

拡張ヒュッケル法による分子構造最適化の並列処理 分子構造の簡易高速生成の試み

田島 澄恵^a, 片桐 孝洋^b, 金田 康正^c, 長嶋 雲兵^{d*}

^a お茶の水女子大学人間文化研究科複雑系科学専攻, 〒 112-8610 文京区大塚 2-1-1

^b 東京大学大学院理学系研究科情報科学専攻, 〒 113-8658 文京区弥生 2-11-16

^c 東京大学情報基盤センタースーパーコンピュータ研究部門, 〒 113-8658 文京区弥生 2-11-16

^d 産業技術融合領域研究所計算科学研究グループ, 〒 305-8562 つくば市東 1-1-4

*e-mail: umpei@nair.go.jp

(Received: November 29, 1999; Accepted for publication: December 27, 1999; Published on Web: April 10, 2000)

大規模分子軌道計算のための初期分子構造パラメータを簡便に得るために、拡張ヒュッケル法で用いられるスレータ軌道のイクスポネントを水素、炭素、窒素、酸素について求めた。また、全エネルギーのみを用いて分子構造パラメータを最適化するプログラムを開発した。これを用いると、炭化水素分子の分子構造パラメータを実験値と 10% 程度の誤差で得ることができた。

拡張ヒュッケル計算では、計算時間の 99% 以上が行列要素の生成と対角化に用いられるため並列計算の効果が大きい。そのため、メッセージパッシングライブラリ MPI をもちいて並列化された対角化のルーチンをもちいたプログラムの並列化を行った。DEC Alpha2116A/533MHz 8 台の PC クラスタを用いた DNA の Helix 構造 (115 原子: 286 次元) 分子の構造最適化を実行したところプロセッサ数 8 台まで 90% 以上の並列化効率を得られた。

キーワード: Extended Hückel method, Molecular Geometry Optimization, Matrix Diagonalization, Parallel Processing, PC Cluster

1 はじめに

最近の計算機の急速な発展を背景に非経験的分子軌道法や分子動力学法など分子科学計算機シミュレーションで取り扱われる分子のサイズは急速に拡大している。最近では、数百原子によって構成される分子の非経験的分子軌道計算も珍しくなくなっている。

分子軌道計算には、初期データとして分子軌道を記述する基底関数と分子構造が必要である。基底関数の多くは、すでにひろく利用されている様々なプログラム中にデータベースとして格納されており、超精密計算の場合など特別な基底関数を用いた計算を試みる場合をのぞいて、

簡単なキーワードを入力することで基底関数の指定を行うことができる。ところが、分子構造パラメータの指定は、いくつか分子構造の入力を支援するシステムが存在するとはいえ、基本的に人手かまたは簡便な分子軌道計算を用いた最適化によって目的とする系に適切な分子構造パラメータを得ることになる。もちろん小さな分子であれば、標準的な結合距離、結合角、二面角を初期構造として入力し、最適な分子構造を非経験的分子軌道法を用いて計算することも可能であるが、このような方法は演算量が基底関数の数 N の4乗で増加するため、タンパク質など数百原子の系では非常に難しい。そのため、たとえ比較的小さな分子の場合でもその初期構造は分子力学法や半経験的分子軌道法を用いて計算されることが多い。しかしながら、これらの方法は、金属原子を含む系では適用が難しいことが知られている。

本研究では、数百原子によって構成される分子の分子軌道計算のための分子構造パラメータを簡便に得る方法を開発することを目的として、金属原子でも適用可能な拡張ヒュッケル法を用い、良好な分子構造パラメータを得るためのスレータ軌道のイクスポネントを得た。さらにそれを並列処理することで高速化を図った。

