0, 1, 2$ for the $O_2$ dimer. Binding energies for various geometric arrangements of two $O_2$ molecules are found to be less than 0.1 kcal/mol, with the splitting between the three possible $S$ values being about 0.005 kcal/mol. At some geometries, the $S = 0$ state is lowest, whereas at others $S = 2$ is lowest, but all three states are always very close together. This quasi-degeneracy justifies the assumption of high-spin coupling for any FMO dimer calculations.

The current approach could also be applied to open-shell fragments connected by covalent bonds, for instance, multiple radical centers on a polymer chain, such as unterminated polystyrene and polyethylene. If two open-shell fragments are close and strongly interact, the two open-shell centers should merge into one fragment, and therefore, one might prefer to treat it by the FMO-MCSCF (multiconfiguration self-consistent field) method.

The starting dimer orbitals for an open shell system are constructed using the converged molecular orbitals obtained from monomer calculations. As is shown in eq 8, the formulation of the ROHF Fock matrix is different for each virtual, singly occupied, and doubly occupied molecular orbital. Therefore, the three different shells are treated separately when the initial molecular orbitals are constructed. This is shown schematically in Figure 1. If one has two triplet oxygen molecules, there are two different sets of singly occupied orbitals, one set localized on fragment $K$ and the second set localized on fragment $L$. These four singly occupied orbitals are separated from the doubly occupied orbitals, and reordered so as to be the dimer open shell orbitals. Because there may be

![Figure 1. Schematic illustration of how to construct the initial molecular orbital for an open-shell dimer if both monomers are open-shell molecules. KFG and LFG denote the monomer fragments, and the dimer is constructed from KFG and LFG.](image)
many open shell dimer calculations, the preparation of the starting molecular orbitals for dimer open shell fragments is important to gain good convergence behavior and efficient computational timings as is shown below.

3. COMPUTATIONAL DETAILS

FMO-ROHF calculations for multiple open-shell fragments have been implemented in the GAMESS electronic structure program, and parallelized with the generalized distributed data interface (GDDI). First, the accuracy of the analytic FMO-ROHF gradient method is evaluated by comparison to its corresponding numerical energy gradient and to the analytic gradient of the fully ab initio method. The numerical energy gradient is calculated using double differences with a step size of 0.0005 Å. It is important to analyze the effect of multiple open-shell fragments. Therefore, four different systems have been constructed: (1) 41 oxygen molecules, (2) a mixture of 31 nitrogen and 10 oxygen molecules, (3) 41 nitrogen molecules, and (4) an unuterminated polyethylene 18 mer (−(C\(_2\)H\(_2\))\(_{18}\)) that has two radical centers at both ends. The latter system was divided into 9 fragments (two ethylene units per fragment). Hybrid projection operators were used to describe fragment boundaries. The ratio of nitrogen:oxygen in the second example is 3:1. This corresponds to the ratio in air. These four systems were optimized using the FMO-ROHF or FMO-RHF method, as appropriate, as well as with the corresponding fully ab initio method. The final geometries and energies are compared. If the multiple open-shell FMO-ROHF method is reasonable, all of the different kinds of systems, which possess various multiplicities, should reproduce the corresponding ab initio calculations with similar accuracies.

For a more rigorous test of the quality of the analytic energy gradient, NVE MD simulations were performed, and the energy conservation evaluated. For this purpose, 6 ps NVE MD simulations were performed on the above three systems starting from the optimized geometries using a 0.5 fs time step. The root-mean-square error of the energy (RMSE) was evaluated for each system. Velocity-Verlet was used for the time integration of all the MD simulations.

Second, MD simulations on liquid oxygen have been performed, and the calculated radial distribution function (RDF) have been compared to experimental results. Experimentally, liquid oxygen is a paramagnetic liquid between 90.2 and 54.4 K. Therefore, the total multiplicity of liquid oxygen is not high-spin, but rather some intermediate coupling of high- and low-spin structures due to random thermal motions. However, as a simple test of whether the ROHF calculation works correctly, and to make the analysis simple, the high-spin electronic structure is chosen throughout this study.

Finally, the computational timing is evaluated using spherical clusters of air. The diameters of the spheres are 6 Å (82 atoms), 9 Å (234 atoms), 12 Å (540 atoms), 15 Å (994 atoms), 20 Å (2244 atoms), and 25 Å (4100 atoms).

