

PC クラスタによる大規模分子軌道計算(1)

— 分散メモリ上での効率的な Fock 行列生成と対角化 —

(NEC ソフト、NEC 基礎研*)○山本純一、坂倉耕太、山口直人、中田一人、高田俊和*

【序】 比較的小さな分子の非経験的分子軌道計算は従来から多数行われており、バイオ関連では大きな分子の計算もハイエンドコンピュータを使用して試みられるようになってきた。しかしながら、創薬への実際の応用には、タンパク質などの生体高分子のより大規模な計算を必要とし、これを目指し分子軌道計算の大型化が進んでいる。

大規模計算には高性能コンピュータが必要になるが、最新のコンピュータを利用しても、数千原子規模の非経験的分子軌道計算は困難な状況である。また、ハイエンドなコンピュータは一般に高価であり、コスト面から利用が制約されることもある。

現在、このような課題を克服するため、低価格で且つ高性能なシステムとして PC クラスタが利用され始めている。これは、著しい価格低減と性能向上により、低価格でありながら、ハイエンドクラスの性能を出せるコンピュータとして期待されているからである。しかしながら、PC クラスタは小規模なメモリを分散して搭載しており、これが大規模な Hartree-Fock 計算への適応の障害となっている。

Hartree-Fock 計算では、Fock 行列と密度行列をメモリ上に保持しなければならない。Fock 行列は比較的容易に分散保持できるため、これを利用した並列化アルゴリズムは多数提案されている[1-3]。しかし、Fock 行列の1要素の計算に、密度行列の全要素を必要とするので、通常密度行列全体を1台のマシンのメモリ上に保持することが求められる。即ち、PC クラスタでは、PC1台のメモリ容量によって計算規模全体が限定されてしまうことになる。

この問題を解決するため、Fock 行列及び密度行列を分散保存し、計算規模が PC クラスタシステム全体のメモリ容量に依存する Hartree-Fock 解法を開発したので、報告する。

【方法】 本研究では、行列計算ライブラリとして ScaLAPACK、通信ライブラリとして BLACS を利用している。まず、ScaLAPACK のアルゴリズムに従って、Fock 行列及び密度行列を2次元サイクリックブロックに分割した。Fock 行列要素の計算には密度行列の全要素が必要であるが、部分密度行列を巡回的にプロセッサ間で通信することで、メモリの問題を解決した。また、Fock及び密度行列を分割・保存したことにより、基底関数の交換対称性による2電子積分の計算量削減が困難になるが、密度行列の転送アルゴリズムの改良と保持する部分行列を増やすことにより、計算量を削減することに成功した。保持する行列の増加により、必要なメモリ容量も増加するが、増加した行列も分割しているため、このメモリ増はPC1台当たりでは問題にならない程度である。また、並列化率の向上のため、計算すべき2電子積分の数と基底関数の

展開項数を考慮して、粒度をできるだけ均一に分散・処理する並列化アルゴリズムも開発している。

【結果】 これまでの所、本アルゴリズムによるテストを2500軌道まで行い、アルゴリズムが有効であることを確認しているが、今後計算規模を拡大して、計算方法の信頼性を確認していく予定である。必要なメモリ容量は、100台で分割した場合、1台あたり1000軌道で20MB、2500軌道で80MB 程度である。このメモリ量の内訳として、二電子積分の値を格納する配列が殆どを占めていることが判明しており、アルゴリズムの改良により更に削減することが可能であると考えている。従って、大規模なHartree-Fock計算で必要とされるメモリ容量の問題は克服できたと考えている。以下に1000軌道のベンチマーク結果を示す。

ノード数	SCF 時間(sec)	速度向上比	並列化率 (%)
1	20488.65		
4	5392.65	3.80	98.24
9	2606.59	7.86	98.19
16	2467.54	8.30	93.82
30	1572.44	13.03	95.51
64	519.32	39.45	99.01
121	433.80	47.23	98.70

これは SCF 1 回分の計測結果である。並列化率が伸びないのは、テスト環境が 100Base Ethernet であることと、PC間に設置したHUBによる通信性能の劣化が原因であると考えており、Myrinetなど高速通信環境でのテストを現在検討中である。

【考察】 今回開発した方法により、PC クラスターのマシン1台のメモリサイズにより、計算規模が制限されることはなくなり、廉価な分散メモリの有効活用が可能になった。創薬や機能性食品の開発に、非経験的分子軌道計算への期待が高まっているが、本手法によりタンパク質などの生体高分子の計算が、PC クラスターで簡便に行えるようになると期待している。

【参考文献】

1. Foster, I. T., Tilson, J. L., Wagner, A. F., Shepard, R. L., Harrison., R. J., Kendall, R. A., and Littlefield. R. J., *J. Comp. Chem.*, **17**, 109-123 (1996).
2. Furlani, T.R., King, H.F., Implementation of a parallel direct SCF algorithm on distributed memory computers, *J. Comp. Chem.*, **16**, 91-104 (1995).
3. Hajime Takashima, So Yamada, Shigeru Obara, Kunihiro Kitamura, Shinjiro Inabata, Nobuaki Miyakawa, Kazutoshi Tanabe, Umpei Nagashima, *J. Comp. Chem.* **23**, 1337-1346 (2002)