

## 科学研究費助成事業（学術研究助成基金助成金）研究成果報告書

平成 25 年 5 月 27 日現在

機関番号：63903

研究種目：若手研究(B)

研究期間：2011～2012

課題番号：23700037

研究課題名（和文） 量子化学高精度理論に関する大規模並列計算アルゴリズム開発

研究課題名（英文） Development of Parallel Algorithm for Highly Accurate Quantum Chemistry Calculations

研究代表者

石村 和也 (ISHIMURA KAZUYA)

分子科学研究所・理論・計算分子科学研究領域・特任研究員

研究者番号：80390681

研究成果の概要(和文):京コンピュータを始めとするスーパーコンピュータでナノサイズ分子の電子状態とエネルギーを高精度で求める大規模並列量子化学計算アルゴリズムの開発と実装を行い、炭素材料の相互作用計算などを行った。ノード間、ノード内それぞれでの計算負荷と中間データの均等な分散、ノード間通信におけるデータ量削減、行列演算 BLAS ライブラリを用いた効率的な演算により、20000 ノード以上の計算でも非常に高い実行性能と並列性能を示した。

研究成果の概要(英文): I developed and implemented a parallel algorithm of high-accuracy quantum chemistry calculations for nano-sized molecules which efficiently works on supercomputers such as the K computer, and calculated interaction energies of carbon materials. The algorithm shows high performance of computation and parallelization on more than 20000 nodes because of equal distributions of computation and data, a small amount of communication data, and efficient matrix operations with the BLAS library.

交付決定額

(金額単位:円)

	直接経費	間接経費	合計
交付決定額	2,600,000	780,000	3,380,000

研究分野：総合領域

科研費の分科・細目：情報学・ソフトウェア

キーワード：並列処理、分散処理、量子化学、スーパーコンピュータ、電子状態計算

## 1. 研究開始当初の背景

物質が変化する化学反応や合成などの研究において、分子の電子状態を求める量子化学計算は化合物の安定性や物性の解析・制御・設計など様々な面で重要な役割を果たしている。取り扱われる分子サイズは年々大きくなり、大きな置換基を付けることによる立体障害や、分散力などいわゆる弱い相互作用を多用することで、新たな機能を持った物質・反応の設計が期待される。しかし、より大きな分子をより精度良く計算するためには、計算量とメモリ使用量の増加は避けられない。そこで、近似を導入して計算コストを削減する方法と、並列化により計算コストを

分散する方法の2つが提案されている。近似を導入する方法では、計算対象分子が限られる、もしくは並列化には向いていないなどの課題があり、並列化については、従来のアルゴリズムのままでは計算負荷及びデータ分散が効率的にできない場合が多く、新たなアルゴリズムや式の導出が必要になる。

一方、計算機に関しては、消費電力や発熱の問題により1CPUコア当たりの計算性能は頭打ちになり、コア数を増やすことで計算能力を向上させるようになった。今後も利用できるコア数はさらに増えると予想され、この増加はスーパーコンピュータだけではなく、研究室レベルのPCクラスタも同様である。

そのような計算機に対応した並列性能の良い高速計算アルゴリズム及びプログラムが求められている。

私はこれまでに量子化学計算の基礎である Hartree-Fock エネルギー計算について、ノード間を MPI、ノード内を OpenMP で負荷分散をしたハイブリッド並列化アルゴリズムの開発と実装を行った。計算時間の 9 割以上を占める 2 電子クーロン反発項計算での 4 重ループをノード間とノード内それぞれで分散させ、また初期値計算など計算時間の割合が少ない部分もすべて並列化することで、2000CPU コアでも効率的に計算できることを実証してきた。

## 2. 研究の目的

実験結果の解析、計算からの予測を正確に行うためには、より精度の高い方法での計算が必要になる。本研究では Hartree-Fock 計算で得られた分子軌道を使う 2 次の摂動 (MP2) 法、露わに相関した MP2(MP2-F12) 法など高精度理論の大規模並列エネルギー計算アルゴリズム・プログラム開発を行う。MP2 法は、炭素シート間やベンゼン環-水素間などのいわゆる弱い相互作用を取り扱える高精度理論の中で最も計算量が少ない方法である。

京コンピュータでは数万ノード (数十万 CPU コア) の利用が可能になり、その他のスーパーコンピュータでも数千ノードの利用が容易になりつつある。これからの計算機を効率的に使うため、計算全体を並列化し、さらに計算負荷を全ノードでできる限り均等に分けるアルゴリズムを開発する。さらに、これらの方法は計算量だけでなく計算途中のデータ量も多く、データ分散、ネットワーク通信の最適化も必須である。

