

分子軌道計算プログラム Gaussian03 のパーソナル コンピューター(Linux)での実行とネットワーク並列計算

理学部基礎化学科 坂本 章, 藤森 一希

1. はじめに

米国 Gaussian 社の Gaussian プログラムは最も広く用いられている分子軌道計算プログラムのひとつである。Gaussian プログラムの最新版は 2003 年にリリースされた Gaussian03 であり, 埼玉大学では 2003 年 10 月に Gaussian98 から Gaussian03 (Revision B.04)へアップデートし, 2004 年 12 月に最新のリビジョン(Revision C.02)にアップデートした。Gaussian03 のライセンスは Site-Wide License になっており, ライセンスを取得した機関の中でいくつのコンピューターにインストールして使っても良いことになっている。Gaussian03 の新しい特徴に関しては, Gaussian 社のホームページ("What's New in Gaussian03", http://www.gaussian.com/g_brochures/g03_new.htm)を参照してほしい。Gaussian03 は Linux を含む多くの Unix 系オペレーティング・システム(OS)をベースとしたコンピューターシステムで動作する。ここでは, 最近特に進歩が著しいパーソナルコンピューター上で Linux を OS にした計算環境を作り上げる手順を詳しく述べるとともに, それをベースに我々が製作したネットワーク並列計算システムについて述べる。

2. Gaussian03 のパーソナルコンピューターへのインストール

[1] システム要求条件

Gaussian 社から購入した Gaussian03 のソース・コードは, さまざまな Unix 系 OS でコンパイルできる。Gaussian03 をさまざまな Unix 系コンピューターで動作させるために必要なシステム及びソフトウェアは Gaussian 社のホームページ("Gaussian03 System Requirements", http://www.gaussian.com/g03_plat.htm)にまとめられている。いずれのプラットフォームで動作させるにしても, Gaussian 社の要求する条件を満たすことは重要である。以下に Intel 社の Pentium 系の CPU を用いたシステムで計算環境を作り上げる手順を OS のインストールから順に述べる。

[2] Red Hat Linux 8.0 のインストール

Intel 社の Pentium 系 CPU を用いたシステムの OS としては, SuSE Linux 8.2, 9.0, 9.1 または Red Hat Linux 8.0 が指定されている。我々は Red Hat Linux 8.0 を用いた。インストールを行う際に Install Type で Custom を選択し, Components to Install では Everything を選択しておくことを勧める。

[3] Gaussian03 のインストール手順

Linux のインストール後に Gaussian03 をコンパイルする。キーボードから入力する必要がある事項は, すべてリスト 1 にまとめた。以下, リスト 1 の番号にそって解説する。ただし, ここでは Gaussian03 をコンパイルしたコンピューターを計算専用を使うことを前提としている。

(1) Fortran コンパイラーなど必要なファイルをダウンロードしてそろえる。The Portland Group 社の PGI Fortran Compiler (PGF77) Ver.5.1-6 は, The Portland Group 社(<http://www.pgroup.com/>)から購入及びダウンロードができるほか, 日本での代理店のソフテック社(<http://www.softek.co.jp/>)とベストシステムズ社(<http://www.bestsystems.co.jp/>)から購入及びダウンロードができる。リスト 1 の手順は必要なファイルがすべて CD-ROM(CD-R)にまとめてあるものとして書かれているので, それ以外の場合は適当に読みかえる必要がある。

リスト1 Gaussian03 のインストール手順

(2) root でログインして Terminal を起動する.

```
# adduser gauss03
# passwd gauss03      パスワードを設定する.
```

(3) # chown gauss03 /usr または
chown gauss03 /home # chmod 777 /usr
chgrp gauss03 /usr # chmod 777 /home
chgrp gauss03 /home

(4) # vi /etc/passwd
最終行の gauss03: _____/bash を gauss03: _____/csh に変える.

(5) root をログアウトして, gauss03 でログインして Terminal を起動する.
CD を入れると自動的に /mnt/cdrom にマウントする.

```
$ cp /mnt/cdrom/linux86.tar.gz /tmp/
$ setenv LANG C
$ cd /tmp/
$ gunzip linux86.tar.gz
$ tar xpf linux86.tar
$ ./install
   インストーラーの指示どおりにインストール作業を行う.
$ setenv PGI /usr/pgi
$ set path = ($PGI/linux86/5.1/bin $path)
```

