

# グローバルコンピューティング環境による汎用MCSCFソルバーの開発

日本電気株式会社 基礎研究所 ○高田 俊和

Development of an MCSCF Program System for Global Network Computing Environments

Toshikazu Takada, Fundamental Research Laboratories, NEC Corporation

Since chemical reactions are widely used in productions of industries even in semi-conductor device processes, it is strongly needed to develop a molecular simulation package that is able to analyze the chemical reaction mechanisms based on quantum mechanics. In this project, therefore, an MCSCF (Multi Configuration Self Consistent Field) code with energy gradient calculations is newly developed for parallel computers. Furthermore, to make it easy to use the code, a GUI is developed as well, by which one makes input data, carries out calculations and analyzes results through Internet.

## 1. はじめに

産業における製造プロセスがプレスなどの物理的加工から、半導体の微細加工に見られるように、化学反応を利用した化学的加工へと工業技術の高度化と共に移行している。バイオインフォマティクスやナノテクノロジー分野における将来的活用もスコープに入れながら、このような化学反応のメカニズムを原子・分子レベルで解明できるシミュレーションシステムを開発することは、新素材の開発や生産工程の効率化などに極めて有効である。量子力学に則り、化学反応をシミュレートするには、化学反応の進行と共に変化する電子配置の混合を許すMCSCF (Multi Configuration Self Consistent Field) 法による計算が不可欠である。しかしながら、既存のMCSCFプログラムの適用範囲は、アルゴリズム的な制約から小型分子に限られる場合が多く、新素材開発を行っている研究現場での要請に応えられる段階には至っていない。

一方、コンピュータのハードウェア性能の向上は、並列化をベースに進められている。また、インターネットの著しい普及に伴い、シミュレーションを含めコンピュータの利用形態が大きく変化している。これらのハードウェア動向に着目しながら、本研究開発では、1) CAS-SCF (Complete Active Space SCF) 法により大型分子のMCSCF計算をヘテロジニアスなコンピュータ環境で実行する数値計算部の開発、2) インターネットを介して計算結果を複数の研究者が会話を通して解析できるステアリングシステムの開発、の2点を中心に行った。このようなシステムはまだ報告されておらず、分子シミュレーションシステムの将来的イメージを世界に先駆けて提案できたと考えている。

## 2. MCSCF理論の概略とプログラム開発

MCSCF計算スキームについて、その概略を述べる。MCSCF法におけるエネルギー表式は、

$$E = \sum_{ab} \gamma_{ab} h_{ab} + \frac{1}{2} \sum_{abcd} \Gamma_{abcd} (ab | cd) \quad (1)$$

で与えられる。ここで、 $h$  は1電子積分、 $(ab|cd)$  は2電子積分、 $\gamma$  と  $\Gamma$  は求めるべき解である。また、エネルギー勾配、即ち原子核に働く力は、(1) 式を原子核の座標  $q$  で微分して、次式のように得られる。

$$\frac{\partial E}{\partial q} = \sum_{ab} \gamma_{ab} \frac{\partial h_{ab}}{\partial q} + \frac{1}{2} \sum_{abcd} \Gamma_{abcd} \frac{\partial (ab | cd)}{\partial q} \quad (2)$$

電子配置関数の生成には、symmetric group approach (SGA) を適用した。即ち、

$$\psi_{SM,k}^L = \frac{1}{\sqrt{N!}} \sum_P \varepsilon_P P \{ \Phi_L^0 \Theta_{SM,k}^N \} \quad (3)$$

となり、MCSCF法における configuration interaction (CI) 行列は、

$$H_{lk}^{LR} = \langle \phi_{SM,k}^L | H | \phi_{SM,l}^R \rangle \quad (4)$$

となる。

並列コンピュータでの高速処理を実現するには、2電子積分の原子軌道基底から分子軌道基底への変換、

$$(ab|cd) = \sum_r^N \sum_s^N \sum_t^N \sum_u^N c_{ra} c_{sb} c_{tc} c_{ud} (rs|tu) \quad (5)$$

