

1A14

フラグメント分子軌道法における post-HF 計算の高速化

(立教大¹、JST-CREST²、NECソフト³、国立衛生研⁴、神戸大院⁵)

望月祐志^{1,2}、山下勝美³、中野達也^{2,4}、田中成典^{2,5}

【序】 北浦ら[1]によって提案されたフラグメント分子軌道(FMO)法は、タンパク質や水和凝集系といった巨大分子を量子論的なフレームワークで実行時間内に扱える手法の一つであり、計算対象内部の相互作用の様態を解析することも直截なため、FMO法の有効性は多数の実応用計算によって示されてきている[2]。現時点でFMO計算は、Fedorovらによって推進されているGAMESS[3]、ならびに、私たちが100%自主開発しているABINIT-MP[4]で利用出来るようになっている。この講演では、post-HFと物性値評価の機能を強化した開発/実験版であるABINIT-MPXの最近の高速化と拡張についてまとめて報告する。

【MP2計算】 よく知られているように、MP2計算のコストは N^5 オーダーの積分変換の処理によって完全に支配される。積分変換については並列化による加速がきわめて有効であり、これまでに数多くのアルゴリズムと計算エンジンが開発されてきたが、私たちのもの[5,6]は実効演算数と通信コストを最小化しつつ、ベクトル-ベクトル型BLASライブラリであるDAXPYとDDOTを最深部に用いることで、汎用のクラスター計算機上でも高い実用性を示す特徴があり、実のところABINIT-MP(X)による応用計算ではFMO-MP2が最もよく使われている。タイミング事例としては、198残基のHIV-1プロテアーゼ+ロピナビル複合体のFMO-MP2/6-31G*ジョブ(原子数:3,196、基底総数:26,801)が、典型的なインハウスのPCクラスター資源と言える16コアのIntel Core2 Duoマシン(2.93GHz, 4GB/core)上では36.1時間で完了する。しかし、地球シミュレータ(ES)等のベクトル処理プロセッサを積む超並列スーパーコンピュータ上の実行では、メモリ転送/演算効率の観点から行列-行列型のDGEMMライブラリを使う方が性能の向上が期待出来るため、以下のようなループ構造を持つMP2変換ルートを新設し、ES(1プロセッサの理論速度は8GFLOPS)上で4,096プロセッサまでを用いた一連のベンチマーク計算を行った[7]。

Loop over ij -batch ! size depending on available memory

Loop over s ! to be parallelized

Loop over r

Preparing (pq,rs) ! for canonical pq -pair

$$(iq,rs) = \sum_p C_{pi}(pq,rs)$$

Forming (iq,rs) ! DGEMM, fixed rs , running over p

Forming (ia,rs) ! DGEMM, fixed rs , running over q

$$(ia,rs) = \sum_q C_{qa}(iq,rs)$$

Forming (ia,js) ! DGEMM, fixed s , direct-product for fixed r

$$(ia,js) = \sum_r C_{rj}(ia,rs)$$

End of loop over r

Forming (ia,jb) ! DGEMM, direct-product for fixed s

$$(ia,jb) = \sum_s C_{sb}(ia,js)$$

End of loop over s ! all-reduce operation as barrier

Calculate partial MP2 energy with respect to ij -batch

End of loop over ij -batch

ここで、並列処理は原子軌道添字 s によって駆動され、同期を伴う通信処理は4段階目変換が終わった後に只一度だけ発生することに注意していただきたい。ベンチマークの結果、この新しい

FMO-MP2 エンジンでは ES 上で高いパフォーマンスを示した。その一例を挙げれば、921 残基から成るインフルエンザ抗原と抗体のモデル系の FMO-MP2/6-31G ジョブ(原子数：14,086、基底総数：78,390)の実行時間は 2,048 プロセッサで 91.4 分、4,096 プロセッサでは 53.4 分で、後者では ES の理論ピーク性能値に対しても効率 11.5%を保持しており、種々のアミノ酸から構成される実際の巨大タンパク質に対する実効性能としては十分に及第レベルにあると考えている。

【MP3 計算】 形式 N^6 オーダーとなる MP3 では電子対間の相互作用が考慮され、MP2 での相関エネルギーの過大評価傾向の是正が見込まれる。また、Zn などの金属原子を含むタンパク質への適用も期待される。実装面では、MP3 は MP4(SDQ)や CCSD への“準備”として位置付けられる。今回の MP3 計算の並列・ファイルレス実装では、文献[5-7]でのアルゴリズムを発展させ、4 粒子-2 空孔項に現れる (ac, bd) 積分を露に生成しないで済ますなどの工夫と複合的な並列処理により、実タンパク質での実用性を確保した。前述の HIV-1 の系を同じ 16 コアの PC クラスタで FMO-MP3 ジョブとして実行すると計算時間は 302.0 時間で、FMO-MP2 ジョブとの相対コストは 8.4 倍である。もしメモリがより潤沢な環境(6GB/core 等)で実行するのなら、積分処理のバッチ回数が減るために相対コストのさらなる低下が望める。

【2 次相関密度計算】 文献[5,6]段階でも 2 次相関密度の実装を行い、DNA-シスプラチン複合体の MCP 計算[8]などに供してきた。しかし、この古い密度エンジンでは応答要素に随伴する (ib, ac) リストを露に作るために系によってはバッチ回数が増大してしまう難点があった。そこで、MP3 の (ac, bd) の処理と同様の技法を使ってこの問題を回避し、DGEMM を多用する新エンジンを開発した。もちろん、このカーネルは MP2 エネルギー微分でのラグランジアン構築に転用出来る。

【CIS(D)系計算】 CIS(D)系の励起状態計算[9]では、差分相関エネルギー (T_2U_1 項)を部分再規格化し、さらに余剰緩和エネルギー (T_1U_1 項)を追加補正する PR-CIS(Ds)[10]、さらに 2 次自己エネルギーのシフトにより高次相関を繰込んで緩和エネルギー (U_2 項)を修飾する CIS(D)_{SS(2)}[11]も並列・ファイルレスでの実装が済んでいる。修正 CIS(D)の効果だが、例えば DNA の励起状態計算のテストとして行ったシトシン分子単体の最低 *励起エネルギー評価の場合、6-31G*基底の計算では元の CIS(D)では 5.04eV となって実験値の 4.5-4.6eV と比べて差があるが、再規格化と自己エネルギーを補正する PR-CIS(D)_{SS(2)}では 4.51eV となるなどの改善が得られる。

【謝辞】 本研究は、JST-CREST プロジェクト「フラグメント分子軌道法による生体分子計算システムの開発」、科研費-特定領域「実在系の分子理論」公募研究の一環として行っている。また、東大生産研・文科省 IT プロジェクト「革新的シミュレーションソフトウェアの研究開発」、海洋機構・地球シミュレータセンター(ESC)からもご支援をいただいたことを記しておく。

【文献】 [1] 北浦ら CPL 313(1999)701. [2] Fedorov ら JPC A111(2007)6904. [3] Fedorov ら JCC 25(2004)872. [4] 中野ら CPL 351(2002)475. [5] 望月ら TCA 112(2004)442. [6] 望月ら CPL 396(2004)473. [7] 望月ら CPL 457(2008)396. [8] 石川ら CPL 427(2006)159. [9] 望月ら TCA 117(2007)541. [10] 望月ら CPL 443(2007)389. [11] 望月 CPL 453(2008)109.