Development of efficient computational techniques and codes for second-order Møller–Plesset perturbation calculation of extended systems

Michio Katouda [1]

[1] RIKEN Advanced Institute for Computational Science, Japan

1 Introduction

Electronic structure theory such as ab initio molecular orbital (MO) theory is the powerful tool in elucidating chemical phenomena such as electronic states, molecular structures, properties, and reaction mechanisms. The high-level quantum chemical calculations excellently reproduce the properties for small molecules in the same or better accuracy with the experiments. However, the computational costs drastically increase with respect to the size of molecules when the high-level quantum chemical calculations are performed. To elucidate chemical phenomena of nano molecules and biological molecules, the development of fast and robust theories and computational techniques of the ab initio quantum chemistry is desired. Especially, non-covalent interactions such as van der Waals forces play important roles in the chemical phenomena of such systems. Those interactions come from the electron correlations. The generally accepted density functional theory (DFT) functionals fail to describe non-covalent interactions because they suffer from self-interaction problems and do not incorporate the long-range correlation effect.
Therefore, the electron correlation theory based on the ab initio MO theory is indispensable for the robust reproduction of non-covalent interactions. Among these theories, second-order Møller–Plesset perturbation theory (MP2) [1] is the simplest method to account for the electron correlations at the ab initio level and often used for the practical chemical applications. However, even the computational cost of MP2 calculations scales $O(N^5)$ with respect to the size of molecules (N), and practical applications are limited to molecules of moderate size. To make the MP2 calculations applicable to the extended systems such as nano molecules and biological molecules, development of efficient computational techniques is desired. In this study, we present an efficient parallel resolution-of-identity (RI) MP2 (RI-MP2) algorithm and fragmentation based $O(N)$ parallel RI-MP2 methods. These methods are suitable for the massively parallel computations on the supercomputers such as "K computer".

2 Parallel RI-MP2 Algorithm

We have developed a parallel RI-MP2 algorithm for large molecules [2] and a RI-MP2 method for periodic systems [3]. These methods are based on the RI approximation of four-centre Coulomb integrals, and highly reduce the computational costs. The parallel RI-MP2 algorithm is designed for the efficient parallel calculations by reducing the I/O and network communication overheads and considering the uniform task distribution for the efficient load balancing.
Recently, we have modified our parallel RI-MP2 algorithm from original one. In the original algorithm, the occupied orbital pairs are distributed to processors for the calculation of four-centre integrals. However, the number of occupied orbital pairs is small, and this makes the load balancing problems in the cases of the massively parallel computations. In order to use more large number of CPU cores with the efficient load balancing, we have changed to use the virtual orbital pairs for the parallel task distribution. Generally, the number of virtual orbitals is four times larger than the number of occupied orbitals, and the load valancing is improved from the original algorithm. We have implemented the modified version of parallel RI-MP2 algorithm into GAMESS [4] program and NTChem program. In these implementations, we have performed MPI/Open-MP hybrid parallelization for the efficient usage of the memories and the network devices in the multi-core architectures.
Figure 1 presents the speedups of parallel RI-MP2 calculations. The speedups scale almost liner with respect to the number of CPU cores. These results demonstrate the efficiency of our parallel algorithm and implementations. Using the parallel RI-MP2 codes developed by us, MP2 calculations of large molecules can be performed in modest times with massively parallel supercomputers and low-cost personal computer (PC) clusters.