の並列処理が不可欠であり、並列化アルゴリズムを開発した。エネルギー勾配計算においては、次式のように分子軌道基底の  $\gamma$  及び  $\Gamma$  を原子軌道基底に逆変換することで並列化を実現した。

$$\frac{\partial E}{\partial q} = \sum_{rs}^{AO} \gamma_{rs} \frac{\partial h_{rs}}{\partial q} + \sum_{rstu}^{AO} \Gamma_{rstu} \frac{\partial (rs|tu)}{\partial q} - \sum_{rs}^{AO} W_{rs} \frac{\partial S_{rs}}{\partial q} \quad (6)$$

以下に、開発したプログラムのベンチマーク結果を示す。

### 3. 開発経過

CASSCF法による全エネルギーと原子核に働く力について、既存プログラム GAMESS (米国製) の結果と比較しながらデバッグを行った。まず、全エネルギー (原子単位) であるが、

分子	GAMESS	本プロジェクト
H <sub>2</sub> O	-75.9615241386	-75.9615241386
H <sub>2</sub> CO	-113.8014219135	-113.8014219135
NaOH <sub>2</sub>	-237.8331721124	-237.8331721124

のようにハードウェアの精度で完全に一致している。この結果から、CI行列要素及び一般化Fock行列の計算など数値計算のコアになる部分について正しく動作していることが確認された。

一方、大型分子のMCSCF計算を実現するためには、(5)式の変換を効率的に行うことが不可欠である。この積分変換の並列化における困難さは、1) 原子軌道基底での2電子積分は相互に独立なので、積分計算自体は容易に並列化できるが、分子軌道基底での2電子積分をひとつ計算するのに全ての積分が必要なため、膨大な通信がプロセッサ間で発生し著しい通信ネックとなる。2) 通信ネックを避けようとするれば、原子軌道基底での2電子積分全てをプロセッサ毎に計算しなければならず、並列計算による性能向上は望めない。という相反する問題に遭遇することである。詳細は成果報告書を参照して頂きたいが、この変換をパラレル処理するアルゴリズムを考案し、並列化率についてのベンチマークを行った。測定環境は、1GBローカルメモリ搭載のPentiumIII (933MHz) プロセッサ33台、OSと通信ライブラリはRedHat Linux 6.2とLAM/MPI 6.5.4である。33台のCPUの内、1台をマスタ、残りをスレーブとして、並列化率の測定を行っている。その結果は、

プロセッサ	1	2	4	8	16	32
401原子軌道						
経過時間(秒)	55330	27750	13830	6998	3609	1823
並列効率	1	1.99	4.00	7.91	15.33	30.35
並列化率(%)	-	99.69	100.01	99.83	99.71	99.82

となり、極めて効果的な並列処理であることが確認された。尚、分子は、dobambc (C<sub>31</sub>H<sub>43</sub>NO<sub>3</sub>) である。既存のCASSCFプログラムでは、200~300軌道の計算が通常行われている程度である。本システムを用いて、クロロフィル2量体 (166原子、684電子、994軌道) の計算を現在遂行している。我々の知る限り、ポストHartree-Fock法によるこの規模の計算は報告されておらず、当初の目的を達成できたと考えている。

エネルギー勾配計算部については、2次の密度行列の分子軌道基底から原子軌道基底への逆変換により、(6)式の2電子積分の各座標による微分を、通常の2電子積分と同じように並列処理できるアルゴリズムを開発した。これにより、波動関数計算部同様、効率的な並列処理が可能になった。CASSCF全エネルギーとグラディエントの値を、GAMESSの計算結果と照合すると、

	GAMESS			本プロジェクト		
H <sub>2</sub> O 全エネルギー						
エネルギー勾配	0	0.0000000	0.0666227	0.0000000	0.0000000	0.0666232
H	-0.0547396	0.0000000	-0.0333113	-0.0547400	0.0000000	-0.0333116
H	0.0547396	0.0000000	-0.0333113	0.0547400	0.0000000	-0.0333116

