

課題名 (タイトル) :

原子間相互作用摂動解析による生体分子機能の解析と制御

利用者氏名 : ○小山 洋平*, 篠原 雄太**

所属 : * 生命システム研究センター 計算分子設計研究グループ

**生命システム研究センター 合成生物学研究グループ

<p>1. 本課題の研究の背景、目的、関係するプロジェクトとの関係</p> <p>現在、生命システム研究センター計算分子設計研究グループにおいて分子動力学シミュレーション専用計算機 MDGRAPE-4 の開発が行われており、次年度中には稼動予定である。MDGRAPE-4 では通常の並列計算機に比べて数十倍以上長時間のシミュレーションが実行可能であるが、これに伴い解析すべき構造データ数もより大規模なものとなる。また、報告者は構造データの解析手法として、分子内相互作用摂動解析であるポテンシャルエネルギー主成分分析 (PEPCA; Y. M. Koyama et al., Phys. Rev. E 78, 046702 (2008)) と分子間相互作用摂動解析 (DIPA; Y. M. Koyama et al., Phys. Rev. E 84, 026704 (2011)) を開発してきた。生体分子系に対して PEPCA や DIPA を実行するためには数十万次元以上の高次元の主成分分析 (PCA) を実行する必要がある。このように、データの変数およびデータ数の両者が高次元 (数十万以上) である場合には、解析データがメモリやハードディスクに直接入りきらなくなるため、out-of-core (external-memory) アルゴリズムといわれる手法を用いて全ての計算過程がメモリに収まるように工夫する必要がある。</p> <p>本課題では特異値分解 (SVD) や主成分分析の out-of-core アルゴリズムの一種である randomized SVD (PCA) (N. Halko et al., SIAM Rev. 53, 217 (2011)) と行列成分を必要に応じて計算する on-the-fly evaluation の実装を行い、MDGRAPE-4 に PEPCA や DIPA を適用するためのプログラムの開発を行った。</p>	<p>システム用に MPI、共有メモリ用に OpenMP で実装し、評価を行った。</p> <p>3. 結果</p> <p>単一コアでの randomized PCA および on-the-fly evaluation の実装を行った結果、200 万次元の変数、4 万データからなるタンパク質のシミュレーションに対して 3 日弱で PEPCA が実行できることが確認できた。また、データ数に対して計算時間はほぼ線形であること、ポテンシャルエネルギーの計算部分が律速であること、主成分分析の結果が妥当であることが確認できた。この結果から、単一コアで 100 万データを処理するためには 75 日ほどかかると推測でき、解析を数日以内で終了させるためには 32 コア程度の並列化が必要であると見積もられた。</p> <p>最初に分散メモリシステムでの MPI 実装を行った。この結果、並列性能が良いことは確認できたが、行列成分をその都度計算する on-the-fly evaluation の実装では扱うことのできるデータ次元が利用できる全メモリではなく、各コア当たりのメモリによって決定されるため、扱うことのできるデータ数がコア数を増やしても増加しないことが分かった。一方で、共有メモリシステムでは 1 ノード当たりの全メモリを有効に活用できることが分かったため、OpenMP での実装を行った。RICC の mpc クラスタでは 1 ノードが 8 コアからなるが、8 コアまでほぼ線形に高速化され、8 コア並列で 200 万次元・1 万データの PEPCA が 48 分で終了した。この結果から 8 コアでは 100 万データ処理するのに 80 時間必要であると見積もられ、実用的な計算時間で PEPCA の実行が可能であることが分かった。</p>
<p>2. 具体的な利用内容、計算方法</p> <p>最初に randomized PCA および on-the-fly evaluation を単一コア用に C++ で実装し、その評価を行った。また、この結果に基づき分散メモリ</p>	<p>4. まとめ</p> <p>Randomized PCA、on-the-fly-evaluation、OpenMP による並列化の実装、により高次元・大規模デー</p>

タの PEPCA が実用的な計算時間で実行できるようになった。

5. 今後の計画・展望

MDGRAPE-4 では構造パラメーターファイル(トポロジーファイル)として Gromacs フォーマットが採用される予定である。本報告でのプログラムは現時点で Amber フォーマットしか対応していないため、Gromacs フォーマットへの実装を行っており、対応予定である。また、分子内の静電ポテンシャルの計算にはカットオフを用いていないが、カットオフを導入することによりさらに計算を高速化させる予定である。Randomized-PCA を実行するために必要な行列積や QR 分解を行うために、マルチスレッド版 BLAS・LAPACK を利用しているが、共有メモリシステムに特化した線形代数ライブラリである PLASMA の利用などを検討する予定である。以上の実装・テストを行った後、次年度中にプログラムの公開を目指す。