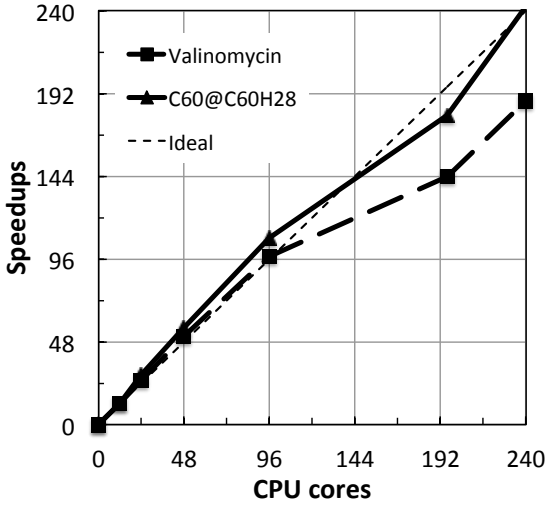

Figure. 1. Speedups of parallel RI-MP2/cc-pVDZ calculations of valinomycin and bucky-catcher $C_{60}@C_{60}H_{28}$. Parallel RI-MP2 calculations were performed on the Xeon X5670 cluster connected with InfiniBand.

3 $O(N)$ RI-MP2 methods for Extended Systems

We have developed efficient implementations that are suitable for the massively parallel computations on the supercomputers. However, even for employing RI-MP2 method, the computational costs of these codes are still $O(N^5)$ with respect to the size of molecules N, and its practical applications are limited to molecules of moderate size. To overcome this limitation of

computational scaling, we have applied the parallel RI-MP2 codes to the fragment molecular orbital (FMO) method [5], the divided-and-conquer (DC) method [6], and the molecular tailoring approach (MTA) [7]. In these methods, a target molecule is divided spatially into small fragments and then the energies and properties are obtained by accumulating contributions of fragmented subsystems. The application of RI-MP2 to those fragment based linear scaling methods considerably speeds up MP2 calculations because the computational scaling and prefactor of RI-MP2 are smaller than those of conventional MP2.

3.1 RI-MP2 with FMO Method (FMO-RI-MP2)

The FMO method is efficient approach for the rapid ab initio quantum chemical calculations of large biological molecules and nanomolecules [8,9]. FMO reduces the computational costs considerably by dividing the target system into fragmented subsystems and performing ab initio quantum chemical calculations for each monomer and dimer (and trimer, if necessary) of fragments. FMO can treat the electron correlation successfully in the MP2 level (FMO-MP2). This opens up many practical chemical applications of FMO to biological molecules such as protein–ligand bindings [8,9].

We have developed a new implementation of FMO-RI-MP2 based on the GAMESS version of parallel RI-MP2 program mentioned in Section 2. In this implementation, we have extended FMO-RI-MP2 to treat the three-body FMO method (FMO3-RI-MP2) as well as the two-body FMO method (FMO2-RI-MP2).

FMO-RI-MP2 considerably speeds up FMO-MP2 calculations of biological molecules without sacrificing chemical accuracy. Table 1 presents the computation times and errors of total energy of Trp-cage protein (PDB ID: 1L2Y). FMO2-RI-MP2 and FMO3-RI-MP2 calculations of 1L2Y protein were finished within 0.5 and 8.2 h, respectively, whereas FMO2-MP2 and FMO3-MP2 calculations took 1.1 and 23.9 h, respectively. The errors of FMO2-RI-MP2 and FMO3-RI-MP2 are at most 3,745 and 0.104 mHartree, respectively.

Recently, we have ported our FMO-RI-MP2 code to K computer and have performed MPI/Open-MP hybrid parallelization. The code was tested for an FMO-RI-MP2 calculation of real protein on K computer up to 86016 CPU cores. The FMO-RI-MP2 code has been supplied as a library program in K computer [10].

Table 1. MP2 energies, errors of MP2 energy, and wall times of FMO-MP2/6-31G* and MP2/6-31G* calculations of 1L2Y protein.

	Energy [H]	Error [mH]	Time [h]
FMO2			
RI-MP2	-7461.560235	3.745	0.5
MP2	-7461.560450	3.530	1.1
FMO3			
RI-MP2	-7461.563877	0.104	8.2
MP2	-7461.564100	-0.120	23.9
Full			
RI-MP2	-7461.563758	0.222	33.8
MP2	-7461.563980		59.2

*cc-pVTZ auxiliary basis set was employed for RI-MP2 calculations. Calculations were performed on 32 nodes of Pentium 4 640 PC cluster connected with the gigabit ethernet.