For all of the calculations, the 6-31G(d) basis set with spherical harmonics is used. The geometry optimizations were performed using the default geometry optimizer in GAMESS based on numerical updates of the Hessian. The MD simulation was performed with an NVT ensemble at 77 K. A spherical boundary potential was applied to maintain the sphere as in eq 20:

\[
V = \frac{1}{2} \sum_A {S_{\text{force}}(R_A - R_{\text{off}})^2} \quad (\text{If } R_A > R_{\text{off}})
\]

In eq 20 \(R_A\) is the distance between each atom \(A\) and the origin, and \(R_{\text{off}}\) is the radius of the spherical boundary. \(S_{\text{force}}\) is a force constant (2 kcal/(mol Å\(^2\))) designed to keep the atoms within a distance \(R_{\text{off}}\) from the center of the sphere.

In the FMO method, the ES dimer and point charge ESP approximations were applied with the thresholds \(R_{\text{ES-DIM}} = 2.0\) and \(R_{\text{ESP-PC}} = 2.5\), respectively. Note that both thresholds are unitless, because they are applied to the interatomic distances normalized by the sum of the atomic van der Waals radii. The geometry optimizations and MD simulations were performed with an empirical Grimme dispersion correction version 3 (ROHF-D3). The FragIt program was used to make the FMO input files.

4. RESULTS AND DISCUSSION

4.1. Evaluation of Gradient Accuracy. To evaluate the accuracy of the FMO-ROHF method for multiple open shell fragments, the analytic energy gradient was compared to the numerical energy gradient, and geometry optimizations were compared to the results of fully ab initio calculations.

In Figure 2, a comparison between the analytic and numerical energy gradients is shown. Without the response terms (dotted black line), the RMSD and Max errors relative to the numerical gradient are 0.000234 and 0.000067 hartree/bohr, respectively. When the response terms are included (thick red line), the RMSD and Max errors relative to the numerical gradient are 0.000012 and 0.000004 hartree/bohr, respectively, an order of magnitude improvement. In fact, the analytic gradient agrees with the numerical gradient to within the expected error in the latter calculation. The improved gradient is especially important for energy conservation in MD simulations. Therefore, the analytic energy gradient including the response terms is used hereinafter.

Four different systems were geometry optimized as noted above, 41 O\(_2\) molecules, 31 nitrogen and 10 oxygen molecules, 41 N\(_2\) molecules, and the polyethylene 18 mer. A summary of the optimization results is shown in Figure 3 and Table 1. In Figure 3, the equilibrium geometry differences between FMO and ab initio calculations are shown for both 41 O\(_2\) molecules, and 31 nitrogen + 10 oxygen molecules. Gray colors represent the positions of the fully ab initio atoms, and blue and red
colors show the FMO atoms. Therefore, the differences between the blue and gray colors represent the errors in the nitrogen atom positions, whereas the differences between the red and gray colors represent the errors in the O atom positions. The positions of most N2 and O2 molecules in FMO are very close to those in full ab initio calculations. Some deviations are observed at the outside surface of the system, because the potential energy surface is very flat. The RMSD FMO errors relative to the ab initio atom positions are 0.111 Å for the oxygen system, 0.175 Å for air, 0.126 Å for the nitrogen cluster, and 0.029 Å for the 18-mer of unterminated polyethylene. Thus, the size of the error for open shell systems is the same as that for closed shells. The average energy error at the stationary points is ∼1 kcal/mol.

4.2. Molecular Dynamics Simulations for Liquid Oxygen. First, the accuracy of the MD simulation is evaluated by testing the energy conservation in NVE MD simulations. For this purpose, MD simulations of oxygen molecules, nitrogen molecules, and mixtures of nitrogen and oxygen molecules were performed to evaluate the effect of multiple open-shell fragments. The energy conservation for ROHF should be comparable to that found previously for the RHF method.13,18 The computational details are described in section 3.

The energy conservation results are shown in Figure 4. The energy conservation for a 6 ps NVE MD simulation is shown in Figure 4a. A time step of 0.5 fs was used in these simulations. The deviations of the ROHF simulations are similar to those of the RHF simulations, thereby confirming acceptable energy conservation for the open shell systems.

