

# 量子化学计算程序包GAUSSIAN

王涛

上海超级计算中心 上海 201203 twang@ssc.net.cn

## 摘要：

GAUSSIAN是一款商业化的量子化学计算程序包。本文对Gaussian的发展背景和程序使用的方法进行了初步介绍，并描述了GAUSSIAN在上海超级计算中心曙光4000A上的安装过程，同时给出了该软件在曙光4000A上的使用方法。

## 背景介绍

Gaussian是目前计算化学领域内最流行、应用范围最广的商业化量子化学计算程序包。它最早是由美国卡内基梅隆大学的约翰·波普(John A Pople, 1998年诺贝尔化学奖)在60年度末、70年代初主导开发的。其名称来自于该软件中所使用的高斯型基组。最初, Gaussian的著作权属于约翰·波普供职的卡内基梅隆大学; 1986年, 约翰·波普进入美国西北大学后, 其版权由Gaussian, Inc.公司所持有。Gaussian软件的出现降低了量子化学计算的门槛, 使得从头计算方法可以广泛使用, 从而极大地推动了其在方法学上的进展。

到目前为止, Gaussian已经推出了12个版本, 包括Gaussian70、Gaussian76、Gaussian80、Gaussian82、Gaussian86、Gaussian88、Gaussian90、Gaussian92、Gaussian92/DFT、Gaussian94、Gaussian98、Gaussian03等, 其版本数字也是该版本发布的年份。其中, 每个版本发布后, 还陆续发布了一些这些版本的修订版。目前最新的版本是Gaussian03 Revision D.01/D.02。

Gaussian程序是用FORTRAN语言编写的, 它从量子力学的基本原理出发, 可计算能量、分子结构、分子体系的振动频率以及大量从这些基本计算方法中导出的分子性质。它能用于研究不同条件下的分子和反应, 包括稳定的粒子和实验上难以观测的化合物, 例如瞬时的反应中间物和过渡结构。

Gaussian的并行模式是采用OpenMP来实现的。OPENMP的并行实现是针对共享内存的机器的, 实现方法简单。因此Gaussian在共享内存的机

器上, 能获得很好的性能。对于跨节点的计算, Gaussian使用TCP Linda软件来实现。TCP Linda 是一个虚拟共享内存的并行执行环境, 它可以把一个通过网络连接的分布式内存的机群或工作站虚拟成共享内存环境, 从而使像Gaussian这样的用OPENMP实现并行的程序能够在分布式内存的机器上运行。

## 软件功能

Gaussian基本实现了量子化学和分子动力学的各种计算方法, 它可以通过ONIOM(最早在Gaussian98中采用)技术来研究生物大分子和蛋白质的反应和光谱、通过计算自旋偶合常数来决定构型、通过周期性边界条件方法来预测聚合物和结晶体这样的周期性体系、计算各种光谱和光谱性质、模拟溶剂对化学反应和分子性质的影响等等。它的基本特性算法如下:

- 可对任何一般收缩的gaussian函数计算单电子和双电子积分。这些基函数可以是笛卡尔高斯函数或纯角动量函数, 并且有多种基组存储于程序中, 可以通过名称调用。积分可储存在内存和外接存储器上, 或在用到时重新计算。对于确定类型的计算, 计算的花费可以使用快速多极方法(FMM)和稀疏矩阵技术实现线性化。

- 将原子轨道(AO)积分转换成分子轨道基的计算, 可用的方法有“in-core”(将AO 积分全部存在内存里)、“直接”(不需储存积分)、“半直接”(用磁盘储存部分积分)或“传统”方法(把所有AO 积分储存在磁盘上)。

- 使用密度拟合, 加速纯DFT计算的库仑部分。
- 数值求积分, 计算DFT 的XC 能量及其微分。

目前在Gaussian03中实现的功能包括如下几个方面:

### 1. 能量

- 使用AMBER、DREIDING和UFF力场的分子力学计算。

- 使用CNDO、INDO、MINDO/3、MNDO、AM1和PM3模型的哈密顿量的半经验计算。

- 使用闭壳层(RHF)、自旋非限制开壳层(UHF)、自旋限制开壳层(ROHF)Hartree-Fock 波函数的自洽场(SCF)计算。

- 使用二级、三级、四级和五级Møller-Plesset 微扰理论计算相关能。MP2 计算可用直接和半直接方法, 有效地使用可用的内存和磁盘空间。

- 用组态相互作用(CI)方法计算相关能, 既可使用全部双激发(CID), 也可使用全部单激发和双激发(CISD)。

- 双取代的耦合簇理论(CCD)、单双取代耦合簇理论(CCSD)、单双取代的二次组态相互作用(QCISD)和Brueckner 轨道双取代耦合簇理论(BD)。还可以计算非迭代的三取代(QCISD 和BD 还有四取代)的贡献。

