

2E06

RI 密度汎関数法の新しい並列化方法とその性能

(東大・情報基盤¹, 東大・生研², 日立・機械³) ○佐藤文俊^{1,2}, 稲葉亨³,
井原直樹², 恒川直樹², 西野典子², 西村康幸², 平野敏行², 吉廣保², 柏木浩²

【序】本研究グループではタンパク質の性質を電子レベルで理解することを目的に、タンパク質全電子計算に基づくシミュレーションシステムを開発している[1]。そのエンジンとなるプログラムは ProteinDF[2]であり、本手法では resolution of identity (RI) 法に基づくガウス関数を基底関数とした密度汎関数法 (DFT) による手法を採用した[3]。タンパク質のような大規模分子の電子状態解析を達成するためには、特に次の二点が重要となる。一つは巨大 SCF 計算を収束させるために、非常に精度のよい初期値を作成することである[4]。もう一つは計算が大規模になるため、効率の良い並列化を行うことである[5]。本発表では後者の並列化について報告する。

【方法】本プログラムで計算律速となる計算ルーチンは、分子積分演算、交換相関項の数値積分、そして行列演算である。分子積分、交換相関項の積分には、大量のメモリ領域を要するすべての行列要素を各 CPU に分散保存させる、独自のアルゴリズムにて並列化した。ここで、本手法における分子積分計算並列効率向上の鍵は、積分カットオフで生き残ったインデックスをシェルタイプ毎に均等圧縮分散処理させることである。行列演算は対角化、行列積、逆行列であり、密行列用の並列版行列演算ライブラリである ScaLAPACK [6]を用いた。本ライブラリも行列要素を分散保持するため、全ての演算は 1CPU 毎に超巨大メモリを必要としない。

【結果と考察】インスリン 6 量体 (306 残基、原子数 4,716、電子数 18,552、基底関数 26,766、補助基底関数 48,600) の全電子状態計算を行なった。使用した計算機は 64 個の Itanium2 プロセッサが NUMalinkTM (3.2 GB/s) で結合された SGI Altix3700 で、ピーク性能は 333 GFLOPS である。全電子計算は 17 回の SCF 計算を要し (Figure 1)、計算時間は 65 時間で (Table 1)、86% の並列化効率が得られた (Table 2)。133 残基のインターロイキンを用いて、Altix3700 と Gbit イーサーで接続した PC クラスタで同様の計算見積もりを行った結果、ネットワークスピードが遅い場合 (PC クラスタ)、行列演算が大きく影響を受けて全体の効率が下がるが、本手法による分子積分、交換相関並列計算に関してはネットワークスピードの影響が少ないことが明らかになった。本研究により、100CPU クラスで有効に並列化された RI-DFT カノニカル分子軌道法プログラムが完成した。

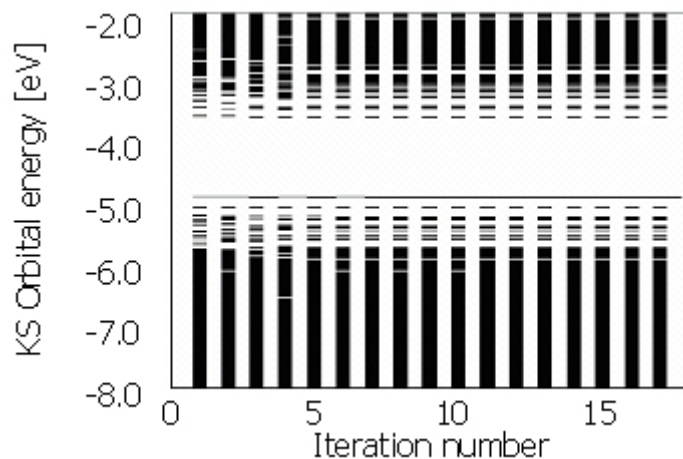


Figure 1. SCF convergence history of the Kohn-Sham orbital energy distribution of insulin hexamer.