大規模並列高精度量子化学計算のプログラムができれば、これまで困難であったナノサイズ分子の高精度エネルギー計算が可能になり、フラレンなどの炭素材料や生体分子への適用が可能になる。さらに今回開発するエネルギー計算アルゴリズムは、様々な計算手法への展開することができる。例えば、エネルギー微分計算による安定構造やスペクトルなど物性計算、液体中の化学反応計算のための溶媒効果を取り入れた計算、化学反応の中心部分を量子化学計算で、周囲をパラメータ利用による計算コストの少ない分子力場計算で行う複数の方法を組み合わせる QM/MM 法が挙げられる。このように、本研究はナノサイズ分子の量子化学計算の基盤になると考えられる。

## 3. 研究の方法

MP2 エネルギーは式(1)により計算される。

$$E_{MP2} = \sum_{ijab} \frac{(ai|bj)(2(ai|bj) - (aj|bi))}{\epsilon_i - \epsilon_j - \epsilon_a - \epsilon_b} \quad (1)$$

$$(ai|bj) = \sum_{\mu\nu\lambda\sigma} C_{\mu a} C_{\nu i} C_{\lambda b} C_{\sigma j} (\mu\nu|\lambda\sigma) \quad (2)$$

$(ai|bj)$  は分子軌道 2 電子クーロン反発項、 $\epsilon$  は分子軌道エネルギーである。 $i, j$  は電子が入っている分子軌道、 $a, b$  は電子が入っていない分子軌道である。従来  $(ai|bj)$  は、式(2)のように原子軌道 2 電子クーロン反発項  $(\mu\nu|\lambda\sigma)$  に分子軌道  $C$  を掛けて求めており、このステップが MP2 エネルギー計算時間の 9 割以上を占める。

本研究では、近年天能が提案した求積法を用いた分子軌道 2 電子クーロン反発項計算 (式(3)) と式(1) を組み合わせることで、大規模並列 MP2 エネルギー計算アルゴリズム・プログラムを開発する。

$$(ai|bj) = \sum_{\mathbf{r}} \omega(\mathbf{r}_a) \varphi_a(\mathbf{r}_a) \varphi_i(\mathbf{r}_a) (\varphi_j(\mathbf{r}) | r_{ij}^{-1} | \varphi_b(\mathbf{r})) \quad (3)$$

$$(\varphi_j(\mathbf{r}) | r_{ij}^{-1} | \varphi_b(\mathbf{r})) = \int d\mathbf{r}_1 \varphi_j(\mathbf{r}_1) \varphi_b(\mathbf{r}_1) |\mathbf{r}_1 - \mathbf{r}_i|^{-1} \quad (4)$$

ここで、 $\omega$  は求積点の重み、 $\varphi_a$  は求積点での分子軌道値、 $(\varphi_j | r_{ij}^{-1} | \varphi_b)$  は 3 中心項である。

求積法を導入したアルゴリズムの概要を図 1 に示す。

```
do g = 1, NGrid (MPI parallel)
  $OMP do
  do i = 1, NOcc
     $\varphi_i(\mathbf{r}_g)$  計算+メモリに保存
  enddo
  $OMP do
  do a = 1, NVir
     $\varphi_a(\mathbf{r}_g)$  計算+メモリに保存
  enddo
  $OMP do
  do jb = 1, NOcc*NVir
     $(\varphi_j(\mathbf{r}) | r_{ij}^{-1} | \varphi_b(\mathbf{r}))$  計算
    +メモリに保存
  enddo
  enddo

  mpi_bcast( $\varphi_i(\mathbf{r}_g)$ ,  $\varphi_a(\mathbf{r}_g)$ )
  do ab = 1, NVir*NVir (MPI parallel)
    mpi_sendrecv( $(\varphi_j(\mathbf{r}) | r_{ij}^{-1} | \varphi_b(\mathbf{r}))$ )
    $OMP do
    do ij = 1, NOcc*NOcc
       $(ai|bj)$  計算 + MP2 エネルギー計算
    enddo
  enddo
  mpi_reduce(MP2 エネルギー)
```

図 1 MP2 エネルギー計算アルゴリズム

重要な点は、

- ・式(3)で必要な項に依存関係が無いから、ループ構造がシンプルになり MPI/OpenMP ハイブリッド並列による負荷分散が容易になる
- ・多数の計算機を使用する場合、全体としてのメモリ量は膨大になることから、従来ハードディスクに保存していた中間データを、各ノードのメモリに分散して保存する
- ・式(3)の各項を計算する前半部分と各項を掛け合わせる後半部分を異なるループ構造にして MPI による並列化を 2 段階に分ける。ノード間の多くの通信が 1 対 1 通信になり、全体の通信量はプロセス数に依存せずほぼ一定になる

