

課題名 (タイトル) :

大規模並列化に適した MD コアプログラム開発

利用者氏名 :

○大野 洋介*

小山 洋*

長谷川 亜樹*

舛本 現*

森本 元太郎**

似鳥 啓吾**

古石 貴裕***

理研での所属研究室名 :

* 和光研究所 次世代計算科学研究開発プログラム

次世代生命体統合シミュレーション研究推進グループ 生命体基盤ソフトウェア開発・高度化チーム

** 和光研究所 基幹研究所 先端計算科学研究領域

システム計算生物学研究グループ 高速分子シミュレーション研究チーム

*** 本所 知的財産戦略センター

産業界との融合的連携研究プログラム 界面ナノ構造研究チーム

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

次世代生命体統合シミュレーション研究推進グループでは次世代スーパーコンピュータを活用するグランドチャレンジとして生命科学分野でのソフトウェア開発を行なっている。その中で生命体基盤ソフトウェア開発・高度化チームは他のチームのソフトウェアの開発支援や共通で利用できるソフトウェア基盤の開発を担当している。本課題の目的は開発中の基盤ソフトウェアの一つである分子動力学(MD)コアプログラムの性能評価、改良点の分析、並列性能改良である。

2. 具体的な利用内容、計算方法

本課題では分子動力学計算で最も計算負荷のかかる二体相互作用計算の性能評価を行なった。クーロン相互作用のような二体相互作用は直接計算では演算量が粒子数の二乗に比例して増加する。そのため、原子間相互作用が遠距離では弱いことに着目し、直接計算はカットオフ距離内のみで行うカットオフ法と、原子の物理量をメッシュ点に分配して空間メッシュで遠距離相互作用を計算する Particle Mesh (PM) 法を組み合わせる

のが主流であり、開発中の MD コアプログラムでも採用した。カットオフ法では近距離粒子の探索の効率化のため空間座標で粒子を分割・グループ化するセルインデックス法を採用した。計算も空間分割の単位であるセルを単位として並列化する。セルインデックス法による並列効率を向上させるとともにカットオフ計算の効率化を図る。計算対象は水分子の系とした。水分子内の結合力による運動のタイムスケールは二体相互作用によるものより短いため、結合力として計算せず結合距離を固定する拘束条件として処理する SETTLE 法を使用した。

並列化は MPI のみでスレッド並列は使用していない。

3. 結果

図 1 は 2400 原子の水の系をカットオフ計算で行なった場合の処理時間が並列数によってどのように減少するかを表示したものである。総処理時間(total)だけでなく、純粋な計算時間(Calculation)と通信時間(Communication)もプロットしてある。通信時間はばらつきが大きいいため多数の試行から最小の結果を選んである。

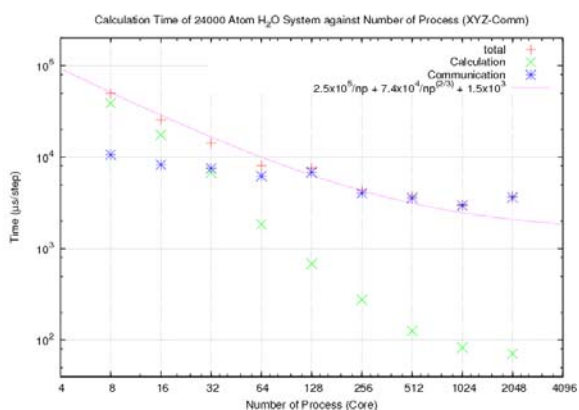


図 1

計算時間は 500 並列程度の並列性があるが、32 並列を超えると通信時間が支配的で、かつ、プロセス数依存がほとんどなくなっている。これは並列度が上がるとカットオフ距離で到達する空間領域が 1 プロセスの分割範囲を超えることで通信するプロセスが増加し、1 プロセスに分割された領域が小さくなる効果を相殺するためだと考えられる。したがって、これ以上の並列性が確保するには、より粒子数が多く計算空間が大きい系を扱う必要がある。