(6) CD を入れると自動的に /mnt/cdrom にマウントする.

```
$ setenv mntpnt /mnt/cdrom
$ setenv g03root /home
$ cd $g03root
$ cat $mntpnt/tar/*.taz | zcat | tar xvf -
$ chgrp -R gauss03 g03
$ cd g03
$ ./bsd/install      gau-machine: Command not found. メッセージが3つ出たが無視した.
$ source $g03root/g03/bsd/g03.login gau-machine: Command not found. メッセージが3つ出たが無視した.
$ bsd/bldg03 >& bldg03.log &
$ tail -f bldg03.log
   正しくコンパイルされると終了までに15-20分ほどかかる(P4: 3.2 GHzの場合).
   エラーメッセージがなく,
   endif
   chmod -R o-rwx ..... wrappers.F xcind.inc
   という終わり方をすれば正常にインストールできている. Ctrl + C で tail コマンドを終了する.
$ ls $g03root/g03/*.exe      実行ファイルが80個できていれば良い.
```

(7) \$ vi /home/gauss03/.login
.login ファイルに以下の5行を追加する.

```
setenv PGI /usr/pgi
set path = ($PGI/linux86/5.1/bin $path)
setenv g03root /home
source $g03root/g03/bsd/g03.login
setenv GAUSS_SCRDIR /scr
```

(8) \$ su
chown gauss03 /scr または
chgrp gauss03 /scr # chmod 777 /scr
exit

一度ログアウトして, gauss03 で再びログインすれば計算実行可能.

- (2) root でログインしてユーザーを設定する.
- (3) 以下の作業をユーザー(ここでは gauss03)で実行できるように, /usr および Gaussian03 をインストールしようとするディレクトリー(ここでは/home)のパーミッションを root で変更する.
- (4) 以後のインストールに csh を用いるために gauss03 のデフォルトのシェルを root で変更する.
- (5) PGF77 をインストールする(ソフテック社のホームページに, 日本語によるインストール手順があるので参考にするといよい. その際, 環境変数の設定において, MANPATH と LM_LICENSE_FILE は設定しなくてもよい).
- (6) Gaussian03 のコンパイルを行う.
- (7) ユーザー(gauss03)のホームディレクトリーに存在する.login ファイルに 5 行追加する.
- (8) スクラッチファイルを書き出すディレクトリー(/scr)のパーミッションを root で変更する. 一度ログアウトして, 再びユーザー(gauss03)でログインすれば Gaussian03 で計算実行可能になる.

3. ネットワーク並列計算システム

Gaussian03 は, いろいろなコンピューターシステムで並列計算を行うことができる(詳しくは Gaussian 社のホームページ "Available Linda Versions", http://www.gaussian.com/linda_plat.htm を参照). 我々はこれまでに, Scientific Computing Associates 社(<http://www.lindaspaces.com/about/index.html>)の TCP Linda プログラムを用いて, Dual CPU コンピューター(Linux)4 台を用いた Gaussian98 プログラムのネットワーク並列計算システム((Intel Pentium III 500 MHz×2)×4)を製作した. 本稿では, Gaussian03 プログラムの導入にともない, 2 台のパーソナルコンピューター(Linux)を用いた Gaussian03 プログラムのネットワーク並列計算システムを新たに製作し, 以前に製作したシステムとの性能の比較などを行った.

[1] 製作したシステム

製作したネットワーク並列計算システムの模式図を図 1 に示す. それぞれのコンピューターは, CPU として Intel Pentium IV 3.2 GHz (Extreme Edition)を搭載している. ほぼ同じ構成のマシン 2 台が 1000Base-T ネットワークカード(オンボード)でスイッチング・ハブによってネットワーク並列に接続されている. キーボード, マウス, ディスプレイは, CPU 切替え器を用いることによって 1 組ですませた. TCP Linda を用いて Gaussian03 をネットワーク並列で実行する場合には 1 台が親機となり, このマシンが他のマシンにネットワークを介して計算を振り分けながら全体の計算を進めていく. 図 1 のシステムでは, それぞれの CPU を用いた計算を独立に 2 つ実行することもできる.

[2] ネットワーク並列計算のパフォーマンス

Gaussian03 と TCP Linda を用いてネットワーク並列で計算できる計算方法とその種類を以下に示す(詳しくは Gaussian 社のホームページ "Using Gaussian 03 with Linda", http://www.gaussian.com/g_tech/linda_use.htm を参照).

