一次元ポリマーのエネルギーバンド計算における並列処理

寺前 裕之

NTT基礎研究所 〒243-0124 厚木市森の里若宮3-1

*e-mail : teramae@will.brl.ntt.co.jp

(Received : August 27, 1997 ; Accepted for publication : September 17, 1997 ; Published on Web : December 5, 1997)

TCGMSG メッセージパッセンジャーを用いた、一次元ポリマーのエネルギーバンド計算の並列処理について述べる。計算機として、Cray T932 (32 CPU)、IBM RS6000 4 台によるワークステーションクラスターならびに Pentium/Pentium-Pro を使用したパーソナルコンピューター 4 台からなる PC クラスターを使用した。並列化は二電子積分ならびに波数ベクトル各点における対角化に対して行なった。特に波数ベクトルに対する並列化処理はエネルギーバンド計算に対しては有効であることが判明した。対象としている問題が大容量の行列演算を必要とせず長いベクトル長が得られないため、スーパーコンピューターの高速性が十分に発揮されず、PC クラスターにおいても、スーパーコンピューターに匹敵する速度が得られることがわかった。

Keywords: Parallel processing, Crystal orbital, Energy band, PC cluster

1 はじめに

一次元ポリマーに関する非経験的結晶軌道法による計算はポリアセチレンなどの導電性高分子の電子状態研究でその有効性が示されたが、CPU 占有時間およびファイルの入出力共に非常に大規模な計算となるため、いくつかの単位セルが小さなポリマー以外の計算はあまりなされていない [1-4]。

近年、単一 CPU でのコンピューターの処理能力も限界に近くなりつつあり、スーパーコンピューター (SC) やワークステーション (WS) も以前ほどの短期間での飛躍的な性能向上は見込めなくなってきた。そこで、現在では複数の CPU を装備した並列処理システムを用いて並列計算を行なうのが大規模計算では主流になりつつあり、SC や WS もクラスター構成ないしは並列型の構成が主流となりつつある。またポリマー計算のように大容量のファイル入出力を伴う場合にはこのファイル入出力待ち時間も大きくなるため並列化することが望ましい。しかし、WS レベルにおいても CPU を 20-30 台の単位で確保することは以前と比べて価格が下がってきたとは言え、コスト的に見て難しい。

一方で比較的安価なパーソナルコンピュータ (PC) は近年の性能向上がめざましく、WS との差が縮小してきた。例えば Pentium-Pro 150MHz の Linpack 実測値 (n=100) では浮動小数点演算速度は最高約 30Mflops にも達し、RS6000/590 の 35Mflops や Cray T932 の単一 CPU での 90Mflops と比較しても遜色がない値となっている。ただし、loop 長が 1000 の場合には最速値で約 11Mflops に低下してしまい、RS6000/590 の 50Mflops や Cray T932 の 450Mflops には及ばない値となる。また近年 PC 上で利用できる LINUX や FreeBSD などの UNIX が普及し、ethernet 用の拡張ボード類も安価に入手可能であるので、PC に UNIX と network board を追加することで、WS 形態での利用も可能になってきた。また WS と違い台数を確保するのも比較的容易と考えられる。

本研究では一次元ポリマーのエネルギーバンド計算プログラムをソケット通信を用いた TCGMSG を利用して並列計算用に書き直しパフォーマンスの計測などを行なった。SC として、Cray 社の T932、WS として IBM 社 RS6000/590 4 台からなるクラスターを使用し、CPU としてそれぞれ Pentium 133MHz/166MHz (P5-133, P5-166), Pentium-Pro 150MHz/200MHz (P6-150, P6-200) を使用し、OS として FreeBSD 2.2.2R を使用した PC 4 台で構成される PC クラスターとの速度比較を行なった。

2 計算方法

ポリマーのエネルギーバンドの計算理論については文献に詳しいので並列計算に必要な部分のみの記述にとどめる [1-2]。一次元ポリマーの計算理論である結晶軌道法は有限の分子系における分子軌道法を無限系のバンド計算に拡張したものである。

Hartree-Fock 方程式は、

$$\epsilon(\mathbf{k})\mathbf{S}(\mathbf{k})\mathbf{C}(\mathbf{k}) = \mathbf{F}(\mathbf{k})\mathbf{C}(\mathbf{k}) \quad (1)$$