After the 6 ps NVE MD simulation, the RMSE was evaluated at several time steps. The RMSE is defined as

![Figure 4. Summary of the ROHF-D3/6-31G(d) NVE MD simulation, with RESDIM = 2.0, RESPPC = 3.5, 0.5 fs time step. (a) Energy trajectory for 6 ps NVE MD simulation. The red solid line is air, the blue dashed line denotes the oxygen system, and the black dotted line denotes the nitrogen system. The vertical axis is the energy difference from the initial energy, and the horizontal axis is the elapsed time from the beginning of the MD simulation. (b) RMSE energy (kcal/mol), relative to the average energy during the MD simulation (log−log plot). The red solid line with rectangles denotes the oxygen system, the blue dotted line with circle denotes the air system, and the black dashed line with triangle denotes the nitrogen system. The horizontal axis is the time step (fs), and the vertical axis is the RMSE energy (kcal/mol) relative to the average energy during the 10 fs MD simulation after 6 ps is elapsed.](image-url)
The pure oxygen system (123 oxygen molecules) including response terms can be used reliably from 0.1 to 0.7 fs time steps. The MD simulations were performed using different time steps, 0.1, 0.3, 0.5, and 0.7 fs, to evaluate the gradient quality. In Figure 4b, the red solid line denotes the RMSE for the air-like mixture, and the blue dashed line denotes the RMSE for pure oxygen, both using the FMO-ROHF method. The black dotted line denotes the FMO-RHF RMSE for the nitrogen molecules, all relative to the fully NVT energy conservation with respect to the time steps is about 2.11, 2.00, and 1.95 for air, oxygen molecules, and nitrogen. The slope around the theoretically expected value$^{94}$ of 2.0 suggests that the analytic FMO-ROHF energy gradient including response terms can be used reliably from 0.1 to 0.7 fs time steps.

NVT MD simulations were performed to describe the geometry and electronic structure of air and liquid oxygen.$^{24,95,96}$ The pure oxygen system (123 oxygen molecules) and three kinds of N$_2$/O$_2$ systems were studied, in which the numbers of oxygen and nitrogen molecules are different: (1) 99 oxygen molecules and 24 nitrogen molecules, (2) 62 oxygen molecules and 61 nitrogen molecules, (3) 36 oxygen molecules and 87 nitrogen molecules.

The distribution of both distance and angle between the molecules are shown in Figure 5a. The distance between fragments is the distance between the closest pair of atoms (oxygens in Figure 1). The angle between the molecules is measured as depicted in Figure 1. The angle between KFG and LFG is defined as the angle constructed by the two vectors associated with KFG and LFG as shown in Figure 1. Because all four N$_2$/O$_2$ systems reveal similar trends, results are shown only for the O$_2$:N$_2$ = 3:7 mixture. The results for the MD simulations for the other three mixtures are available in the Supporting Information (Figures S1–S3). The populations of the intermolecular distance (horizontal axis) and angle (vertical axis) are shown in Figure 5. For the intermolecular distance, the maximum population is ∼3.5 Å, and the population decreases steeply as the intermolecular distance increases. When the intermolecular distance increases, the population decreases slowly. Angles around 90–100° appear to be preferred.

The simulation results are compared to the experimental and other calculations$^{24}$ of the radial distribution function (RDF) in Table 2 and Figure 5b for liquid oxygen. The distances between the molecules monotonically decrease with the increase of the population of oxygen molecules, and the computational results are in good agreement with the experimentally measured RDF. In Figure 5b, the prominent peak of 1.16 Å is the O–O distance within a single oxygen molecule. The distance between atoms of different oxygen molecules has two peaks around 3.5 and 6 Å separated by a minimum around 5 Å. The overall structure of this RDF agrees well with the previously reported liquid oxygen simulation in ref 24. The first peak predicted in this work, the previous theoretical study$^{24}$, and experiment$^{84}$ are at 1.16, 1.24, and 1.24 Å, respectively. The second peak is at 3.28, 3.6, and 3.41 Å, respectively. The second peak corresponds to the intermolecular distance between two O$_2$ molecules. The somewhat shorter distance of the peak compared with experiment may be caused by the general over binding of ROHF. One would expect that the agreement with experiment can be improved with the incorporation of electron correlation in the calculations.

### 4.3. Computational Timing

The computational timings of a single point analytic energy gradient calculation for the air (mixed N$_2$/O$_2$) and pure N$_2$ systems are evaluated, using the ROHF and RHF methods, respectively. The calculations, the results of which are shown in Figure 6, were performed with 16 2.93 GHz Xeon nodes (8 cores per node and 12 GB memory per node). The horizontal axis is the number of atoms in the system, and the vertical axis is the computational timing in hours. For 4100 atoms, only 6 h were required.