- (1) Hartree-Fock 法: シングルポイントエネルギー計算, 構造最適化, 振動数計算
- (2) CIS=Direct 法: シングルポイントエネルギー計算, 構造最適化, 振動数計算
- (3) 密度汎関数(Density Functional Theory)法: シングルポイントエネルギー計算, 構造最適化, 振動数計算, TDDFT エネルギー計算
- (4) Moller-Plesset 2 次摂動(MP2)法: シングルポイントエネルギー計算, 構造最適化

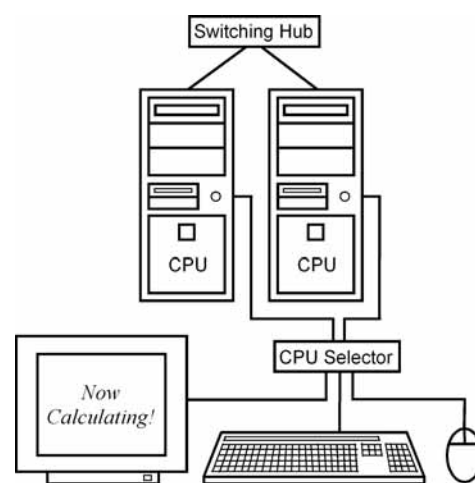


図 1 新たに製作したネットワーク並列計算システムの模式図(Pentium IV (E.E.) 3.2 GHz×2).

次に、我々が製作したシステムで実際に計算を行って測定した、計算速度の向上に対する並列計算の効果を図 2 に示す。計算時間は CPU タイムの合計とした。黒丸(●)は Pentium IV 3.2 GHz を 1 個用いて計算した時の計算速度を 1 として、Pentium IV 3.2 GHz を 2 個用いて計算した時の計算速度をノード数(CPU の数: 2)に対してプロットしたものである。計算は、タンパク質中のアスパラギン酸残基のモデル分子 ($\text{CH}_3\text{CONHCHRCONHCH}_3$, $\text{R}=\text{CH}_2\text{COO}^-$) に対して、計算レベルは HF 法で、SCRF(Self-Consistent Reaction Field)法を適用し、基底関数には 6-31+G** を用いて構造最適化を行った(#P HF/6-31+G** SCRF=Dipole OPT)。全部で 302 個の basis functions から成る。初期構造は、Protein Data Bank に登録されているカルシウム結合タンパク質 (pike parvalbumin, pI = 4.10) のアスパラギン酸残基のまわりの構造を用いた。図 2 から明らかなようにノード数 (CPU 数) が 2 までは、驚くほど良い線形性で計算速度が向上する。図 2 の白丸(○)は参考データとして、以前に製作したネットワーク並列計算システム(Gaussian98 プログラム, (Intel Pentium III 500 MHz×2)×4)において、Pentium III 500 MHz をそれぞれ 1, 2, 4 個用いて計算した時の計算速度を、ノード数として 500 MHz / 3.2 GHz = 0.156 の 1, 2, 4 倍、すなわち 0.156, 0.313, 0.625 に対してプロットしたものである。Pentium III クラスター(図 2 ○)においてもノード数に対する計算速度の向上は良い線形性を示しているが、その傾きは Pentium IV クラスター(図 2 ●)の傾きより小さい。これはキャッシュ量の違いなどを含む Pentium III と Pentium IV の基本的な性能の差が現れていると考えられる。最後に、Pentium IV クラスターにおいてもノード数が 4 程度までは良い線形性を保つと予想されるので、今後ノード数を増やしていきたいと考えている。

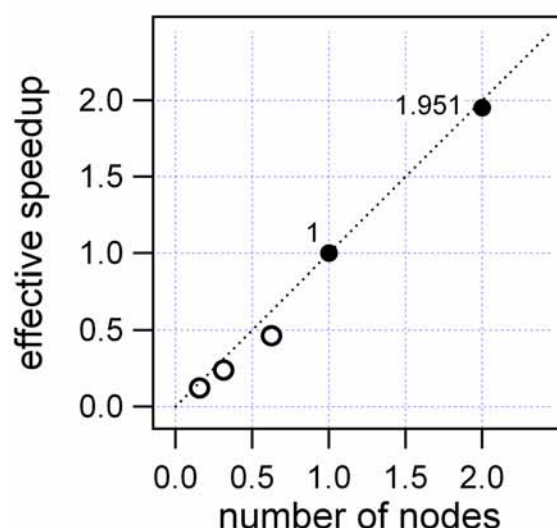


図 2 ネットワーク並列計算におけるマシン数と計算速度の関係(●: Pentium IV 3.2 GHz×ノード数). 参考データ ○: Pentium III 500 MHz×4. 点線はノード数に対し完全に線形で計算速度が向上した場合.

本文中に記載されている会社名、製品名は、各社の登録商標または商標です。