Table 1. Timing-data in the whole calculation of insulin hexamer using 1.3 GHz 64 CPU Itanium2 cluster

Calculation routines	Elapsed time [hour:minute]
Integrals before SCF loop	0:27
X matrix	0:56
Grid generation	8:27
Inverse matrix	0:34
SCF loop: Average/ First / Converged(17th)	3:12 / 3:49 / 2:36
Other	0:21
Total	64:58

Table 2. Timing-data in first SCF calculation of insulin hexamer using 1.3 GHz Itanium2 cluster

Calculation Routine	8 CPUs				16 CPUs			32 CPUs			64 CPUs		
	Elapsed time [sec]	Elapsed time [sec]	Speed up ratio ^a	Parallelization efficiency [%]	Elapsed time [sec]	Speed up ratio ^a	Parallelization efficiency [%]	Elapsed time [sec]	Speed up ratio ^a	Parallelization efficiency [%]	Elapsed time [sec]	Speed up ratio ^a	Parallelization efficiency [%]
Molecular integrals	59 353	31 247	15.20	94.97	15 923	29.82	93.19	8 671	54.76	85.56			
XC potential fitting	4 225	2 695	12.54	78.39	1 386	24.39	76.21	760	44.47	69.49			
Diagonalization	11 977	6 901	13.88	86.78	3 558	26.93	84.16	1 899	50.46	78.84			
Multiplication	12 766	6 630	15.40	96.28	3 418	29.88	93.37	1 840	55.50	86.73			
Other	3 553	2 350	12.10	75.60	2 207	12.88	40.25	756	37.60	58.75			
Total	91 874	49 823	14.75	92.20	26 492	27.74	86.70	13 926	52.78	82.47			

^a Speed up ratio: serial time estimated by elapsed time of eight CPUs divided by eight.

本研究の一部は、文部科学省次世代 IT 基盤構築のための研究開発プログラム「革新的シミュレーションソフトウェアの研究開発」の支援の下に行われた。計算は統計数理研究所の統計科学スーパーコンピュータシステムを利用した（2006-ISM-CRP-1016）。

【参考文献】

- [1] <http://www.rss21.iis.u-tokyo.ac.jp/theme/life/reaction/index.html>; 西野典子, 平野敏行, 佐藤文俊, 分子構造総合討論会 2006, 2P074, (2006); 恒川直樹, 伊藤宏比古, 佐藤文俊, *ibid*, 2P079, (2006); 井原直樹, 佐藤文俊, *ibid*, 3P064, (2006); 西村康幸, 吉廣保, 石川寛人, 佐藤文俊, *ibid*, 4P067, (2006).
- [2] F. Sato, Y. Shigemitsu, I. Okazaki, S. Yahiro, M. Fukue, S. Kozuru, H. Kashiwagi, *Int. J. Quant. Chem.*, **63**, 245 (1997).
- [3] J. Andzelm, E. Wimmer, *J. Chem. Phys.*, **96**, 1280 (1992).
- [4] H. Kashiwagi, H. Iwai, K. Tokieda, M. Era, T. Sumita, T. Yoshihiro, F. Sato, *Molecular Physics*, **101**, 81 (2003); T. Inaba, S. Tahara, N. Nishikawa, H. Kashiwagi, F. Sato, *J. Comput. Chem.*, **26**, 987 (2005).
- [5] T. Inaba, F. Sato, "Development of Parallel Density Functional Program using Distributed Matrix to Calculate All-Electron Canonical Wavefunction of Large Molecules", *J. Comp. Chem.*, submitting (2006).
- [6] L. S. Blackford, J. Choi, A. Cleary, E. D'Azevedo, J. Demmel, L. Dhillon, J. Dongarra, S. Hammarling, G. Henry, A. Petit, K. Stanley, D. Walker, R. C. Whaley, "ScaLAPACK User's Guide", PA, (1997).