Now, consider a timing comparison to the full HF energy gradients calculation on the smallest two oxygen–nitrogen mixtures, corresponding to air filling a sphere with a diameter of 6 Å (82 atoms) and 9 Å (234 atoms). The full HF calculation is impossible for the larger systems because of the large number of atoms. The calculations were performed on a supercomputer with 4100 atoms. The calculations were performed using 16 2.93 GHz Xeon nodes (8 cores per node and 12 GB memory per node).

<|table|>

<table>
<thead>
<tr>
<th>system</th>
<th>first peak</th>
<th>second peak</th>
<th>first peak</th>
<th>second peak</th>
</tr>
</thead>
<tbody>
<tr>
<td>O$_2$:N$_2$ = 3:7</td>
<td>1.10</td>
<td>3.52</td>
<td>1.16</td>
<td>3.69</td>
</tr>
<tr>
<td>O$_2$:N$_2$ = 5:5</td>
<td>1.12</td>
<td>3.44</td>
<td>1.18</td>
<td>3.58</td>
</tr>
<tr>
<td>O$_2$:N$_2$ = 8:2</td>
<td>1.14</td>
<td>3.32</td>
<td>1.22</td>
<td>3.47</td>
</tr>
<tr>
<td>all O$_2$</td>
<td>1.16</td>
<td>3.28</td>
<td>1.24</td>
<td>3.41</td>
</tr>
</tbody>
</table>

$^a$Both experimental$^{84}$ and simulation results are shown. The first peak corresponds to the average distance inside the oxygen and nitrogen diatomics. The second peak corresponds to the nearest intermolecular distance. The temperature is 77 K.

### Figure 5.

(a) Summary of the geometric information obtained from the MD simulation of O$_2$:N$_2$ = 3:7. The horizontal axis is the distance between molecules (Å), and the left vertical axis is the angle between two molecules (degrees). The color denotes the frequency for each given angle and bond length. (b) Radial distribution function (RDF) for pure oxygen as a function of the interatomic separation (red solid line) and the distance between the centers of mass of O$_2$ molecules (blue dashed line).
large memory requirement. The computational timing for FMO and full HF are 0.48 and 6.2 min, respectively, for 82 atoms; whereas for 234 atoms the timings were 3.3 and 132.7 min, respectively. Using FMO, the time required for the calculation decreased by a factor of 12.9 and 40.2 for the systems consisting of 82 and 234 atoms, respectively.

The difference between the FMO-ROHF and the FMO-RHF time requirements for the analytic energy gradient is small, less than half an hour for the largest system in this study. The difference is primarily due to the fact that the ROHF calculation took 22 iterations compared with 14 iterations for the RHF calculation to obtain the ESP by self-consistent monomer cycles.

5. CONCLUSIONS

The energy and fully analytic gradient for the multiple open shell FMO-ROHF method has been derived, implemented, and evaluated. The FMO-ROHF method can reproduce conventional ab initio results with an accuracy that is comparable to the accuracy of FMO-RHF calculations. The accuracy of the analytic energy gradient is excellent, sufficient to run MD simulations with good energy conservation. FMO-ROHF MD simulations have been used to calculate the RDF of liquid oxygen, obtaining good agreement with experiment. Populations of various geometric configurations have been studied for several mixtures of nitrogen and oxygen.

The computational timings demonstrate that the FMO-ROHF method has low scaling. A further reduction of the scaling, approaching linear scaling, can be achieved with the multipole expansion of the electrostatic interaction. For 234 atoms, the FMO-ROHF calculation is 40.2 times faster than full ROHF; an FMO-ROHF calculation of a nitrogen–oxygen mixture containing 4000 atoms took only 6 h on 16 Xeon nodes containing 128 cores.

ASSOCIATED CONTENT

Supporting Information

A detailed mathematical derivation of the gradients, and probability maps as a function of geometry for several mixtures of oxygen and nitrogen from FMO-MD simulations are available free of charge via the Internet at http://pubs.acs.org.

REFERENCES


ACKNOWLEDGMENTS

D.G.F. thanks the Next Generation Super Computing Project, Nanoscience Program (MEXT, Japan), and Computational Materials Science Initiative (CMSI, Japan) for financial support. Calculations were performed on TSUBAME2.5 at the Global Scientific Information Computing Center of Tokyo Institute and RIKEN Integrated Cluster of Clusters (RICC) at RIKEN. M.W.S. and M.S.G. were supported by a grant from the Air Force Office of Scientific Research under AFOSR Award No. FA9550-11-1-0099. H.N. thanks JSPS for supporting of this work by KAKENHI Grant Number 262235.