2 拡張ヒュッケル法による分子構造最適化

炭化水素分子の系では、拡張ヒュッケル法はすでに過去の物であるが、金属錯体などの金属原子を含む分子の経験的分子軌道計算法として現在もなお広く使われている。分子軌道計算のための分子構造パラメータの初期値を計算するために拡張ヒュッケル法を選択した最大の理由は、拡張ヒュッケル法は、広く使われている半経験的分子軌道法や非経験的分子軌道法のように非線形方程式を自己無撞着(SCF)に解く必要はなく、線形の方程式を解けばよいことにある。つまり近似ハミルトニアン行列を1度対角化すればその固有値と固有ベクトルが解であり、それぞれの固有値に電子占有数を乗じた物の総和が全エネルギーとなることである。言い替えれば、半経験的分子軌道法や非経験的分子軌道法では、系の全エネルギーを得るために最低2回以上(実際は系の大きさにほぼ比例する回数)の行列対角化が必要であるのに比べ、拡張ヒュッケル法は一度の対角化で全エネルギーが計算できる。また、すでに金属原子のパラメータがそろっているため、金属錯体を含む広い範囲での応用が可能であることである。しかしながら、古典的な拡張ヒュッケル法では、分子構造を決める際に重要な核間反発が明示的に評価されていない。そのため、基本となる方程式にはAndersonらの方法[1-3]を採用した。本研究では、分子構造パラメータの再現に非常に敏感なスレータ型軌道のイクスポネントを分子構造パラメータを良く再現するように最適化した。

最適化されたスレータ型軌道のイクスポネントは

H: 1s 1.46 C: 2s 2.01 2p 1.765 N: 2s 2.55 2p 1.95 O: 2s 3.584 2p 2.050

である。通常のスレータ則で決められるイクスポネント(H: 1s 1.00 C: 2s2p 1.625 N: 2s2p 1.95 O: 2s2p 2.275)に比較すると、若干原子核上に局在化している。

分子構造パラメータの最適化方法は以下の通りである。次の説で説明するように、片桐と金田[4]によれば、固有値・固有ベクトルの計算コストは、固有値の計算が1に対して、固有ベクトルの計算が100となることが知られている。この固有ベクトル計算の比率が固有値計算に比較して非常に大きいという傾向は、並列計算の際用いることのできるプロセッサ数が増加するに従って顕著になる。

本研究では対角化の演算を固有値計算にのみ限るために、微分を必要とする Newton 法ではなく、非線形最適化の技法を用いた。具体的な最適化法は以下の通りである。

- 1) 初期構造として、核間距離 CH, NH, OH を 1.08 とし、さらに CC, CN, CO 1.34 とする。角度および二面角は、常識的に原子間反発が少なくなるように適当に決め、これらを用いて全エネルギーを計算する。
- 2) 最適化されるパラメータのランダム順列を作成し、その順にパラメータを 2 つ選択する。
- 3) そのパラメータの値 (X_i, X_j) と $(-0.05X_i, X_j)$ 、 $(+0.05X_i, X_j)$ 、 $(X_i, -0.05X_j)$ 、 $(X_i, +0.05X_j)$ の 5 点の全エネルギー計算をおこなう。
- 4) 5 点を用いた 2 次曲面近似により新たな (X_i, X_j) を得る。
- 5) 3)、4) を繰り返し、すべてのパラメータに対して新たな (X_i, X_j) を求める。パラメータ数が奇数の場合は、最後のパラメータに関して 2 次曲線近似を行い、新たな X_i を求める。ただし、パラメータの最大の変化は初期値の 20% 以下とする。
- 6) 新たな分子構造パラメータによる全エネルギーと古いそれによる全エネルギーの差を取り収束の判断を行う。収束の場合計算を終了し、そうでない場合は 2) に戻る。

この手順で、全エネルギー計算の都度行列の対角化が実行される。

小さな炭化水素分子の最適化構造パラメータの例を Table 1 に示す。Table 1 には実験値と実験値を比較的良好に再現するといわれている STO3G 基底を用いた非経験的分子軌道法による結果 [5] と本研究の結果と実験値の差、および非経験的分子軌道法と実験値の差を示した。

Table 1 を見る限り、角度・二面角の誤差が大きい印象が得られるが、距離の誤差に比べ距離の誤差が小さいといった誤差の系統的な振る舞いはみられない。これは、STO3G を用いた非経験的分子軌道計算の結果でも同様である。誤差の傾向をみるために、Table 2 に Table 1 に示した分子構造パラメータの実験値との相対誤差の絶対値の平均を示した。括弧の中は標準偏差である。STO3G が実験を 1% 程度の誤差で表現しているのに比べ、拡張ヒュッケルを用いた計算では、1-3% 程度の誤差を持つことがわかる。われわれの経験では原子数 10 程度の炭化水素に限ると、実験値に比べて 5% 程度の誤差があるが、STO3G を用いた非経験的分子軌道法の結果とほぼ同程度の精度で分子構造を再現する。この性能は、分子構造の初期構造を求めるという目的には十分である。もちろん、計算時間は、Gaussian94 を用いた HF/STO-3G 非経験的分子軌道計算では比較にならない。これは、423 軌道のハートリーフォック計算の計算時間の 99.4% が二電子積分の計算であり、フォック行列生成と対角化が 0.1% 程度 [6] であることから想像できよう。つまり、この 0.1% が対角化 1 回であったとしても、拡張ヒュッケル計算は、非経験的分子軌道計算の 1000 倍以上の高速性を持つということである。