である。 $F_{rs}(\mathbf{k})$, $H_{rs}(\mathbf{k})$, $S_{rs}(\mathbf{k})$ は、 $\mathbf{k}$ をあらわに含まない実空間での各行列要素のフーリエ変換で表すことができて、

$$F_{rs}(\mathbf{k}) = \sum_{j=-N}^N \exp(i\mathbf{k}\mathbf{a}j) F_{rs}^{0j} \quad (2)$$

$$H_{rs}(\mathbf{k}) = \sum_{j=-N}^N \exp(i\mathbf{k}\mathbf{a}j) H_{rs}^{0j} \quad (3)$$

$$S_{rs}(\mathbf{k}) = \sum_{j=-N}^N \exp(i\mathbf{k}\mathbf{a}j) S_{rs}^{0j} \quad (4)$$

ここで、

$$S_{rs}^{0j} = \int \chi_r^0 \chi_s^j d\mathbf{r} \equiv \langle r^0 | s^j \rangle \quad (5)$$

$$H_{rs}^{0j} = -\frac{1}{2} \langle r^0 | \Delta | s^j \rangle - \sum_{h=-N}^N \sum_A^{atom} \langle r^0 | \frac{Z_A}{|\mathbf{r} - h\mathbf{a} - \mathbf{R}_A|} | s^j \rangle \quad (6)$$

```

...
ncb=0
DO J= 0,N
DO K=-N+J,N
DO L=-N+J,N
ncb=ncb+1
if(mod(ncb,nproc).eq.me) then
DO R=1,n
DO S=1,n
DO T=1,n
DO U=1,n
Calculate <R(0)S(J)|T(K)U(L)>
or <R'(0)S(J)|T(K)U(L)>
ENDDOO
ENDDOO
ENDDOO
endif
ENDDOO
ENDDOO
ENDDOO
...

```

(a) 二電子積分およびその微分の並列化

```

...
DO K=1,KPOINT
if(mod(K,nproc).eq.me) then
calculate F(K)
diagonalize F(K)
calculate P(K)
endif
ENDDO
call dgop(type+msgdbl,density,KPOINT*n*n,'+')
...

```

(b) 波数ベクトルを伴う部分の並列化

図 1: 並列計算の概要

$$F_{rs}^{0j} = H_{rs}^{0j} + \sum_{h=-N}^N \sum_{l=-N}^N \sum_t^n \sum_u^n P_{tu}^{hl} \{2\langle r^0 s^j | t^h u^l \rangle - \langle r^0 t^h | s^j u^l \rangle\} \quad (7)$$

$$P_{tu}^{hl} = \frac{a}{\pi} \int_{\text{BZ}} \sum_n^{\text{occupied}} \exp\{-i\mathbf{k}\mathbf{a}(h-l)\} C_{tn}^*(\mathbf{k}) C_{un}(\mathbf{k}) d\mathbf{k} \quad (8)$$

これらの関係式からユニットセル当りの全エネルギーは、

$$\frac{E_{\text{total}}}{N} = -\frac{1}{2} \sum_{j=-N}^N \sum_r^n \sum_s^n (H_{rs}^{0j} + F_{rs}^{0j}) P_{rs}^{0j} + \frac{1}{2} \sum_{h=-N}^N \sum_A^{\text{atom}} \sum_B^{\text{atom}} \frac{Z_A Z_B}{|\mathbf{R}_A - \mathbf{R}_B - j\mathbf{a}|} \quad (9)$$

ここで、 n は基底関数の数を N は考慮する隣接セル数を表す。式 (7) より容易にわかるように、二電子積分の数は $N^3 n^4$ に比例する。Fock の行列要素を計算するのに必要な電子密度行列は結晶軌道の係数 $C_{ln}(\mathbf{k})$ から計算されるが、結晶軌道の係数は変分方程式を解かないと得ることができない。従って、分子軌道計算の場合と全く同様に SCF 計算が必要である。二電子積分および二電子積分の核座標に関する微分の計算を隣接セル数である N を用いた並列化ならびにバンド計算において特徴的な波数ベクトル $\mathbf{k}$ に関する計算についての並列化を行なった。(式 (1)-(4), (8)) 図 1 に並列化の概要を示

表 1: Poly-tetrafluoro-ethylene の並列計算の速度比較 (秒)

プロセス数	時間 (秒) ^a	加速率	時間 (秒) ^b	加速率
Cray T932				
1	544.7	1.00	544.7	1.00
2	282.5	1.93	291.2	1.87
4	147.7	3.69	163.6	3.32
8	88.9	6.13	108.3	5.02
16	60.8	8.96	80.4	6.77
RS6000/590 cluster				
1	342.8	1.00	342.8	1.00
2	182.8	1.88	195.8	1.75
4	88.2	3.89	107.9	3.18
8	52.9	6.48	76.3	4.49
16	37.1	9.24	62.6	5.47
PC-cluster				
1	794.4	1.00	794.4	1.00
2	492.0	1.61	556.2	1.42
4	301.7	2.63	468.2	1.69
8	154.6	5.14	337.1	2.35
16	139.3	5.70	301.2	2.63
^a k 並列化を含む				
^b k 並列化を含まない				

