Direct SCF 法と通常 SCF 法の速度比較に関する研究

大田原一成† 下川信祐† 寺前裕之‡*
ATR 適応コミュニケーション研究所† NTT コミュニケーション科学基礎研究所‡

はじめに

Hartree-Fock 分子軌道法は 1970 年代頃から色々な分子の電子状態計算に適用されるようになり、今日では計算機実験として一つの分野を確立するまでになりつつある。より現実的な大きさの分子への適用を考えた場合に Hartree-Fock 分子軌道法では、分子のサイズが大きくなると急激に増加する二電子積分をどのように取り扱うかが大きな問題となる。各分子軌道が基底関数により、

$$\psi_i = \sum_{r=1}^{N} c_{ri}\phi_r$$

と表されるとすると、Hartree-Fock の方程式は、

$$\sum_{s=1}^{N}(F_{rs} - \varepsilon_i S_{rs})c_{ri} = 0$$

$$S_{rs} = \int \phi_r(1)\phi_s(1)dr$$

$$F_{rs} = H_{rs} + \sum_{t=1}^{N}\sum_{u=1}^{N} P_{tu}[2 < rs \mid tu > - < rt \mid su >]$$

と表される。ここで

$$H_{rs} = \int \phi_r^*(1)\left[-\frac{1}{2}\left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial y_1^2} + \frac{\partial^2}{\partial z_1^2}\right) + \sum_{A=1}^{atom}\frac{Z_A}{r_{1A}}\right]\phi_s(1)dr$$

はコアハミルトニアンで一電子積分である。

$$< rs \mid tu >= \iint \phi_r^*(1)\phi_s(1)\frac{1}{r_{12}}\phi_t^*(2)\phi_u(2)dr_1 dr_2$$

が二電子積分であり、添字の $rstu$ はそれぞれ 1 から基底関数の数 N までの値を取るので、全体としての二電子積分の数は N^4 のオーダーとなる。ここで

$$P_{tu} = \sum_{i=1}^{occ.} c_{ti}c_{ui}$$

であって、解くべき変数を含んでいるため Hartree-Fock 方程式は繰り返し演算(SCF 演算)を行って解く必要があり、Fock の行列要素 F_{rs} の計算に、二電子積分を複数回利用する必要がある。

Gaussian70 や Hondo5 のような初期のプログラムでは、二電子積分をファイルに書き込み、SCF 過程における繰り返し計算ではそのファイルを読み込んで利用するようにプログラムされていた[1]。これを通常 SCF 法と呼ぶことにする。一方で増大する計算の規模に応えるべく、この二電子積分をファイルに書き込まず毎回再計算を行う方法が direct SCF 法として提案された[2]。当時はハードディスクディスク容量が足りずに不可能であった計算が時間さえかければ可能となったために急速に普及した。また direct SCF 法では必然的にハードディスクへのアクセスが無くなるために、CPU 使用時間では不利になるものの入出力の待ち時間を含めた経過時間では通常 SCF 法よりも有利となる事が多いようである。現在の分子軌道プログラムは概ねこの direct SCF 法を使用するようになっている。

近年のコンピューター本体や周辺機器の性能向上と価格低下により、パーソナルコンピュータークラスターを使用し、Hartree-Fock 分子軌道計算においても上述の二電子積分ならびに Fock の行列要素 F_{rs} の計算に並列処理を利用する事が可能となってきた。並列処理を行うことにより二電子積分は各ローカルのハードディスクに分割されて保存されるので全体としての容量が増加し、また入出力も分割されたファイルに対して行われるのでこれも並列入出力となり負荷が大幅に軽減される。またオペレーティングシステムにより、ファイルの入出力はメモリー上にバッファーされるので、極端な場合にはファイルの入出力は各１回の書き込みと読み込みのみ行われて、残りの読み込みは実際には行われずにメモリー上に保持されたまま行われるので、非常に高速な処理となる。この場合にはプログラムの書き直しを行う必要もなく簡便に高速化が達成できることとなる[3]。

以上のように、入出力の待ち時間を含めた経過時間においても direct SCF 法の優位性が薄れてきているのではないかと思われ、本研究では、中規模の PC クラスター環境を使用した並列処理計算において 217-274 基底関数までの Hartree-Fock 分子軌道計算を行い、direct SCF 法と通常 SCF 法で経過時間の比較を行ったので報告する。

A comparison study on computational time between direct SCF and conventional SCF methods.
† OHTAWARA Kazushige, SHIMOGAWA Shinsuke, ATR Adaptive Communications Research Laboratories
‡ TERAMAE Hiroyuki, NTT Communication Science Laboratories

計算方法

表 1 に並列計算環境を示す。16 筐体で 16CPU からなる PC クラスターを用いて計算を行った。分子軌道計算には GAMESS[4]を使用した。並列化のライブラリは GAMESS 付属の物をそのまま使用し、並列処理自身も GAMESS の元のコードのままとした。