3 拡張ヒュッケル法による分子構造最適化の並列処理

拡張ヒュッケル法の計算時間の 99% 以上は行列の対角化に要する。行列対角化の計算量は、行列のサイズのほぼ 3 乗に比例する。そのため、タンパク質や高分子などの巨大分子では、拡張ヒュッケル計算といえども多大な計算時間を必要とする。そのため、片桐と金田 [4] が開発した Householder-bisection 法による並列対角化サブルーチンを組み込んで、対角化を並列実行する事で高速化の可能性を調査した。

Table 1. 計算された分子構造

(距離は 10^{-10} m、角度は、度、括弧の中は実験値との相対誤差: [calc.-obs.]/obs.)

		STO3G	This work	Obs.
C2H2	rCC	1.168(-0.029)	1.199(-0.003)	1.203
	rCH	1.065(+0.004)	1.053(-0.008)	1.061
C2H4	rCC	1.306(-0.025)	1.342(+0.002)	1.339
	rCH	1.082(+0.003)	1.086(+0.001)	1.085
	aHCH	115.6(-0.019)	113.0(-0.041)	117.8
CH2NH (Cs)	rCN	1.273(0.000)	1.241(-0.032)	1.273
	rCHsyn	1.091(-0.011)	1.103(0.000)	1.103
	rCHanti	1.089(+0.007)	1.087(+0.006)	1.081
	rNH	1.048(+0.024)	0.990(-0.032)	1.023
	aHsCN	125.4(+0.016)	127.0(+0.029)	123.4
	aHaCN	119.1(-0.005)	119.8(+0.001)	119.7
	aHNC	109.1(-0.013)	122.9(+0.013)	110.5
CH3NH2 (Cs)	rCN	1.486(+0.010)	1.485(+0.010)	1.471
	rCHtr	1.093(-0.001)	1.097(-0.002)	1.099
	rCHg	1.089(-0.009)	1.089(-0.001)	1.090
	rNHa	1.033(+0.011)	1.021(+0.012)	1.010
	aNCHtr	113.7(-0.002)	112.3(-0.014)	113.9
	aNCHgg'	124.0(-0.003)	109.1(-0.130)	124.4
	aHgCHg'	108.2(+0.002)	109.8(+0.017)	108.0
	aHNH	104.4(-0.025)	110.5(+0.032)	107.1
CO	rCO	1.146(+0.016)	1.149(+0.019)	1.128
H2CO	rCO	1.217(+0.007)	1.193(-0.012)	1.208
	rCH	1.101(-0.013)	1.103(-0.019)	1.116
	aHCH	114.5(-0.017)	123.2(+0.058)	116.5
CH3OH*	rCO	1.433(+0.008)	1.473(+0.037)	1.421
	rCHtr	1.092(-0.002)	1.088(-0.005)	1.094
	rCHg	1.095(+0.001)	1.092(-0.002)	1.094
	rOH	0.991(+0.029)	0.974(+0.011)	0.963
	aCOH	103.8(-0.039)	105.5(-0.023)	108.0
C6H6*	rCC	1.384(-0.009)	1.436(+0.028)	1.397
	rCH	1.083(-0.001)	1.089(+0.005)	1.084
C5NH5*	rCaN	1.353(+0.010)	1.351(+0.008)	1.340
	rCaCb	1.388(-0.005)	1.426(+0.022)	1.395
	rCbCc	1.385(-0.006)	1.439(+0.032)	1.394
	rCaH	1.087(-0.003)	1.088(+0.004)	1.084
	rCbH	1.082(+0.001)	1.088(+0.006)	1.081
	rCcH	1.083(+0.006)	1.090(+0.012)	1.077