具体的には、前半は求積点( $g$ )のループが、後半は電子の入っていない分子軌道ペア( $ab$ )のループが最も外になり、どちらも MPI ランクにより各ノードにインデックスが分散される。前半のループ内では、分子軌道( $i, a, jb$ )が、後半のループ内では電子の入っている分子軌道ペア( $ij$ )が OpenMP によりノード内で各 CPU コアに分散される。前半のループ内で、式(3)で必要な項が計算され各ノードのメモリに分散保存される。後半のループが始まる前に  $\varphi_a, \varphi_j$  が MPI\_BCAST によりすべてのノードに配られるが、これらのデータ量は  $(\varphi_j|r^{-1}|\varphi_b)$  に比べて小さい。 $(\varphi_j|r^{-1}|\varphi_b)$  はデータ量が多いため、MPI\_SENDRECV で必要とするノードにのみ送られる。最後に各ノードで求めた部分的な MP2 エネルギーを MPI\_REDUCE でマスターノードに集める。

MPI による送受信を OpenMP 領域の外で行い、アルゴリズムをできる限りシンプルにしながらか MPI/OpenMP ハイブリッド並列を実現する。これにより、Flat MPI に比べてプロセス数が少なく、より負荷分散が均等になるため並列加速率が向上すると期待される。さらにノード内のプロセス数が減り、プロセス当たりの計算用に必要なメモリ量が少なくなるため、中間データ保存用に大きなメモリ領域が使用できる。

MP2-F12 エネルギー計算では、2 電子項だけでなく、3 電子、4 電子項を計算する必要があるが、本研究では 2 電子項と同様に求積法で求める。そのため、計算手順、データ保存方法などアルゴリズムの骨格は MP2 エネルギー計算と同じである。

#### 4. 研究成果

MPI/OpenMP ハイブリッド並列 MP2 エネルギー計算アルゴリズムの実装を行い、SGI Altix ICE 8400EX スーパーコンピュータ(8CPU コア/ノード)8192CPU コアを使って並列性能を測定した。用いた分子は、アントラセン( $C_{14}H_{10}$ )、テトラセン( $C_{18}H_{12}$ )、コロネン( $C_{24}H_{12}$ )、基底関

数はそれぞれ aug-cc-pCV5Z (94 電子、3334 基底)、aug-cc-pCVQZ (120 電子、2514 基底)、aug-cc-pCVTZ (156 電子、1692 基底)である。まず、求積法の精度確認を行った。Medium、Fine グリッドはそれぞれ原子当たり 9216 点、24576 点で、求積法を使わない従来法のエネルギーとの誤差を表 1 にまとめた。Medium グリッドでも誤差は 1mHartree ほどで化学反応などでの議論は十分であり、これ以降の計算では Medium グリッドを使用した。

表 1 各グリッドでの MP2 エネルギー誤差 (mHartree)

	Medium	Fine
テトラセン	0.399	0.095
コロネン	1.704	0.102

256 から 8192 コアまでの MP2 エネルギー計算実行時間とその並列加速率を表 2 に示す。4096 コアまでは、どの分子でも並列性能は高い。8192 コアでは、小さな分子の計算時間の減少幅は小さくなっている。

表 2 MP2 エネルギー計算実行時間(秒)、カッコ内は 256 コアを基準とした並列加速率

分子	CPU コア数		
	256	4096	8192
アントラセン	15180.0 (256.0)	1095.9 (3545.9)	857.2 (4533.7)
テトラセン	15133.8 (256.0)	1063.1 (3644.2)	656.7 (5900.0)
コロネン	14212.9 (256.0)	975.6 (3729.6)	580.7 (6266.1)

この原因を調べるため、演算時間と通信時間それぞれについて表 3 と 4 にまとめた。演算時間については、8192 コアでも並列加速率は 7900 倍以上であり、演算自体はほぼ均等に分散できていることがわかる。

表 3 MP2 エネルギー計算における演算時間(秒)、カッコ内は並列加速率

分子	CPU コア数		
	256	4096	8192
アントラセン	15021.5 (256.0)	947.0 (4060.9)	482.8 (7964.5)
テトラセン	14987.8 (256.0)	934.4 (4106.2)	484.6 (7918.3)
コロネン	14000.6 (256.0)	886.5 (4043.2)	453.0 (7912.5)

通信時間については、2 種類の通信パターン別に調べた。分子軌道値は MPI\_BCAST で全ノードへの送信があるため、ノード(コア)数が増えるにしたがって通信時間は増加している。一方、3 中心項は MPI\_SENDRECV で 1 対 1 通信であり、ノード数が増えるほど 1 回あ

