

Multi-Scale Supercomputing of Large Molecular Aggregates: A Case Study of the Light-Harvesting Photosynthetic Center

Igor V. Polyakov¹, Alexander A. Moskovsky^{1, 2, 3},
Alexander V. Nemukhin^{1, 3}

© The Authors 2015. This paper is published with open access at SuperFri.org

Numerical solution of the quantum mechanical Schrödinger equation is required to model electronic excitations in the light-harvesting photosynthetic complexes composed of up to millions of atoms. We demonstrate that the modern supercomputers can be used to treat electronic structure calculations in large molecular aggregates if proper multi-scale massive-parallel approaches are applied. We show that three-level parallelization scheme based on the novel numerical algorithms assuming fragmentation of a light-harvesting complex allows us to considerably reduce the high scaling of *ab initio* quantum chemistry methods. More specifically, we have applied the time-dependent density functional theory based on the fragment molecular orbital presentation (FMO-TDDFT) implemented in modern supercomputers to obtain realistic estimation of the electronic excitation in the complex. The application shows a good overall scaling.

Keywords: quantum chemistry, fragmentation, multi-scale approaches, parallel algorithms.

Introduction

Application of *ab initio* quantum-chemical algorithms aiming at numerical solutions of the Schrödinger equation for molecules has served as a polygon for testing high-performance computers for a long period of time. One of the reasons is that the computational cost of such simulations increases dramatically with the system size N , namely, up to N^6 , especially if electronically excited states are involved. Presently, quantum-chemical calculations can be performed efficiently through a number of modern software readily available for any of supercomputer architectures. Most of the popular programs are fairly scalable and perform well with the big x86-64 clusters, some of these applications are capable to harness accelerators such as NVIDIA K40 or Xeon Phi devices. Certain algorithms also require vast amounts of memory per core which can be a problem for most computational clusters. That is why the routine calculations are usually conducted for moderate sized systems containing several hundreds of atoms in total. They usually scale well for up to 256 cores on the Infiniband clusters.

In this respect characterization of the photosynthetic molecular machines responsible for conversion of solar energy to chemical energy presents one of the greatest challenges for quantum-based simulations. Relevant model systems mimicking the reaction centers (RC) and the light harvesting (LH) complexes should incorporate up to millions of interacting atoms, and calculations of electronic excitations in the photosensitive part of such complexes constitute the most difficult problem.

In attempts to reduce the high scaling of *ab initio* methods and novel numerical algorithms based on fragmentation of a large molecular aggregate have begun to extend the traditional quantum-chemistry approaches. Recently developed methodology of fragment molecular orbitals (FMO) [2] [6] [7] opens the way for application of algorithms of quantum chemistry for a very

¹Department of Chemistry, M.V. Lomonosov Moscow State University, Leninskie Gory 1/3, Moscow 119991, Russian Federation

²ZAO RSC Technologies, Kutuzovskiy av., 36/23, Moscow 121170, Russian Federation

³N.M. Emanuel Institute of Biochemical Physics, Russian Academy of Sciences, Kosygina 4, Moscow 119334, Russian Federation

large molecular system assuming a properly designed fragmentation scheme. We show here how this method can be applied for a photosynthetic LH-RC complex comprising a half of million atoms. Presumably, the FMO technique belongs to the group of highly parallel methods of quantum chemistry. In this work we check the corresponding options using current implementation of the FMO algorithms in modern facilities utilizing both large and small computational clusters.

1. Models and Methods

It is beyond the scope of this paper to describe construction of a model molecular system for the photosynthetic center. We only mention that such work presents a separate highly demanding computational problem and requires applications of consuming molecular mechanics and molecular dynamics calculations. We refer to our recent work [5] in which the bacterial photosynthetic core complex LH-RC composed of a half of million atoms has been modeled. Here we use the atomic coordinates of this all-atom three-dimensional model system to characterize its electronic excitations.

An elliptical ring of 32 bacteriochlorophyll (BChl) chromophores (each of them being a very large molecule) embedded in the LH polypeptide chains lies in the heart of molecular machinery. Subtle interaction of the BChl molecules between themselves as well as with the surrounding molecular groups plays an essential role in the excitation process. It is unrealistic to expect that a direct calculation of the electronic excitations in such system by using *ab initio* quantum chemistry methods will be carried out in the nearest future. Application of the multi-scale fragmentation technique [2] [7] is an important step toward the quantum-based characterization of the system.