* 文献 [5] 中で、実験値が明らかでない角度・二面角は、表から除いた。

Table 2. 計算された分子構造パラメータの平均相対誤差とその標準偏差

	Expt.-This work	Expt.-STO3G
すべて	0.021(0.018)	0.009(0.007)
核間距離	0.015(0.012)	0.009(0.006)
角度・二面角	0.031(0.024)	0.009(0.008)

本研究で用いた並列化のモデルは、マスター・スレーブ方式であり、マスターが行列を生成し、スレーブが対角化を並列に実行する。スレーブは計算した固有値をマスターに返し、マスターはそれを用いて全エネルギーを計算する。行列対角化のアルゴリズムは、Householder-bisectionである。拡張ヒュッケル法では、軌道占有数が0でない固有ベクトルの固有値が必要であるので、固有値は行列の次元数の1/2程度しか計算しなくても良い。

Householder-bisection法による行列対角化の並列処理は、1) Householder 3重対角化、2) 二分法による固有値計算、3) 逆反復法による固有ベクトル計算とHouseholder 逆変換となる。片桐と金田は、3000次元のFrank行列の全固有値、全固有ベクトルの計算で、6番目の逆反復法による固有ベクトル計算と7) Householder 逆変換に90%以上の計算時間がかかること、さらに並列度が高くなるにつれ、固有ベクトル計算のコストが増加することを報告している。[4]

Table 3. 3000次元の全固有値、固有ベクトル計算時間*

	三重対角化	固有値計算	固有ベクトル計算	合計
実時間(sec.)	51.4	2.45	2172.4	2226.2
割合(%)	2.3	0.11	97.6	100.0

* 計算は、DEC Alpha 2116A/533MHz 8台/100BaseT-Switch構成のPCクラスタを用いて行った。

750炭素原子系(3000次元)の行列の対角化に要する時間をTable 3に示した。計算は、DEC Alpha2116A/533MHz 8台/100BaseT-Switch構成のPCクラスタを用いて行った。メッセージパッシングライブラリーは、MPI、プログラム言語はFortran90を用いた。Table 3に示した計算時間の比率は、片桐と金田の報告[4]と測定環境が違うが、彼らの報告を再現している。

そのため分子構造探索には、前節で述べたように、固有値と固有ベクトルの両方を必要とするエネルギー微分を用いたNewton法ではなく、エネルギー値のみを用いる非線形探索法を採用した。DNAのHelix構造(115原子:286次元)分子の構造最適化場合、同様な精度の分子構造を得るための計算時間は、Newton法に比べ300倍程度の高速化が可能となった。230原子(572次元)のDNA Helix構造最適化の場合の並列化によるスピードアップと効率をTable 4に示す。対角化ルーチンの並列処理効率が非常に高いため、8台による計算でも効率が90%を越えている事がわかる。

Table 4. 並列処理の効率

プロセッサ数	2	4	8
総計算時間	9.6h	4.9h	2.6h
スピードアップ	1.0	1.94	3.73
並列処理効率	—	97%	93%
対角化 1 回	5.0sec.	2.5	1.29
スピードアップ	1.0	2.0	3.87
並列処理効率	—	100%	97%

4 まとめと今後の課題

本研究では、数百原子によって構成される分子の分子軌道計算のための分子構造パラメータを簡便に得る方法を開発することを目的として、金属原子でも適用可能な拡張ヒュッケル法を用い、良好な分子構造パラメータを得るためのスレータ軌道のイクスポーネントを得た。また、拡張ヒュッケル法の計算時間の多くは行列の対角化に費やされ、固有値計算に比べ固有ベクトル計算のコストが非常に大きいので、固有値計算のみ、つまり全エネルギー計算のみで構造最適化を行うアルゴリズムを開発した。

これを用いて炭化水素の分子構造パラメータを計算すると、実験値に対して 1-3% 程度の誤差を持つことがわかる。原子数 10 程度の炭化水素に限ると、実験値に比べて 5% 程度の誤差があるが、STO3G を用いた非経験的分子軌道法の結果とほぼ同程度の精度で分子構造を再現する。このように本研究で得られた方法は、分子構造の初期構造を求めるという目的には十分であることがわかった。