たりのデータ送受信量は減るため、4096 コアまでは時間は減少している。しかし、8192 コアでは 1 回あたりのデータ量が数十 KB 程度まで少なくなったため、データそのものの通信時間よりもレイテンシ(遅延時間)が大きくなり、どの分子でも増加したと考えられる。計算量は基底の 5 乗、通信データ量は基底の 3 乗に比例して増加するため、より大きな分子の計算では通信時間の割合は小さくなり、計算全体の並列性能は向上すると期待される。

表 4 MP2 エネルギー計算における通信時間(秒)

CPU コア数	256	4096	8192
アントラセン			
分子軌道値	31.5	80.5	99.9
3 中心項	125.0	64.5	270.3
テトラセン			
分子軌道値	29.3	77.9	91.8
3 中心項	115.6	48.6	77.5
コロネン			
分子軌道値	24.4	63.8	75.8
3 中心項	187.3	24.2	50.6

さらに、フラレン( $C_{60}$ ) (aug-cc-pCVTZ、3540 基底)の計算は8192 コアで4.8時間であった。この計算では1.9TBの3中心項データが生成されるが、各ノードのメモリに分散保存することで実行可能となった。OpenMPによりノード内のすべてのコアでデータを共有することで、メモリを有効に利用することができたためである。MPIとOpenMP両方を使うことで初めて実行が可能になったと言える。

京コンピュータ24576ノード(196608コア)で( $C_{150}H_{30}$ )<sub>2</sub>のMP2エネルギー計算(1860電子、Dunning-DZ(3120基底))を行ったところ、16.6分で計算が完了し、実行性能28%、並列化効率80%を達成した。

MP2-F12 エネルギー計算について、京コンピュータで $C_{60}$ 分子(cc-pVTZ(1800基底))を用いて並列性能を測定した。表5のように65536コアでは9分弱で計算が終了し、並列性能も高く、MP2-F12計算が実用的になったことを実証した。

表 5 MP2-F12 の計算時間(秒)と並列加速率

CPU コア数	8192	32768	65536
計算時間	2805.0	812.6	524.1
(並列加速率)	(8192.0)	(28279.2)	(43844.7)

今後は、炭素材料や生体分子などの精密な電子状態計算を行い、新たな物質や反応の予測につなげていく予定である。

## 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者に

は下線)

[雑誌論文](計1件)

(1) Kazuya Ishimura, Seiichiro Ten-no, MPI/OpenMP hybrid parallel implementation of second-order Møller-Plesset perturbation theory using numerical quadratures, Theoretical Chemistry Accounts, 査読有, Vol.130, 2011, pp. 317-321.

DOI : 10.1007/s00214-011-0980-2

[学会発表](計8件)

(1) 石村和也、量子化学超並列計算基盤プログラムの開発、第3回計算物質科学イニシアティブ研究会、2012年12月3日、自然科学研究機構岡崎コンファレンスセンター(愛知県)

(2) Kazuya Ishimura、Development of Massively Parallel algorithm for MP2-F12 calculations、Theory and Applications of Computational Chemistry-2012、2012年09月04日、University of Pavia(イタリア)

(3) 石村和也、量子化学計算の並列化に向けた取り組み、日本応用数理学会 2012年度年会、2012年08月31日、稚内全日空ホテル(北海道)

(4) 石村和也、量子化学計算の高速化と並列化手法、近畿化学協会コンピュータ化学部会第84回例会、2012年06月20日、大阪産業創造館(大阪府)

(5) 石村和也、大規模量子化学計算手法の開発、第4回データ工学と情報マネジメントに関するフォーラム、2012年3月4日、シーサイドホテル舞子ピラ神戸(兵庫県)

(6) Kazuya Ishimura、Parallelization and Acceleration for Large-Scale Quantum Chemistry Calculations、International Workshop on Quantum Chemistry Massively Parallel Programming Now in Supercomputers、2012年2月28日、東京大学駒場ファカルティハウス(東京都)

(7) 石村和也、大規模量子化学計算に向けた取り組み、CMSI若手技術交流会 第一回、2011年7月7日、計算科学研究機構(兵庫県)

(8) 石村和也、天能精一郎、求積法を用いた超並列MP2計算手法の開発、第14回理論化学討論会、2011年5月12日、岡山大学創立五十周年記念館(岡山県)

## 6. 研究組織

(1) 研究代表者

石村 和也 (ISHIMURA KAZUYA)

分子科学研究所・理論・計算分子科学研究領域・特任研究員

研究者番号：80390681