According to the fragment molecular orbital (FMO) method a molecular system is divided into a number of fragments (N). The quantum Hamiltonian of the fragment I is written in the conventional form [6] (Eq.1)

$$H_I = \sum_i^{n_I} \left\{ -\frac{1}{2} \bar{V}_i^2 - \sum_s^{all-atoms} \frac{Z_s}{|r_i - r_s|} + \sum_{J \neq I}^N \int dr' \frac{\rho_J(r')}{|r_i - r'|} \right\} + \sum_{i>j}^{n_I} \frac{1}{|r_i - r_j|} + \sum_{s>t}^{fragment} \frac{Z_s Z_t}{|r_s - r_t|} \quad (1)$$

Here n_I - number of electrons in the fragment I, Z_s - nuclear charge of atom s, $\rho_J(r)$ - electron density of fragment J. After solving the Schrödinger equations at an appropriate level by using either the Hartree-Fock (HF) or the density functional theory (DFT) one obtains the fragment energies E_I and fragment pair (in the second-order FMO theory, FMO2) energies E_{IJ} . The total energy of the system can thus be expressed as follows [1] (Eq. 2)

$$E = \sum_{I=1}^N E_I + \sum_{I=1}^N (E_{IJ} - E_I - E_J) \quad (2)$$

To treat the electronic excitations the time-dependent density functional theory (TDDFT) based upon the fragment molecular orbital presentation [1] is applied. In the FMO-TDDFT method the excitation is assumed to be localized mostly on one fragment (M), other fragments contribute the additive two-body corrections (Eq. 3)

$$E^* = E_M^* + \sum_{I \neq M} E_I^0 + \sum_{I \neq M} (E_{MI}^* - E_M^* - E_I^0) + \sum_{\substack{I>J \\ I, J \neq M}} (E_{IJ}^0 - E_M^* - E_I^0) \quad (3)$$

where E^0 notation is for the ground and E^* for the excited state energies. The excitation energy ω in the FMO2-TDDFT approach is given as a sum of the target monomer excitation energy ω_M and a sum of dimer corrections:

$$\omega \equiv E^* - E^0 = \omega_M + \sum_{I \neq M} (\omega_{MI} - \omega_M) \quad (4)$$

The FMO algorithms are implemented in modern versions of the GAMESS(US) quantum-chemistry package [10] which are currently running on Moscow State University Lomonosov supercomputers [9] and on MVS-10P cluster in Joint Supercomputer Center of the Russian Academy of Sciences. An essential feature of this software is the use of generalized distributed data interface (GDDI) [3] which allows the 2-level hierarchical parallelization scheme. This approach allows individual tasks to be assigned to groups that execute them nearly independently of each other approaching linear scalability at the group level. A commonly used parallel library MPI is also capable to divide nodes into groups, but it is a very basic and static interface. The GDDI supports distributed memory operations localized within groups, whereas distributed memory implementation in GAMESS cannot use the original MPI group-divided nodes.

After the initialization in a typical FMO job the single-point HF or DFT calculation (see Eq.(1)) is performed on each fragment separately in the mean Coulomb field of other monomers. This calculation is performed by one DDI group nearly independently of the others. One cycle of the single-point monomer self-consistent field runs generates a new density, thus changing the Coulomb field; these cycles are run until convergence is obtained. This yields the monomer densities that are then used during the dimer runs, also done independently by DDI groups. During this second step the dimers are constructed from each pair of monomers, and the initial density is taken to be the sum of the two monomer densities. Then, a self-consistent field (SCF) calculation is performed for each dimer.

Our previous experience to employ the FMO method to simulate properties of a biomolecular system [8] has resulted in the encouraging conclusions. Linear scalability up to 1536 cores on the RSC Tornado computational cluster has been shown and a limit of 5000 cores has been reported for the system of 2570 atoms.

In this work, FMO2-TDDFT calculations have been carried out for a model photosynthetic center using Lomonosov-2 supercomputer and MVS-10P cluster. GAMESS(US) version 5 (from December 2014 (R1)) has been compiled with Intel Fortran compiler version 14. Intel MKL BLAS library has been used in both cases; however, OpenMPI-1.8.4 has been used on Lomonosov-2 and Intel MPI-4.1 on MVS-10P.

2. Application

Even within the FMO construct it is unpractical to assign all 32 BChl chromophores to a single fragment, to directly compute its excitation energy, and to estimate the contributions from other fragments representing the environmental molecular groups (see Eq.(4)). Therefore, we rely on effective Hamiltonian technique [4] which is often applied to compute excitonic levels of light-harvesting complexes. Usually the empirical or semi-empirical formulae are employed within this approach [4] [12] but in this work we evaluate diagonal and near-diagonal elements of the entire 32 x 32 Hamiltonian matrix by the *ab initio* type quantum calculations. This requires to conduct 32 FMO2-TDDFT calculations which can be run independently, thus, adding a 3rd parallel level on top of regular GDDI 2-level hierarchical parallelization scheme of FMO2-

TDDFT [3]. The remaining off-diagonal elements are small and can be calculated through the formula for dipole-induced dipole coupling as given in Ref. [4].