- 密度泛函理论(DFT), 一般还包括用户可以自行组合的Hartree-Fock 和DFT的混合方法。

- 自动的高精度能量方法: G1理论、G2理论、G2(MP2)理论、G3理论、G3(MP2)理论以及其它的变体; 完全基组(CBS)方法: CBS-4、CBS-q、CBS-Q、CBS-Q//B3和CBS-QCI/APNO以及一般的CBS外推方法; Martin 的W1 方法(加上了微小的修改)。

- 一般的MCSCF, 包括完全活性空间SCF(CASSCF)方法, 并可选择包含MP2相关能计算。可处理14 个以内的活性轨道, 还支持RASSCF变体。

- 广义价键-完全电子对(GVB-PP) SCF 方法。

- 对Hartree-Fock和DFT方法计算在取消限制后, 测试SCF波函数的稳定性。

- 使用单激发组态相互作用(CI-Singles)方法、HF和DFT的含时方法、ZINDO半经验方法以及Nakatsuji 等人的对称性匹配簇/组态相互作用(SAC-CI)方法, 计算激发态能量。

### 2. 梯度和几何优化

- 解析计算RHF、UHF、ROHF、GVB-PP、CASSCF、MP2、MP3、MP4(SDQ)、CID、CISD、CCD、CCSD、QCISD、密度泛函和激发态CIS能量的核坐标梯度。以上的后-SCF方法可以利用冻结核近似。

- 使用内坐标、笛卡尔坐标或混合坐标, 自动几何优化到能量最小或鞍点结构。不论输入使用何种坐标系, 执行优化计算默认使用冗余内坐标。

- 使用同步过渡引导的准Newton方法, 自动进行过渡态搜索。

- 使用内反应坐标(IRC)计算反应路径。

- 对能量和几何优化进行二或三层的ONIOM计算。

- 同时优化过渡态和反应路径。

- 使用态平均CASSCF进行圆锥截点的优化。

- 沿着指定的反应路径, 定位过渡结构能量最大点的IRCMAX计算。

- 经典的动力学轨迹计算, 其中的经典运动方程集成了解析二阶导数, 可以使用:

- Born-Oppenheimer分子动力学(BOMD)。它可以使用任何可以计算解析梯度的方法, 也可以利用Hessian的信息。

- 通过原子为中心的密度矩阵传播分子动力学模型, 传播电子自由度。这一方法与有关的Car-Parrinello方法既有相似, 又有不同。这可以用AM1、HF和DFT方法完成。

### 3. 频率和二阶导数

- 对RHF、UHF、DFT、RMP2、UMP2和CASSCF方法, 以及用CIS方法计算的激发态等, 解析计算力常数(对核坐标的二阶导数)、极化率、超极化率和偶极矩的解析导数。

- 通过对MP3、MP4(SDQ)、CID、CISD、CCD和QCISD方法的能量梯度的数值微分, 计算力常数、极化率和偶极矩导数。

- 使用任意同位素、温度和压强做谐振分析和热化学分析。

- 对内坐标分析简正模式。

- 确定振动跃迁的红外和拉曼强度。也可以确定预共振拉曼强度。

- 谐性振-转耦合。

- 非谐性振动和振-转耦合。非谐性振动可用于能计算解析二阶导数的方法。

### 4. 分子性质

- 使用SCF、DFT、MP2、CI、CCD、QCISD方法求解各种单电子性质, 如Mulliken布居分析、多极矩、自然布居分析、静电势以及用

Merz-Kollman-Singh方法、CHelp方法或CHelpG方法计算静电势导出电荷。

- 用Hartree-Fock和DFT方法计算静态的或含频的极化率和超极化率。
- 用SCF、DFT和MP2方法，计算NMR屏蔽张量和分子的磁化系数；用GIAO方法计算磁化系数；在Hartree-Fock和DFT级别计算自旋-自旋耦合常数。
- 振动圆二色性(VCD)强度。
- 用传播子方法计算电子亲和能和电离势。
- CASSCF 计算中可计算两自旋状态间的近似自旋-轨道耦合。
- 电子圆二色性。
- 用GIAO 计算旋光性和旋光色散。
- 超精细谱：g 张量、核的电四极常数、转动常数、四阶离心畸变项、电子自旋转动项、核自旋转动项、偶极超精细项以及Fermi接触项。输入可以用广泛使用的H. M. Pickett 的程序产生。