速度の測定に用いた分子は、実際に市販されているベンゾジアゼピン系およびチエノジアゼピン系の向精神薬で、フルトプラゼパム($C_{19}H_{16}ClFN_2O$)、トリアゾラム($C_{17}H_{12}Cl_2N_4$)、クロチアゼパム($C_{16}H_{15}ClN_2OS$)、エチゾラム($C_{17}H_{16}ClN_4S$)、フルタゾラム($C_{19}H_{18}ClFN_2O_3$)、ロラゼパム($C_{15}H_{10}Cl_2N_2O_2$)とした。基底関数には 3-21G 基底を使用した。計算時間の計測は SCF 計算およびエネルギー勾配の一点計算を各 10 回行い、最も速い値を採用した。

結果と考察

各分子の原子数、基底関数の数、計算に必要であった二電子積分の容量を表 2 にまとめた。二電子積分の容量が 1.5 ギガバイトを超え、単一の CPU で計算を行う場合には搭載メモリーを越える値となっている分子もあるが、並列処理環境下では 2 台程度以上の PC クラスター環境で、メモリー上での処理が可能な値となっている。

並列処理のノード数を変化させて計測した経過時間は、いずれの分子の場合にも、適当なノード数となってメモリー上に二電子積分が展開されると、direct SCF 法に比べて通常 SCF 法の方がより経過時間が短くなった。

特に、最も規模が大きい計算となるフルタゾラムの場合は、1 ノード時にはメモリー上に展開されずファイル入出力を伴うために direct SCF 法での 635 秒に比べて 1559 秒と約 2.46 倍の時間を要したが、4 ノード時ではオペレーティングシステムによりファイル入力がメモリー上に展開され物理的なファイルの入出力を伴わなくなるために、逆に 192 秒に対して 141 秒と約 0.73 倍と 27%程度の経過時間の縮小となった。ノード数が 4 以上の場合も同様に通常 SCF 法で経過時間が短縮される。1 回あたりの計算が約 15 秒程度（16 ノード時）の短縮であったとしても、我々が使用している Hartree-Fock 分子軌道法を用いた分子ダイナミクスのように 1000 回の繰り返し計算では、約 4.2 時間となり決して無視できない値となる。計算規模がより大きくなった場合にはノード数の増加によって対応することができる。現在では、例えば 10 ギガバイトのメモリーと 1 テラバイトのディスク容量を持った PC クラスターを構築するのは困難ではない。今後、並列計算が主流となってくれば、再び通常 SCF 法が主流となってくるかもしれない。

今回の計算では経過時間の短縮は 20%程度に留まったが、CPU 占有時間を比べると通常 SCF 法では direct SCF 法に比べて相当小さいため、さらなる計算時間短縮の可能性があると思われ、今後の検討課題である。また今後のハードディスクの高速化やデータバスの高速化によって、自然と短縮される可能性もあることを指摘しておきたい。

以上のように本研究では PC クラスターを用いた並列計算環境で、Hartree-Fock 分子軌道法における direct SCF 法およびファイルの入出力を伴う通常の SCF 法の経過時間の比較を行った。計算対象として、向精神薬である 6 種類の薬剤の分子を使用し、3-21G 基底を用いて計算を行った。最も規模の大きいフルタゾラムの場合で、16CPU を用いた場合にはファイル入出力がメモリーにバッファーされるために、direct SCF 法に比べ、約 20%程度通常 SCF 法が速く計算されることがわかった。

本研究のうち、ATR での研究は通信・放送機構の研究委託により実施したものである。

表 1.　PC クラスターの構成

CPU	16CPU Pentium4 2.0GHz, 512K Cache
Memory	1.5GB SD-RAM / board
Hard Disk	120GB / 5400rpm
OS	Linux kernel 2.4
Compiler	Intel FORTRAN ver.6.0
Parallel Library	MPICH ver.1.2.4

表 2.　各分子の基底関数数と二電子積分容量

[分子名]	[分子式]	[基底関数数]	[二電子積分]
フルトプラゼパム	$C_{19}H_{16}ClFN_2O$	248	1.5GB
トリアゾラム	$C_{17}H_{12}Cl_2N_4$	239	1.3GB
クロチアゼパム	$C_{16}H_{15}ClN_2OS$	227	1.0GB
エチゾラム	$C_{17}H_{16}ClN_4S$	245	1.2GB
フルタゾラム	$C_{19}H_{18}ClFN_2O_3$	274	2.9GB
ロラゼパム	$C_{15}H_{10}Cl_2N_2O_2$	217	0.9GB

参考文献

[1] M. Dupuis, J. Rys, and H. F. King, *J. Chem. Phys.*, **65**, 111 (1976).
[2] J. Almlf, J. K. Faegri, and K. Korsell, *J. Comput. Chem.*, **3**, 385-399 (1982)
[3]H. Teramae and K. Ohtawara, *J. Chem. Software*, **8**, 55-61 (2002)
[4] M. W. Schmidt, et.al., *J. Comput. Chem.* **14**, 1347-1363 (1993)