More specifically, each of 32 calculations is centered around one particular BChl molecule from the BChl ring. This is a target fragment for the FMO-TDDFT calculation (see Eq.(3)). The corrections due to the environmental molecular groups computed within the FMO2-TDDFT method (Eq.(4)) are evaluated as follows. All atoms lying within a certain limit from the selected BChl (grouped to the corresponding fragments) are included into a model subsystem.

To illustrate this computational scheme in more details we refer to the model system of the photosynthetic center described in Ref. [5]. We have selected BChl No.1 as a target chromophore created the subsystem centered around BChl No.1 by cutting all molecules and residues lying outside the limit of 10 Å from BChl No.1. The subsystem size was 3624 atoms including those from a group of 5 BChl molecules (BChl No.31,32,1,2,3). These atoms were combined to 172 fragments for the FMO scheme; also 97 water molecules were added. All broken peptide bonds on the edges of the models were capped with hydrogen atoms. Protein chains were split as one amino acid residue per fragment. Carotenoid residues were split into four equal parts, membrane residues were split into six parts. BChl molecules were divided as follows - the ring with the magnesium atom was assigned to one fragment and the remaining parts to three other fragments. The histidines coordinating the magnesium ions were added to BChl ring fragments.

The two-layer model was employed in every FMO2-TDDFT calculation. Layer 1 was treated at the HF/3-21G level whereas layer 2 - at the more accurate DFT level CAM-B3LYP/6-31G*. The 2nd layer consisted of BChl molecules No.32,1,2 and of all fragments in a close contact with target BChl No.1 fragment.

The excitation energy of 1.658 eV for the central BChl No.1 fragment was obtained. The FMO2-TDDFT corrections (-0.107eV) shifted the resulting excitation energy to 1.551 eV. This value is very close to the results of the excitonic model [4] obtained with the empirically adjusted parameters for the effective Hamiltonian, 1.48 eV, thus, providing support to our computational scheme.

3. Technical aspects

This application shows good overall scaling of FMO algorithms as illustrated in Table 1. Calculation time (minutes) of the FMO-TDDFT algorithms attainable by Lomonosov-2 super-computer.

It takes little time to achieve convergence for the 1st layer consisting of 172 fragments when using the 3-21G basis set in HF calculations. Correspondingly, this part of the calculation does not benefit from a large number of cores per group (240 vs 480) but thrives from the increased number of groups (3 and 6 vs 1). The 2nd layer consists of only 42 fragments but basic set (6-31G*) is considerably larger which means a greater computational effort per fragment, longer runtime and permits the better intergroup scaling. It should be noted the SCF calculations for monomers in both layers obviously benefit from the increased GDDI group count. The excitation energy calculation (Eq.(4)) for our system requires computations of several large (with 1592-877 basis functions) TD-CAM-B3LYP dimer independent tasks. It takes the greatest walltime (~ 70%), while the scaling is good - 81% for 480 vs 130 cores in a single group.

In this application the nodes were evenly assigned to each group which is an obvious drawback as can be seen from timing numbers for 2nd layer SCF calculations for 480/3 and 480/6 tests. This is due to a low count of the computational tasks per FMO iteration (42). Also there

Table 1. Calculation time (minutes) of the FMO-TDDFT algorithms attainable by Lomonosov-2 supercomputer

Number of cores/groups	Layer1	Layer2	TDDFT	Total
130/1	54.6	210	730.7	995.3
240/1	37.1	127	423.2	587.3
240/3	28.1	109.1	403.6	540.8
480/1	28.2	82.4	243.6	354.5
480/3	16.9	58.9	211.5	287.3
480/6	14.4	89.7	209.5	313.6
720/6	10.4	61.9	142.9	215.2

are five very large BChl fragments that require large calculation efforts while other fragments are considerably smaller. However, GDDI implementation in GAMESS(US) is very flexible and allows one to use different number of groups for different kinds of computational tasks by varying the `ngrfmo` parameter. It also enables manual node division into groups using the `mannod` parameter. Correspondingly, computational scheme should benefit from employing ~ 6 groups with 512-128 cores per group. By using a total of ~ 2000 thousands of CPU cores with almost linear scalability it should take less than 100 minutes to complete the calculation. One should note that GAMESS(US) TDDFT code does not use distributed memory algorithms. The replicated memory demands are quite high for large tasks; in our application, it took up to 2 Gigabytes per core.

As mentioned earlier, in order to fully describe the light-harvesting antenna using the effective Hamiltonian technique 32 independent calculations have to be conducted to obtain all diagonal and most important off-diagonal matrix elements. On the one hand, this scheme enables an efficient use of tens of thousands of processor cores simultaneously, and on the other hand it should still be operational with only several hundreds of cores used for a long period of time. This is a very important issue because large computational clusters are not often available for a user.

