

氏 名 石村 和也

学位（専攻分野） 博士（理学）

学位記番号 総研大乙第 178 号

学位授与の日付 平成 19 年 9 月 28 日

学位授与の要件 学位規則第 6 条第 2 項該当

学位論文題目 Development of efficient algorithms for quantum chemistry  
calculations of large molecules

論文審査委員 主 査 教授 小杉 信博  
教授 永瀬 茂  
准教授 信定 克幸  
准教授 柳井 毅  
教授 天能 精一郎（名古屋大学）

Quantum chemistry plays an important role in elucidating molecular geometries, electronic states, and reaction mechanisms, because of the developments of a variety of theoretical methods, such as Hartree-Fock (HF), Møller-Plesset (MP) perturbation, configuration interaction (CI), coupled-cluster (CC), and density functional theory (DFT) methods. Electronic structure calculations have been carried out by not only theoretical chemists but also experimental chemists. DFT is currently most widely used to investigate large molecules in the ground state as well as small molecules because of the low computational cost. However, the generally used functionals fail to describe correctly non-covalent interactions that are important for host-guest molecules, self-assembly, and molecular recognition, and they tend to underestimate reaction barriers. Many attempts have been made to develop new functionals and add semiempirical or empirical correction terms to standard functionals, but no generally accepted DFT method has emerged yet.

Second-order Møller-Plesset perturbation theory (MP2) is the simplest method that includes electron correlation important for non-covalent interactions and reaction barriers nonempirically. However, the computational cost of MP2 is considerably higher than that of DFT. In addition, much larger sizes of fast memory and hard disk are required in MP2 calculations. These make MP2 calculations increasingly difficult for larger molecules. Since workstation or personal computer (PC) clusters have become popular for quantum chemistry calculations, an efficient parallel calculation is a solution of the problem. Therefore, new parallel algorithms for MP2 energy and gradient calculations are presented in this thesis. Furthermore, an efficient algorithm for the generation of two-electron repulsion integrals (ERIs) which is important in quantum chemistry calculations is also presented.

For the calculations of excited states, different approaches are required: for example, CI, multi-configuration self-consistent field (MCSCF), time-dependent DFT (TDDFT), and symmetry adapted cluster (SAC)/SAC-CI methods. One of the most accurate methods is SAC/SAC-CI, as demonstrated for many molecules. In this thesis, SAC/SAC-CI calculations of ground, ionized, and excited states are presented.

This thesis consists of five chapters: a new algorithm of two-electron repulsion integral calculations (Chapter I), a new parallel algorithm of MP2 energy calculations (Chapter II), a new parallel algorithm of MP2 energy gradient calculations (Chapter III), applications of MP2 calculations (Chapter IV), and SAC/SAC-CI calculations of ionized and excited states (Chapter V).

In quantum chemistry calculations, the generation of ERIs is one of the most basic subjects and is the most time-consuming step especially in direct SCF calculations. Many algorithms have been developed to reduce the computational cost. In Pople-Hehre algorithm, Cartesian axes are rotated to make several coordinate

components zero or constant, so that these components are skipped in the generation of ERIs. In McMurchie-Davidson algorithm, ERIs are generated from  $(ss|ss)$  type integrals using a recurrence relation derived from Hermite polynomials. By combining these two algorithms, a new algorithm is developed in Chapter I. The results show that the new algorithm reduces the computational cost by 10 - 40%, as compared with the original algorithms. It is notable that the generation of ERIs including  $d$  functions is considerably fast. The program implemented officially in GAMESS in 2004 has been used all over the world.

In quantum mechanics, perturbation methods can be used for adding corrections to reference solutions. In the MP perturbation method, a sum over Fock operators is used as the reference term, and the exact two-electron repulsion operator minus twice the average two-electron repulsion operator is used as the perturbation term. It is the advantage that the MP perturbation method is size consistent and size extensive, unlike truncated CI methods. The zero-order wave function is the HF Slater determinant, and the zero-order energy is expressed as a sum of occupied molecular orbital (MO) energies. The first-order perturbation is the correction for the overcounting of two-electron repulsions at zero-order, and the first-order energy corresponds to the HF energy. The MP correlation starts at second-order. In general, second-order (MP2) accounts for 80 - 90% of electron correlation. Therefore, MP2 is focused in this thesis since it is applicable to large molecules with considerable reliability and low computational cost.