## 5. 溶解模型

所有这些模型均使用自洽反应场(SCRF)方法模拟溶液中的体系。

- Onsager模型（偶极和球反应场），包括在HF和DFT级别的解析一阶和二阶导数，在MP2、MP3、MP4(SDQ)、CI、CCD和QCISD级别的单点能计算。
- 对于解析HF、DFT、MP2、MP3、MP4(SDQ)、QCISD、CCD、CCSD、CID和CISD的能量，以及HF和DFT的梯度和频率、Tomasi 等人的极化连续（重叠球）模型(PCM)：
  - 对激发态计算溶剂影响。
  - 计算溶剂出现时的很多特性。
- 在HF和DFT级别，能量的IPCM（静态等密度曲面）模型。
- 在HF和DFT级别上的SCI-PCM（自洽等密度曲面）模型，用解析方法计算能量和梯度，数值方法计算振动频率。

## 移植和安装

Gaussian系列程序可通过向Gaussian, Inc公司及其代理商购买获得。对于Gaussian98及其以前的版本，Gaussian, Inc公司出售源代码和已编译好的可执行程序给全世界用户；而对于目前的Gaussian03，出于某种原因，Gaussian, Inc公司只出售已编译好的可执行程序给中国大陆用户，而对中国大陆以外的用户，仍然出售源代码和已编译好的可执行程序。用户获得Gaussian03的可执行程序后，只需要做一定的环境变量的设定就可以使用了。下文给出了Gaussian03在上海超级计算中心曙

光4000A机群上的安装过程。

### 1. 安装的软硬件要求：

- 根据系统的不同，Gaussian目录下可执行文件需要大约150-750 MB 磁盘空间。
- 对于Linux系统，同时计算的每个Gaussian任务需要使用大约64 MB交换空间，使用默认的48 MB的动态内存分配。当然，如果一个任务通过使用Link 0 命令%Mem，或者在Default.Route 文件使用-M-命令来占用更多的内存，将会需要更多的交换空间。
- 对于特殊的操作系统或编译需要，参看随CD 一起的发布宣传品，或者联系Gaussian 公司。

### 2. 上海超级计算中心曙光4000A软硬件环境：

硬件环境：曙光4000A机群系统，每节点四路Opteron 850 CPU、8G内存，Myrinet2000互联计算网络

操作系统：Turbo Linux 8.0

### 3. 软件的安装

#### 3.1 设置Gaussian 运行环境：

在将Gaussian03解压在用户目录下并取名为g03后，修改用户目录下的.bash\_profile文件，加入下列内容：

```
g03root=~
export g03root
. $g03root/g03/bsd/g03.profile
export PATH=$PATH:$g03root/g03:$g03root/g03/bsd
export GAUSS_SCRDIR=~/.scratch/gaussian
export LD_LIBRARY_PATH=$LD_LIBRARY_PATH:$g03root/g03
```

执行source .bash\_profile命令。

#### 3.2 修改~/g03/bsd/ g03.profile

将该文件中的LD\_LIBRARY\_PATH=\$GAUSS\_EXEDIR改为LD\_LIBRARY\_PATH=\$LD\_LIBRARY\_PATH:\$GAUSS\_EXEDIR

3.3 在用户目录下建立文件夹scratch/gaussian作为临时文件存储目录：

```
mkdir -p ~/.scratch/gaussian
```

## 使用方法

使用Gaussian03计算，用户只需要准备一个输入文件就可以了。

### 1. 建立输入文件

Gaussian的输入文件由一个ASCII文件组成，包含一系列的行，其基本结构如下：

输入文件部分	用途	是否需要空行作为结束？
Link 0命令行（以%开头）	说明计算时的系统信息，如内存、硬盘大小、CPU情况等	否
计算方法说明（以#开头）	指定需要的计算类型，模型化学，以及其它选项。	是
标题行	计算的简要说明	是
分子说明行	定义要研究的分子体系，包括分子电荷、多重态和分子结构等	是
可选的附加部分	通常用于特殊任务类型的输入需要	通常需要

一般情况下，Gaussian 输入遵循下面的语法规则：

- 输入是自由格式，且大小写无关。
- 空格，TAB 键，逗号，正斜杠“/”都可以作为一行内不同项之间的连接符。多个空格作为一个分隔符处理。

- 关键字的选项可以用以下格式指定：

keyword = 选项

keyword ( 选项 )

keyword= ( 选项1, 选项2, ..... )

keyword ( 选项1, 选项2, ..... )