す。なお二電子積分の並列化のアルゴリズムは分子軌道法において使われているものと同様であるが、単位セルの番号を用いて並列化を行なっている点が異なる。並列計算を行なうためのライブラリとしては、TCGMSG ライブラリ [5] を用いて各プロセス間のソケット通信を行なった。

速度比較のために poly-tetrafluoro-ethylene (C_2F_4)_x を対象に選んだ。基底関数は STO-3G[7] を用いた。(n=30) 隣接セル数 N は 5 として、二電子積分のカットオフ法には Namur Cutoff 法を用いた [8]。波数ベクトルのサンプリングは 41 点で行ない、Simpson の公式を用いて式 (8) の積分を行なった [6]。

3 結果と考察

表 1 に計測された CPU 占有時間 (秒単位) を示す。二電子積分と波数ベクトルの両方を並列化したものと、波数ベクトルについては並列化を行なわなかった物の双方について示した。ここで並列計算における CPU 占有時間は並列計算時に最も CPU 時間

を消費したプロセスの CPU 時間である。速度の計測は各 10 回行ない、最速値をとった。WS および PC クラスタはそれぞれ 4CPU しか持たないため、各 CPU について 1、2、または 4 プロセスずつ発生させて、それぞれ 4、8、16 並列計算の値をシュミレートした。ただし同一 CPU 上のプロセス間通信ではセマフォを用いた共有メモリーが使用されるので、純粋に 8CPU、16CPU でソケット通信のみを使用した場合に比べて速度的に有利になっている可能性は残る。なお、Cray T932 は 32CPU を持ち原理的には 32 並列まで可能であるが、センター共用マシンであり運用上のメモリー確保の関係から 16 並列以上の計算が今回は不可能であった。

ここで加速率を 1CPU 使用時の CPU 占有時間を並列計算時の CPU 時間で除算したものと定義すると、SC ならびに WS では、4 プロセスまではほぼ直線的に加速されて、1CPU 使用時の 3.69 倍ならびに 3.89 倍と理論値 4 にかなり近い値が得られていることがわかる。これに対して、PC クラスタでは速度の違う CPU を用いているため最も遅い CPU が律速となり見かけ上の加速率は小さくなるものの、4 プロセスで 2.63 倍となり、並列計算を行なうことにより SC や WS を単一で用いた場合の性能を上回っていることがわかる。また PC クラスタ 8 プロセス時の CPU 時間を見てみると Cray のほぼ 4CPU、RS6000/590 クラスタの 3CPU と同等の性能が得られている。ただし、PC クラスタでは 8 プロセスから 16 プロセスへプロセス数を増加させても 5.14 倍から 5.70 倍とわずかな速度の向上にとどまっている。

次に Cray T932 および RS6000/590 クラスタでは並列度が 8 または 16 になった場合には 4 までの場合に比べて加速率が上がらなくなり、並列度 8 の場合で 6-6.5 倍程度、並列度 16 の場合で約 9 倍程度に留まっている。図 2 に並列度 16 の場合の各ノードでの CPU 占有時間を示した。これから容易にわかるように、並列度が上がるにつれて、各ノードへの計算の効率的な分散がうまくいかなくなっている。特にノード 12、14 はほとんど使用されていないことがわかる。さらに PC クラスタでは最も計算速度の速い P6-200 CPU がノード 12-15 に割り当てられており、並列度 8 と 16 の場合でほとんど速度の向上が見られなかった原因であると考えられる。今後はさらに各 CPU への負荷分散がより平均化されるような並列化の方法を検討する必要がある。

波数ベクトルに関する並列化の効果は並列度が上がると相対的にかなり大きくなり、16 並列計算時には、最も効果が少ない Cray T932 の場合でも 80.4 秒から 60.8 秒と約 30% の節約になっていることが表 1 からわかる。WS クラスタにおいても、62.6 秒から 37.1 秒と約 40% 節約されている。特に PC クラスタの場合は対角化およびそれに伴う行列の演算が特に遅いためと思われるが [9]、CPU 時間が 301.2 秒から 139.3 秒と半分以下に減少している。逆にいうと、このようにベクトル長の短い計算においては PC クラスタによる並列計算は特に有効であることがわかる。波数ベクトルのサンプリング点は多ければ多いほど正確になるため、例えば CPU 数の倍数になるようにサンプリングを行なうような方法を用いる事で高精度高能率な並列処理が PC クラスタ上で行なえる事が期待できる。