図 2 は原子数で約 4 倍、空間サイズで約 1.5 倍の系での結果である。

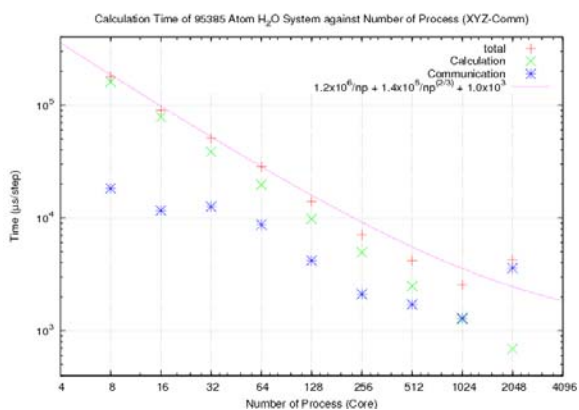


図 2

計算時間は 2048 並列、通信時間は 1024 並列まで減少しており、並列性がよくなっている。通信範囲は 1024 並列から広がっており 2048 並列では通信時間が著しく増加している。現状の通信方式では大規模並列時には隣接通信にとどまるような空間サイズの系でなければ通信性能が著しく劣化すると予想される。

4. まとめ

並列度があがると通信時間が支配的となるため粒子数・空間サイズが大きい系でなければ並列性

の確保は困難である。特に通信対象となるプロセスの増加の影響が大きく、カットオフ距離と空間分割サイズが制限される。

共有環境で通信時間の正確な測定は難しい。

5. 今後の計画・展望

本報告時点では完了していない PME やマルチグリッド法による遠距離相互作用計算等の追加機能の実装と評価を行なう。高並列時には遠距離相互作用計算の負荷も無視できないと考えられる。特に通信負荷が高いと考えられるので通信性能の評価・改善を重点的に行なう。

また、結合力計算機能追加によりタンパク質等の計算が可能となるので、生体分子の実用計算に近い条件での性能評価も行なう。

6. 一般利用で演算時間を使い切れなかった理由

並列性能評価が主目的であったが、通信時間の実測値のばらつきが大きかった。ネットワークの物理的構造上、並列度が上がるほど他のジョブの実行状況の影響が大きいためだと考えられ、一般利用は高並列・長時間の性能測定は困難だと判断し、小規模または短時間の性能評価を主に行なったため予定より利用時間が短くなった。また、システムトラブルがあったこと、開発中のソフトウェアで評価対象の機能を順次追加していることにより、12 月までの利用が少なく、一月以降に利用が偏った。特に年度内に実装完了予定のマルチグリッド法は、本報告時点では評価途中である。機能追加により複雑な処理を含んだ計算を行なうので一回の試行の計算時間は今季より長くなる。主要な追加機能である遠距離相互作用計算の負荷の割合はパラメータ調整でほぼ任意に設定できるが、基本的にはカットオフ計算と同程度の負荷の時に全体の計算時間が最小になるので、最適な条件では計算時間はほぼ倍増する見込みである。また、タンパク質を含む系など計算可能な粒子系の種類も増えるため、テストパターンの種類も増える。これらの要因を考慮すると、今季の評価時間の 4-5 倍の時間が必要になると考えられ、今回持ち越す時間が適切な演算時間だと考えられる。

平成 21 年度 RICC 利用研究成果リスト

【国際会議などの予稿集、proceeding】

Hiroshi Koyama , Gen Masumoto, Yousuke Ohno, Aki Hasegawa, Gentaro Morimoto, and Makoto Taiji, “Development of Highly parallelized Molecular Dynamics Simulation code”, SIAM Conference on Parallel Processing for Scientific Computing (PP10), Seattle, Washington (2010 Feb.)

Hiroshi Koyama, Gen Masumoto, Yousuke Ohno, Aki Hasegawa, Gentaro Morimoto, Makoto Taiji, “cppmd” current status and the future, The 2nd Biosupercomputing Symposium, Tokyo (2010 Mar.)

