

1C17 並列処理による分子構造最適化計算の高速化—グリッドコンピューティング活用に向けて—

(情報通信研究機構) ○奥野好成, 益子信郎

【序】 シミュレーションによって分子の安定構造を求めることは、ナノテクノロジー分野のみならずバイオテクノロジー分野等においても重要である。ところが、化学分野でのシミュレーションの中核を成す量子化学計算によって、大きな（原子数の多い）分子の安定構造を求めるのは、余りに計算時間がかかり困難である。

計算を高速化させる一つの方法は、並列処理を活用することである。特に、最近、インターネットを介して多数のパソコンを連携させて大規模な計算を可能にするグリッドコンピューティングの活用が容易になってきており、並列処理による超高速計算への期待が高まりつつある。

しかしながら、量子化学計算での分子の構造最適化のアルゴリズム自身を並列処理させる研究は、あまりなされていない。従来の分子構造最適化は準ニュートン法に基づいて逐次的に行うアルゴリズムであり、並列処理できなかつたからである。即ち、準ニュートン法では、試行構造 \mathbf{q}_k でのエネルギーの原子座標による一次微分 \mathbf{g}_k とその一次微分の変化に基づき更新していく近似ヘシアン \mathbf{B}_k を用いて、 $\mathbf{q}_{k+1} = \mathbf{q}_k - \mathbf{B}_k^{-1}\mathbf{g}_k$ (式1) に従って、新規試行構造 \mathbf{q}_{k+1} を得る作業を繰り返しながら行うため、新規試行構造は旧試行構造の計算が完了しないと決められなかつたからである。

そこで、本研究では、ネットワーク接続された、複数の異機種計算機を利用して、分子構造最適化を並列処理させるアルゴリズムの開発を試みた。

【アルゴリズム】 本研究では、以下に示すようなアルゴリズムを検討した (図参照)。
ステップ1：ホスト計算機にて多数の初期試行構造を発生させ、それらの情報を、計算を実行する多数の計算機に転送させる。ステップ2：全ての計算機でエネルギー及びエネルギーの原子座標による一次微分の量子化学計算を実行させる。ステップ3：いずれかの計算機で計算が終了したら、直ちにその結果をホスト計算機に転送させる。ステップ4：ホスト計算機にて、送られてきた計算結果から構造最適化の収束条件を満たしているかどうか判断させ、満たしていれば、最適化を終了させる。満たしていなければ、今までの計算結果の情報を基にヘシアンを更新させ、更新したヘシアンと今まで計算した構造の中でエネルギーが最も低い構造から、新規試行構造を発生させる。ステップ5：新規試行構造の情報を、ホスト計算機から、計算結果を送ってきた計算機に返送させる。ステップ6：送り返された新規試行構造について、エネルギー及びエネルギーの原子座標による一次微分の計算を実行させる。ステップ7：ステップ3に戻る。

ステップ4における、ヘシアンの更新法及び新規試行構造の決定法については、種々

試行錯誤を行った。その結果、以下に示すような手法を用いれば、高い並列化効率を得られることが分かった。1) ヘシアン更新に BFGS 法に類似の式を用いる。2) 更新に用いる 2 つの参照構造は、直前に計算された試行構造と計算した全試行構造中で最安定な構造若しくは直前に計算された構造の最近傍にある構造を用いる。3) ヘシアンが正値でなくなる（即ち近似エネルギー曲面が凸曲面になる）ことがしばしば起こるので、強制的に正値にする手法を導入する。4) 式 1 による新規試行構造だけでなく多様な新規試行構造を次々発生させ、それぞれを別々の計算機で計算させる。

【計算結果】 4 台の計算機を用いて分子の構造最適化を行ったところ、下表に示すような高い並列化効率を得られた。

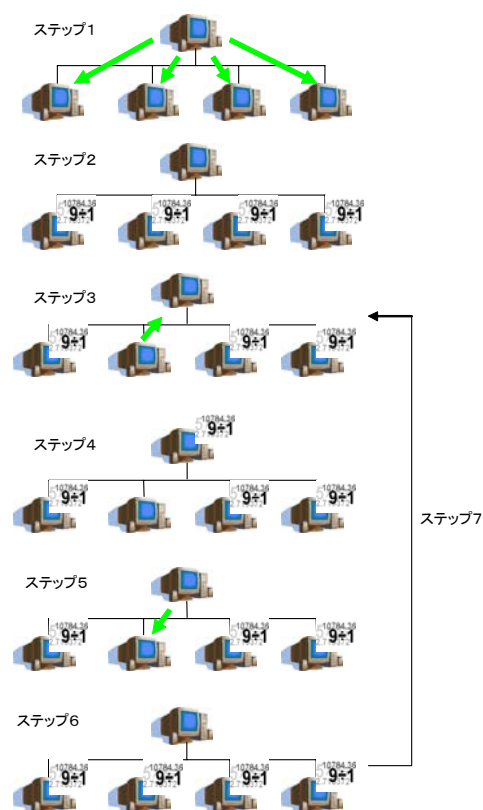


表. 4 台の計算機による並列処理での構造最適化と 1 台の計算機による構造最適化

分子	基底関数 ^{a)}	4 台の計算機 ^{b)} による最適化		1 台の計算機 ^{c)} による最適化		並列化効率 (%) ^{e)}
		一次微分を計算した数 ^{d)}	時間 (秒)	一次微分を計算した数	時間 (秒)	
2H ₂ O	cc-pVTZ	101 (15+22+45+19)	1853	75	3063	54
3H ₂ O	3-21G	146 (27+42+64+14)	1334	105	2218	49
テトラフェニル ポルフィリン	STO-3G	146 (17+18+65+46)	16200	121	29553	69
アゾベンゼン	3-21G	97 (12+16+50+19)	2250	75	3418	53

a) 理論レベルはハートリーフォックレベル; b) 計算機は、Origin2000, Tezro, PowerEdge2600, NEC/SX-6; c) 計算機は、最速である PowerEdge2600; d) 括弧内は、Origin2000, Tezro, PowerEdge2600, NEC/SX-6 のそれぞれで計算した数; e) 並列化効率は、次式で計算した。
$$p = \left(\sum_i n_{parallel}^i - n_{parallel}^{poweredge} \right)^{-1} \left(n_{single}^{poweredge} - n_{parallel}^{poweredge} \right) \times 100$$
 ここで、 $n_{parallel}^i$ 、 $n_{parallel}^{poweredge}$ は、それぞれ、並列計算において、 i 番目の計算機で一次微分を計算した数、Poweredge2600 で一次微分を計算した数であり、 $n_{single}^{poweredge}$ は、Poweredge2600 のみでの計算における一次微分を計算した数である。なお、高精度計算では、一次微分計算に要する計算時間に比べて、データ転送やヘシアン更新・新規試行構造決定等に要する時間はほとんど無視できるので、並列化効率の評価式ではこれらの時間を無視した。