図 2: 16 ノード並列計算時の各 CPU 使用時間

ファイルの並列入出力は今回行った方法によれば自然に行なわれる。1CPU での実行時には 32MB あったファイルが例えば 16 並列処理では最大のファイルサイズが約 3.6MB になりファイルアクセス時間は CPU 使用時間と同等に分散されていることがわかる。この事実は入出力速度が遅い PC クラスタの場合特に重要である。また今回は試みなかったが、PC クラスタにおいてはメモリーを増設するのが価格面から SC や WS に比べてはるかに容易であり、かなりの大きさのポリマーまで全ての積分をメモリーに保存することもある程度の台数が確保できるならば可能である。今後の検討課題であろう。

PC クラスタに関して、P5CPU についてはあまり問題がないが、P6CPU はちょっとしたプログラミングによって大きく計算速度が変わる場合がある。例えば、Linpack ベンチマークでループ数 100 配列数 201 では 10Mflops 程度であるが配列数 200 であると 30Mflops 程度まで速度が向上する。ループ数を 101 として計算してみると、配列数 101 だと 9Mflops 程度まで速度が低下してしまう。ループ数 101 配列数を 102 とすると 15Mflops まで回復する。このようにちょっとしたプログラムの差により、キャッシュメモリーの制御の関係と思われるが、大きく計算速度が変わるようである。本研究においても、例えば積分のプログラムの手直しにより全く関係が無いと思われる、SCF の部分の計算時間が 1.5 倍程度の範囲内で上下した。PC クラスタで最高のパ

パフォーマンスを得るためには今後さらにプログラム全体の検討が必要と思われる。

4 まとめ

一次元ポリマーの非経験的バンド計算の並列処理について、PC クラスタ、WS クラスタならびにスーパーコンピュータを用いてモデルポリマーの計算時間についての検討した。

意外にも単一 CPU での計算においても、SC はそれほどの速度が得られず WS が最速となる結果が得られた。主な原因としては例とした計算が大きな配列を必要とせず、SC のメリットが生かされなかったと考えられるが、これについては詳しい解析が必要であり今後の検討を要する。また PC(P6-200) では予想以上の速度が得られているが、計算の規模等からキャッシュメモリー上でうまく処理されたのではないかと考えられる。そのため常にこのような速度が期待できるかどうかについてはさらに検討する必要がある。

謝辞 本研究を行なうにあたり、お世話になりましたお茶の水女子大学の長嶋雲兵教授に深く感謝いたします。

参考文献

- [1] M. Kertesz, *Adv. Quantum Chem.*, **15**, 161 (1982).
- [2] P. Otto, E. Clementi, and J. Ladik, *J. Chem. Phys.* **78**, 4547 (1983).
- [3] H. Teramae, *J. Chem. Phys.* **85**, 990 (1986).
- [4] H. Teramae and K. Takeda, *J. Am. Chem. Soc.* **111**, 1281 (1989).
- [5] R. J. Harrison, TCGMSG ver. 4.04, Battelle Pacific Northwest Laboratory (1994); R. J. Harrison, *Int. J. Quantum Chem.*, **40**, 847 (1991)
- [6] P. J. Davis and P. Rabinowitz, *Method of Numerical Integration* (Academic Press, New York, 1975), p. 45.
- [7] W. J. Hehre, R. Ditchfield, R. F. Stewart, and J. A. Pople, *J. Chem. Phys.* **52**, 2769 (1970).
- [8] H. Teramae, *Theoret. Chim. Acta* **94**, 311 (1996).
- [9] 日向寺祥子、長嶋雲兵、青柳睦、佐藤三久、関口智嗣、桐山博史、細矢治夫、*IPJS SIG Notes*, Vol. 94, 94-HPC-52, 19 (1994).

Parallel Processing on *ab initio* Crystal Orbital Calculations of One-Dimensional Polymers

Hiroyuki TERAMAE

NTT Basic Research Laboratories
3-1 Morinosato Wakamiya, Atsugi, Kanagawa, 243-0124 Japan
*e-mail : teramae@will.brl.ntt.co.jp

(Received : August 27, 1997 ; Accepted for publication : September 17, 1997 ; Published on Web : December 5, 1997)

We have performed parallel processing on the *ab initio* crystal orbital calculations of one-dimensional polymers using the TCGMSG message passenger. Cray T932 (32 CPU), IBM RS6000 cluster (4 CPU), and personal computer cluster with Intel Pentium and Pentium-Pro (4 CPU) were used for the computational environment. The two electron integrals and the diagonalization of the different k-points were computed in parallel. We found that the parallel processing on the diagonalisation of the different k-points were totally effective for the energy band calculations. The problem solved here does not require massive matrix calculations and we can obtain the speed of the supercomputer with the personal computer cluster.

Keyword: Parallel processing, Crystal orbital, Energy band, PC cluster