

4P062 グリッド技術を用いた フォック行列計算プログラムの開発

(科技振 CREST¹, 産総研 計算科学², 九大 情基³)

○梅田宏明^{1,2}, 稲富雄一³, 渡邊寿雄^{1,2}, 石元孝佳^{1,2}, 長嶋雲兵^{1,2}

序

生体分子のような大規模系の分子軌道計算ではフォック行列の生成に全計算時間の 9 割以上の時間を費やすことが知られている。このような大規模分子軌道計算を可能にするための手法として、ボトルネックとなっているフォック行列の生成を別の高速計算機(例えばスーパーコンピュータや MPP マシンなど)にオフロードする方法がある。これまで我々が行なってきた EHPC (Embedded High Performance Computing)プロジェクト[1]では、フォック行列生成専用計算機を開発し、これにフォック行列生成部分をオフロードすることで高速化を目指し一定の成果を挙げている。一方で近年の PC の高速化やインターネット環境の発達により、インターネット上の複数の計算機資源を使うことのできるグリッド技術が現実的なものとなってきた。グリッド技術を利用すれば (PC クラスタはもちろん専用計算機や MPP マシンなど) 複数の計算機資源を利用することが可能であり、より高速な科学技術計算を実現できる。本研究では、フォック行列の生成をオフロードする対象としてグリッド環境を想定し、広域ネットワークで接続された複数の PC クラスタシステムにおいても効果的に動作するフォック行列計算プログラムを開発した。

GridRPC/MPI ハイブリッド・プログラミング・モデル

グリッド技術を用いた並列計算モデルとしては GridMPI を利用して MPI プログラムをそのままグリッド環境下で動作させるものもあるが、ネットワーク遅延の大きい広域ネットワークを経由したような場合には通信のコストが大きくなるため十分な性能が発揮できない場合がある。そこで疎結合ネットワークを考慮したプログラミング・モデルである GridRPC プログラミング・モデルを採用した。特に GridRPC API をサポートするミドルウェアである Ninf-G [2] では、広域ネットワーク向けの GridRPC だけでなく、高速な内部ネットワーク用に MPI による並列プログラムも利用することができる (GridRPC/MPI ハイブリッド・プログラミング・モデル、図 1)。

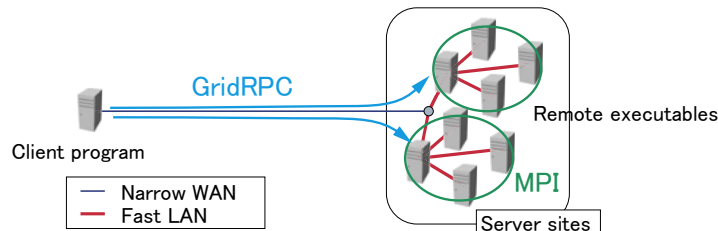


図 1 GridRPC/MPI ハイブリッド・プログラミング・モデル

図 1 に見られるように GridRPC/MPI モデルは、クライアントマシン・リモート実行プログラム間の GridRPC 通信とリモート実行プログラム内の MPI 通信の二つの階層からなっており、これに適したアルゴリズムの開発が必要である。一方、EHPC プロジェクトにおいて我々が開発した

並列フォック行列生成プログラムは、EHPC システムのような階層的な構造を持った並列コンピュータシステムにおいても効果的に並列計算が可能であった [3]。このため EHPC プロジェクトにおいて開発したプログラムは比較的用意に GridRPC/MPI モデルを用いてグリッド実行可能になると考えられる。そこで我々は EHPC システム用並列ルーチン組込み GAMESS [4]プログラムの Ninf-G 環境への移植を行なった。

テスト計算および結果

開発したプログラムのテストとしてクランビン分子(46 残基, 642 原子)の HF/STO-3G 計算(1,974 基底)を行なった。テストはクライアント・プログラムを実行するローカル PC から、WAN 環境に横した 100Base-TX ネットワークを経由して接続された AIST スーパークラスタ(ASC)[5]の P32 および F32 クラスタの合計 256 プロセッサを GridRPC で呼び出す形で行なった。ASC の内部ネットワークはそれぞれ Myrinet および Gigabit ethernet となっている。それぞれのクラスタをさらに異なるプロセッサ数(8-32)の 10 個の MPI リモート実行プログラムに分けることで、性能の異なるクラスタ群を利用した場合の負荷分散の状態を検証した。

図 2 は最初の SCF サイクル付近でのリモート実行プログラムのタイミング・チャートである。青の領域はデータ転送を、赤はフォック行列の計算を、白はクライアントからの要求待ちを意味している。フォック計算直前のデータ転送は密度行列データ等の転送であり、計算途中にある線はパラメータの転送である。密度行列の転送では、比較的大きいサイズのデータを WAN 経由で同時に送っているため、やや時間がかかっている。しかしながら通信の頻度は SCF サイクルにつき十数回非同期に行なうだけであり、遅延の大きい WAN に対しても高い性能を発揮できると思われる。フォック行列計算は全てのリモート実行プログラムにおいてほぼ同時に終了しており、性能の異なるクラスタ群を用いても全てのプロセッサを効率良く利用していることがわかる。

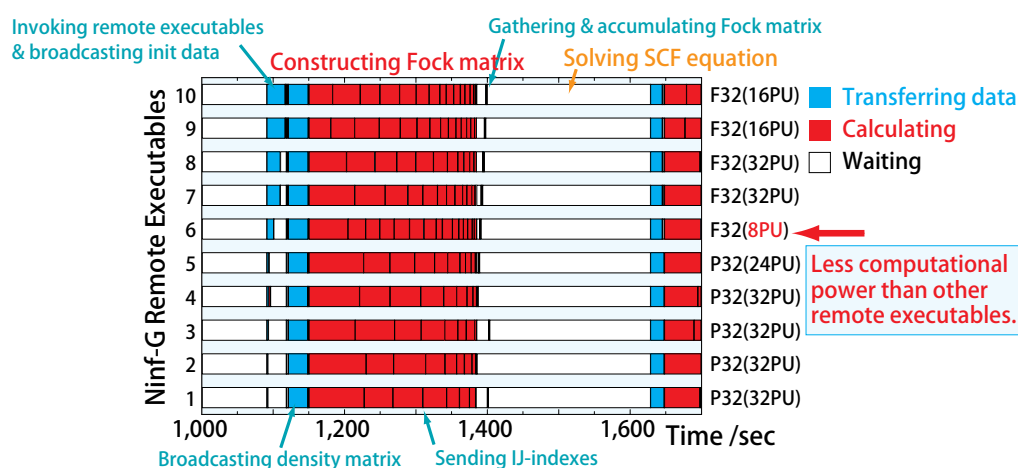


図 2 リモート実行プログラムのタイミング・チャート

- [1] <http://www.ehpc.jp/>. [2] “Ninf: A Global Computing Interface”, <http://ninf.apgrid.org/>.
 [3] H. Umeda et al., *J. Comput. Chem. Jpn.*, **4**, 179(2005). [4] M. W. Schmidt et al., *J. Comput. Chem.*, **14**, 1347(1993); <http://www.msg.ameslab.gov/GAMESS/GAMESS.html>.
 [5] <http://unit.aist.go.jp/tacc/ci/supercluster.html>.