The constantly updated TOP500 list reflects innovations and modern tendencies in supercomputing. Remarkably, current No.1 (Tianhe-2) and 2 (Titan) positions are held by supercomputers using Intel Xeon Phi and NVIDIA K20x. Innovations in quantum chemistry software driven by modern supercomputers design enable efficient use of accelerators [11] to carry out the most computationally demanding and complex tasks in molecular modeling. Although our computations are capable to take advantages of large computational clusters with x86-64 processors currently attainable FMO calculations with GAMESS(US) software are unable to efficiently utilize either NVIDIA CUDA or Xeon Phi devices.

Comparing Lomonosov-2 configuration with that of MVS-10P we see that nodes on Lomonosov can provide only 10 CPU cores to a generic application. RAM bandwidth is not saturated by 10 cores (Intel Xeon E5V2, IvyBridge) and better application performance (per node) can be expected on MVS-10P nodes where two sockets provide together 16 cores of Intel Xeon E5 (Sandybridge) on close clock frequency. However the number of nodes is ~ 200 for MVS-10P and ~ 1000 for Lomonosov-2. Both computational clusters use Mellanox FDR Infiniband. MVS-10P nodes have local SSD storage that can be useful in a number of quantum

chemistry methods for temporary storage (ERI data). Our application is not demanding to the parallel file storage, so in this respect the system differences are of minor importance.

Conclusion

This work demonstrates efficiency of novel quantum-based computational algorithms implemented in the modern supercomputers to compute the most difficult properties of large molecular aggregates such as electronic excitations in the photosynthetic light-harvesting complexes. A notable feature of these calculations is a use of the three-level parallelization scheme which allows one to gain advantages of the proper fragmentation of a model system. A good overall scaling of the fragment molecular orbital algorithms is demonstrated.

This work is supported by the Program No.24 of the Presidium of the Russian Academy of Sciences.

This paper is distributed under the terms of the Creative Commons Attribution-Non Commercial 3.0 License which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is properly cited.

References

1. M. Chiba, D. Fedorov, and K. Kitaura. Time-dependent density functional theory based upon the fragment molecular orbital method. *J. Chem. Phys.*, 27, 2007.
2. D. Fedorov, K. Kitaura, E. Starikov, J. Lewis, and S. Tanaka. *Modern Methods for Theoretical Physical Chemistry of Biopolymers*. September 2006.
3. D. Fedorov, R. Olson, K. Kitaura, M. Gordon, and S. Koseki. A new hierarchical parallelization scheme: generalized distributed data interface (gddi), and an application to the fragment molecular orbital method (fmo). *J. Comput. Chem.*, 25:872–880, 2004.
4. X. Hu, T. Ritz, A. Damjanovic, and K. Schulten. Pigment organization and transfer of electronic excitation in the photosynthetic unit of purple bacteria. *J. Phys. Chem.*, 101:3854–3871, 1997.
5. M. Khrenova, A. Nemukhin, B. Grigorenko, P. Wang, and J.-P. Zhang. All-atom structures and calcium binding sites of the bacterial photosynthetic lh1-rc core complex from thermochromatium tepidum. *J. Mol. Model.*, 20:1–6, 2014.
6. K. Kitaura, E. Ikeo, T. Asada, T. Nakano, and M. Uebayasi. *Fragment molecular orbital method: an approximate computational method for large molecules*. November 1999.
7. T. Nagata, D. Fedorov, and K. Kitaura. Linear-scaling techniques in computational chemistry and physics. *Springer*, 13:17–64, 2011.
8. I. Polyakov, L. Grigorenko, A. Moskovsky, Vl. Pentkovski, and A. Nemukhin. Towards quantum-based modeling of enzymatic reaction pathways: Application to the acetylcholinesterase catalysis. *Chem. Phys. Lett.*, 556:251–255, 2013.
9. V. Sadovnichy, A. Tikhonravov, Vl Voevodin, and V. Opanasenko. Lomonosov: Supercomputing at moscow state university. In *Contemporary High Performance Computing: From*

Petascale toward Exascale, Chapman and Hall/CRC Computational Science, pages 283–307, Boca Raton, United States, 2013. Boca Raton, United States.

10. M. Schmidt, K. Baldridge, J. Boatz, S. Elbert, M. Gordon, J. Jensen, S. Koseki, N. Matsunaga, K. Nguyen, S. Su, T. Windus, M. Dupuis, and J. Montgomery. *General Atomic and Molecular Electronic Structure System.*, volume 14. 1993.
11. B. Scott and B. Levine. Nanoscale multireference quantum chemistry: Full configuration interaction on graphical processing units. *J. Chem. Theory Comput.*, 11:4708—4716, 2015.
12. S. Tretiak, C. Middleton, V. Chernyak, and S. Mukamel. Bacteriochlorophyll and carotenoid excitonic couplings in the lh2 system of purple bacteria. *J. Phys. Chem.*, 104:9540—9553, 2000.