3.2 RI-MP2 with DC Method (DC-RI-MP2)

The MP2 method based on the DC approach (DC-MP2) is one of the efficient methods for the calculation of large nanomolecules [11]. In DC-MP2, the total system is divided into small nonoverlapping subsystems called central regions. Some atoms that are located around a central region are added to the subsystems to account for the environmental effects that are missing by the division of the total system. The MP2 correlation energy of a total system is obtained as a sum of the MP2 correlation energies of subsystems, which are evaluated exploiting the localized molecular orbitals provided by the DC Hartree–Fock (HF) calculations.

Recently, we have developed a new implementation of DC-RI-MP2 code based on the GAMESS version of parallel RI-MP2 program mentioned in Section 2. In this implementation, we also performed multilevel parallelization of DC-MP2, which is suitable for the calculations of extended systems on the massively parallel supercomputers [6]. The parallelization scheme is a combination of the coarse-grain parallelization assigning each subsystem to a group of processors and the fine-grain parallelization where the computational tasks for evaluating the MP2 correlation energy of the assigned subsystem are distributed among processors in the group. The DC-RI-MP2 code has been ported to K computer and supplied as a library program in K computer [10].

4 Summary

We have developed an efficient parallel RI-MP2 algorithm. We also have developed efficient computational techniques and codes of RI-MP2 based on molecular fragmentation based $O(N)$ quantum chemical methods such as FMO-RI-MP2 and DC-RI-MP2. These methods and codes are suitable for the massively parallel computations of extended systems on the supercomputers. The codes developed in this study have been supplied to users of K computer as a library program [10].

Acknowledgements

This work was supported by the Nanoscience Program in the Next Generation Super Computing Project of the MEXT. Part of calculations were performed at K computer, Kobe, Japan and the Research Center for Computational Science, Okazaki, Japan.

References

[1] D. Cremer, Møller–Plesset perturbation theory: from small molecule methods to methods for thousands of atoms, WIREs Comput. Mol. Sci., 1, 509-530, 2011.

[2] M. Katouda and S. Nagase, Efficient parallel algorithm of second-order Møller-Plesset perturbation theory with resolution-of-identity approximation (RI-MP2), Int. J. Quant. Chem., 109, 2121-2130, 2009.

[3] M. Katouda and S. Nagase, Application of second-order Møller–Plesset perturbation theory with resolution-of-identity approximation to periodic systems, J. Chem. Phys., 133, 184103, 2010.

[4] http://www.msg.chem.iastate.edu/gamess/

[5] M. Katouda, Application of resolution of identity approximation of second-order Møller–Plesset perturbation theory to three-body fragment molecular orbital method, Theor. Chem. Acc., 130, 449-453, 2011.

[6] M. Katouda, M. Kobayashi, H. Nakai, and S. Nagase, Two-level hierarchical parallelization of second-order Møller-plesset perturbation calculations in divide-and-conquer method, J. Comput. Chem., 32, 2756-2764, 2011.

[7] A. P. Rahalkar, M. Katouda, S. R. Gadre, and S. Nagase, Molecular tailoring approach in conjunction with MP2 and RI-MP2 codes: A comparison with fragment molecular orbital method, J. Comput. Chem., 31, 2405-2418, 2010.

[8] D. G. Fedorov, K. Kitaura (eds), The fragment molecular orbital method: practical applications to large molecular systems, CRC Press, Boca Raton, 2009.

[9] D. G. Fedorov, K. Kitaura, Extending the power of quantum chemistry to large systems with the fragment molecular orbital method, J. Phys. Chem. A, 111, 6904-6914.

[10] http://pal.ims.ac.jp/pal/

[11] M. Kobayashi and H. Nakai, Divide-and-conquer approaches to quantum chemistry: Theory and implementation in Linear-scaling techniques in computational chemistry and physics: Methods and applications, Springer, Dordrecht, 97-127, 2011.