The formal computational scaling of MP2 energy calculations with respect to molecular size is fifth order, much higher than that of DFT energy calculations. Therefore, less expensive methods, such as Local MP2, density fitting (resolution of identity, RI) MP2, and Laplace Transform MP2, have been developed. However, all of these methods include approximations or cut-offs that need to be checked against full MP2 energies. An alternative approach to reduce the computational cost is to parallelize MP2 energy calculations. A number of papers on parallel MP2 energy calculations have been published. Almost all of them are based on simple parallelization methods that distribute only atomic orbital (AO) or MO indices to each processor. These methods have a disadvantage since intermediate integrals are broadcasted to all CPUs or the same AO ERIs are generated in all processors. Baker and Pulay developed a new parallel algorithm using Saebø-Almlöf integral transformation method. This algorithm parallelizes the first half transformation by AO indices and the second half transformation by MO indices. The advantages are that the total amount of network communication is independent of the number of processors and the AO integrals are generated only once. The disadvantage is the I/O overhead for the sorting of half-transformed integrals. A new parallel algorithm for MP2 energy calculations based on the two-step parallelization idea is presented in Chapter II. In this algorithm, AO indices are distributed in the AO integral generation and the first

three quarter transformation, and MO indices are distributed in the last quarter transformation and MP2 energy calculation. Because the algorithm makes the sorting of intermediate integrals very simple, the parallel efficiency is highly improved and the I/O overhead is removed. Furthermore, the algorithm reduces highly the floating point operation (FLOP) count as well as the required memory and hard disk space, in comparison with other algorithms. Test calculations of taxol ( $C_{47}H_{51}NO_{14}$ ) and luciferin ( $C_{11}H_8N_2O_3S_2$ ) were performed on a cluster of Pentium 4 computers connected by gigabit Ethernet. The parallel scaling of the developed code is excellent up to the largest number of processors we have tested. For instance, the elapsed time for the MP2 energy calculations on 16 processors is on average 15.4 times faster than that on the single-processor.

Determination of molecular geometries and reaction paths is a fundamental task in quantum chemistry and requires energy gradients with respect to nuclear coordinates. In Chapter III, a new parallel algorithm for MP2 energy gradient calculations is presented. The algorithm consists of 5 steps, the integral transformation, the MP2 amplitude calculation, the MP2 Lagrangian calculation, the coupled-perturbed HF calculation, and the integral derivative calculation. All steps are parallelized by distributing AO or MO indices. The algorithm also reduces the FLOP count, the required memory, and hard disk space. Test calculations of MP2 energy gradients were performed for taxol and luciferin on a cluster of Pentium 4 computers. The speedups are very good up to 80 CPU cores we have tested. For instance, the speedup ratios are 28.2 - 33.0 on 32 processors, corresponding to 88% - 103% of linear speedup. This indicates the high parallel efficiency of the present algorithm. The calculation of taxol with 6-31G(d) (1032 contracted basis functions) finishes within 2 hours on 32 processors, which requires only 1.8GB memory and 13.4GB hard disk per processor. Therefore, geometry optimization of molecules with 1000 basis functions can be easily performed using standard PC clusters.

In Chapter IV, several applications of MP2 are performed using the program developed in Chapters II and III. Some molecules that DFT cannot treat well are optimized at the MP2 level. Geometry optimization is also carried out using the spin-component scaled (SCS) MP2 method. In this method, a different scaling is employed for the same and opposite spin components of the MP2 energy, so that SCS-MP2 performs as well as the much more costly CCSD(T) method at a high level of theory.