また、拡張ヒュッケル法の計算時間の多くは行列の対角化に費やされ、行列の対角化の計算量は系のサイズの 3 乗に比例して増大するために、系が大きくなると拡張ヒュッケル計算でも莫大な計算時間を必要とする。そこで、行列対角化を並列処理することで高速化を達成する可能性をみた。メッセージパッシングライブラリを用いた PC クラスタ上での並列処理では、PC 8 台までは 95% 以上の並列処理効率を得た。これは、拡張ヒュッケル計算の並列処理が 8 PE まででは非常に効率よく実行される事を示している。

今後の課題としては、まず各種原子のパラメータの決定があげられる。現在は、H,C,N,O だけであるが、各種金属原子の原子価状態に対応するパラメータを決めていく必要がある。そのためには、各種金属錯体の分子構造のデータベース化が必要である。

また、本研究では分子構造のみに注目したパラメータ最適化を行ったが、励起エネルギーや双極子モーメントなどの各種の物理量をよく再現するためのパラメータ最適化の可能性も考えられる。

本研究で開発した分子構造最適化プログラムの並列性は、行列計算の行列サイズ N の並列性に加え、2 節で説明した計算ステップの 3)、4) で、5 つの点 (または 9) の計算の独立性がある。現在は、マスター 1・スレーブ 8 の 2 階層の並列処理であるが、5 つの点 (または 9) の計算の独立性を考慮すると、3 階層の並列処理が可能となる。そこでマスター 1, 1st スレーブ 5 (または 9)・2nd スレーブ 8 (または 16) のシステムが構築可能な 256 台の並列処理システムを用いて、3 階層並列処理により、さらに 5 (または 9) 倍の高速化が達成できるかを確

認する必要がある。

有益なご議論とご示唆をいただいた、お茶の水女子大学の細矢治夫教授、北海道教育大学の小原 繁助教授、三菱化学の中村振一郎博士、諫田克哉博士、大正製薬の高島 研究員、北村一泰博士、化学技術戦略推進機構土井プロジェクトの吉田元二主任研究員、東京都立大学の三浦信明博士に感謝する。行列対角化のアルゴリズムに関して有益なご示唆をいただいた、図書館情報大学の長谷川秀彦講師、電子技術総合研究所の関口智嗣主任研究官および建部修見博士に感謝する。

参考文献

- [1] A. B. Anderson and R. Hoffmann, *J. Chem. Phys.*, **60**, 4721 (1974).
- [2] A. B. Anderson, *J. Chem. Phys.*, **64**, 1187 (1975).
- [3] A. B. Anderson, R. W. Grims, and J. J. Gajewski, *J. Phys. Chem.*, **91**, 4245 (1987).
- [4] 片桐孝洋, 金田康正, *IPSIJ Sig Notes, 97-HPC- 69*, **97**, 49 (1997).
- [5] W. J. Hehre, L. Radom, P. v. R. Schleyer, J. A. Pople, *Ab Initio Molecular Orbital Theory*, John Wiley & Sons, Inc. (1985).
- [6] 高島 , 私信

Parallel Processing of Molecular Geometry Parameter Optimization by Extended Hückel Method –An Attempt of Simple Fast Generation of Molecular Structure –

Sumie TAJIMA^a, Takahiro KATAGIRI^b, Yasumasa KANADA^c and Umpei NAGASHIMA^{d*}

^aDepartment of Humanculture and Science, Graduate School of Ochanomizu University
2-1-1 Otsuka, Bunkyo-ku, Tokyo 112-8610, Japan

^bDepartment of Information Science, Graduate School of Science, the University of Tokyo
2-11-16 Yayoi, Bunkyo-ku, Tokyo 113-8658, Japan

^cComputer Centre Division, Information Technology Center, the University of Tokyo
2-11-16 Yayoi, Bunkyo-ku, Tokyo 113-8658, Japan

^dNational Institute for Interdisciplinary Research
1-1-4 Higashi, Tsukuba, Ibaraki 305-8562, Japan

**e-mail: umpei@nair.go.jp*

In order to have an initial molecular geometry simply and fast, we modified the extended Hückel method and developed a geometry optimize method using only total energy. The difference between calculated and observed values of small hydrocarbon molecules is less than about 10%.

Parallel processing of the extended Hückel is very efficient up to 8 PC-cluster system because the efficiency for the extended Hückel calculation of DNA(115 atoms and 286 dimension) is bigger than 90%.

Keywords: Extended Hückel method, Molecular Geometry Optimization, Matrix Diagonalization, Parallel Processing, PC Cluster