となり、エネルギー勾配計算部も正常に動作していることが確認されている。

#### 4. ネットワークの活用

シミュレーション結果の解析には、分子グラフィックスは必須である。既存の分子グラフィックスの多くが、高速CGエンジンを搭載しているグラフィックスワークステーションにグリッドデータを送り、回転や拡大をローカルに処理している。この方法では転送すべきデータ量が多く、スーパーコンピュータとグラフィックス端末が近接していなければならない。分子シミュレーションは今後インターネット環境下で行われるので、サーバ上でピクセルデータに変換して画像圧縮技術により転送データ量を削減する手法が、最も適していると考えられる。また、異なる分野の研究者が、同一の分子の性質を色々な観点から研究することは良くある。以上の点を考慮し、多地点の研究者がインターネットを介して、MCSCFの計算結果をグラフィックス画面と会話とにより解析できるGUIシステムを開発した。図1は、そのGUIのブラウザ画面である。協力研究者のそれぞれの拠点から、本システムを稼動し、会話による解析の実験を行った。その結果、本プロジェクトで想定した機能が実現されていることを確認した。計算科学で得られる情報は、将来のITの重要なコンテンツになると考えられ、企業における研究活動や高校・大学での理科教育に、このようなシミュレーションシステムが活用されると期待している。

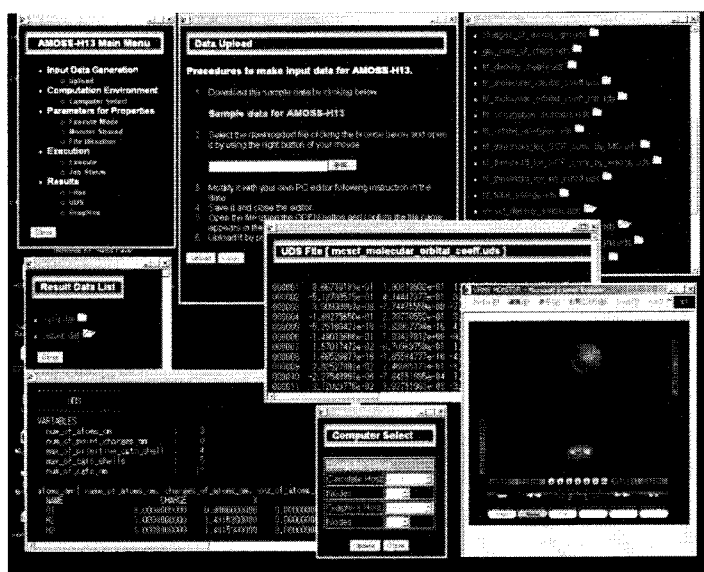


図1 GUIのブラウザ画面

## 5. 今後の課題

本プロジェクトでは、並列コンピュータで効率よく稼動するCASSCFのプログラムと、インターネット環境下で計算を簡便に処理するステアリングシステムの開発を、主たる課題としてきた。何れの開発においても、当初計画した機能を基本的には実現できたと考えているが、実際開発してみて初めて顕在化した問題もある。例えば、分散処理に対応するため、プロセス間通信によりジョブをspawnしているが、その通信が遅く経過時間において大幅な遅延を引き起こしている。本格的な並列分散処理に向うには、ハードウェアを開発している研究者との交流も含め、本プロジェクトで開発した数値計算部の注意深いチューニングが必要と考えている。

## 6. 研究実施体制

研究開発題目：汎用MCSCFソルバー及びインターネットGUIの開発

研究開発題目：汎用MCSCFソルバー及びインターネットGUIの開発

氏名	所属	役職	研究開発項目
高田 俊和	日本電気基礎研究所	主席研究員	MCSCF計算部設計
津田 健一郎	日本電気基礎研究所	研究員	分子グラフィックス

研究開発題目：MCSCF分散処理のための基本アルゴリズム開発及びMnetへの接続によるプログラム評価

氏名	所属	役職	研究開発項目
関口 智嗣	工技院電総研	主任研究官	分散処理アルゴリズム
建部 修見	工技院電総研	研究員	分散処理アルゴリズム

研究開発題目：IMnetへの接続によるプログラム評価

氏名	所属	役職	研究開発項目
柏木 浩	九工大情報工学部	教授	IMnetによる性能評価

研究開発題目：IMnetへの接続によるプログラム評価

氏名	所属	役職	研究開発項目
日向寺 祥子	東海大学計算センター	講師	IMnetによる性能評価