SAC theory is developed for ground states and based on CC theory that describes higher-order electron correlation. The main factor of electron correlation is collisions of two electrons. In CC theory, most collisions of four electrons can be taken in as the product of collisions of two electrons. Only a symmetry adapted excitation operator is used for the SAC expansion. Since the operator of the SAC expansion is totally symmetric, the unlinked terms (the products of the operators) are also totally

symmetric. SAC-CI is developed to treat excited states. SAC and SAC-CI wave functions are orthogonal and Hamiltonian-orthogonal to each other. These orthogonalities are especially important for the calculations of transitions and relaxations. In general, the SAC-CI operators  $R$  are restricted to single and double excitations. This is called the SAC-CI SD-R method. For the calculations of high-spin states and multiple excitation processes, triple, quadruple, and higher excitation operators are included. This is called the SAC-CI general-R method. In Chapter V, the ground, singlet and triplet excited, ionized and electron attached states of ferrocene ( $\text{Fe}(\text{C}_5\text{H}_5)_2$ ) were calculated using the SAC/SAC-CI SD-R method. The calculated results are in good agreement with experimental values. It is found that shake-up processes (one electron ionization and one electron excitation) contribute to the first two ionization peaks.

## 論文の審査結果の要旨

申請者石村和也氏は2002年3月に京都大学大学院工学研究科修士課程を修了し同年4月より分子科学研究所の技官（技術職員）として研究に従事してきた。研究内容は大規模分子の量子化学計算のための効率的アルゴリズムの開発研究と大規模分子の電子構造研究から成る。小さな分子に対しては精緻な第一原理量子化学計算が可能であり、実験値の定量的予測まで可能であるのに対し、大規模分子に対しては近似的な量子化学計算である密度汎関数法（DFT）が広く使われている。DFT法では必ずしも信頼性のある結果が得られるわけではないため、石村氏は信頼性のある結果を得るためには電子相関まで正しく考慮できる第一原理量子化学計算を実現する必要があると考え、研究を進めた。提出された論文は5章からなる。

第1章は分子積分の高速化アルゴリズムについて書かれている。量子化学計算ではガウス関数による基底関数展開が基本としているが、2電子間の反発積分の数は基底関数の4乗個に比例するため、その演算時間は量子化学計算の律速部分のひとつとなっている。石村氏は従来の方法を比較検討した結果、s関数とp関数だけの場合に最も速いアルゴリズムとして知られているPople-Hehre法にMcMurchie-Davidson法を組み合わせ拡張して、d関数を含む場合にも高速となるアルゴリズムを新たに開発した。

第2章は基底状態分子に対して電子相関を効率的に取り込む方法として知られている2次のMoller-Plesset摂動（MP2）法を並列化するためのアルゴリズムについて書かれている。基底関数ベースの積分から分子軌道ベースの積分に積分変換する操作として軌道数の5乗に収める各種演算アルゴリズムが知られているが、石村氏はMP2計算では2段階の積分変換により重複する計算が必要ないことと並列化及びI/O負荷の軽減を前提に考えることで、高速化アルゴリズムを考案した。

第3章ではMP2法を基本にしてエネルギー微分で構造最適化を行う並列アルゴリズムについて書かれている。基本的には2章で構築したような積分変換アルゴリズムの改良とI/O負荷の軽減によって高速化アルゴリズムを考案した。

第4章では開発したアルゴリズムを組み込んだ量子化学プログラムを利用して大規模二核鉛分子の結晶構造と溶液中の構造について研究している。結晶構造的には鉛間結合が単結合であるのに対し、溶液中の構造は多重結合になりうることを示唆する結果を得た。さらに、MP2法より精度の高い3次までの結合クラスター法CCSD(T)をMP2法の一部をパラメータ化することで再現する方法を白金錯体や芳香族二量体に適用し、良好な結果を得た。このような努力によって現在、3000軌道を越える大規模分子のMP2計算が現実的なものとなり、巨大分子の基底状態の分子構造や電子構造について従来より格段に信頼性の高い結果を得た。

第5章では比較的大きなフェロセン分子の励起状態について結合クラスター法に配置間相互作用を組み合わせた方法(SAC-CI)を適用して従来より信頼性の高い結果を得ている。

以上のように石村氏の研究は、大規模分子の量子化学計算に実用上、重要な貢献を果たすものであり、開発した高速化アルゴリズムの一部は汎用量子化学計算パッケージのひとつである GAMESS にすでに実装されて世界中で広く利用されている。以上、審査委員会全員一致で本申請論文は博士（理学）の学位論文として十分と判断した。