多个选项放在括号中，并用任何有效的分隔符分开（习惯用逗号，如上所示）。左括号前的等号可以忽略，这个等号的前后也可以随意加空格。注意有些选项带有数值；本例中，选项名的后面接等号：如CBSEtrap(NMin=6)。

- 在整个Gaussian 03 系统中，所有的关键字和选项都可以使用能相互区分开的缩写形式。因此，SCF 关键字的Conventional 选项可以缩写成Conven，但不能写成Conv（因为还存在Convergence 选项）。无论Conventional 和Convergence 是否同时出现，对任何给定的关键字都要保证是有效选项。

- 在Gaussian 03 输入文件中，可以引用外部文件内容，使用的语法为：@文件名。这使得整个文件放到输入命令串的当前位置。命令加上/N 能防止被包含文件的内容在输出文件的开始部分回显。

- 注释行以感叹号(!)开始，可以出现在一行中的任何位置。输入文件中可以随处加入整行的注释行。

下面以一个简单的输入文件作为例子进行说明。

```
%nproc=4
# SP, UHF/6-311G**
```

example

```
-2 1
H .000000 .000000 2.091575
```

```
C .000000 .000000 1.000991
C -1.223016 .000000 .319391
C 1.223016 .000000 .319391
C -2.481434 .000000 1.044032
C -1.205753 .000000 -1.082635
C 1.205753 .000000 -1.082635
C 2.481434 .000000 1.044032
C .000000 .000000 -1.772474
H 2.147747 .000000 -1.629849
H 3.429332 .000000 .514983
H .000000 .000000 -2.861276
H -2.147747 .000000 -1.629849
H 2.501274 .000000 2.129439
H -2.501274 .000000 2.129439
H -3.429332 .000000 .514983
```

第一行的%nproc=4就是所谓的Link 0命令，它说明本次计算用4个CPU；第二行以#开头的是计算方法说明，SP表示任务类型是单点计算，UHF表示使用非限制性Hartree-Fock计算，6-311G\*\*表示计算时使用的基组；下面要空一行，再给出自己的任务简短说明（这里是example）；任务简短说明后还需要再空一行，然后给出分子构型说明，其中，-2表示分子的带电量是-2电荷，1表示分子的多重态是1，后面是分子的笛卡尔坐标表示。关于计算方法说明中的更多关键字，可参考高斯的使用手册。

## 2. 建立作业提交的脚本

上海超级计算中的曙光4000A使用的是PLATFORM公司的LSF作业调度系统，因此需要建立一个LSF脚本文件来进行作业提交。提交作业脚本如下，假定这个脚本的文件名是gaussian.lsf：

```
#!/bin/sh
APP_NAME= " fourcpus "
NP=4
NP_PER_NODE=4
RUN= " RAW "
$g03root /g03/g03 $HOME/test/gaussian/test.com
```

其中test.com是输入文件。注意：高斯的输入文件要以.com或者.gif结尾。此处的输出文件将默认为test.log

将gaussian.lsf的权限加为可执行的：chmod 755 gaussian.lsf，用bsub ./ gaussian.lsf提交作业。

如果要使用高斯的并行计算功能，则要注意下列事项：

在输入文件的第一行加入下列内容：

```
%nproc=CPUNUMBER
```

其中CPUNUMBER要和gaussian.lsf中的NP的数值一致。此外，在gaussian.lsf中，NP\_PER\_NODE的数目要大于或等于NP的数值。目前，在上海超级计算

中心的曙光4000A上，最多可以进行4CPU的并行。

计算过程中会产生一系列中间文件，这些文件将存储在\$GAUSS\_SCRDIR的位置。如果计算正常结束，这些临时文件会自动删除，如果是非正常结束，这些文件会被保留，因此要注意定时清理\$GAUSS\_SCRDIR这个目录。

### 3. 其它说明

Gaussian03中只有部分方法实现了并行化，并且在整个程序中，不同的功能被独立成功能模块执行文件，也就是用户能够观察到的进程名。下表给出了各个功能模块的执行文件的说明。

功能模块文件名 (以.exe结尾)	实现的功能	是否并行
L0	初始化程序，控制占位	
L1	处理计算执行路径，创建执行链接的列表，并初始化scratch文件	
L101	读取标题和分子说明部分	
L102	FP 优化	
L103	Berny 优化到最小值和TS，STQN 过渡态寻找	
L105	MS 优化	
L106	力常数或偶极矩的数值微分，用以获得极化率或超极化率	
L107	线性同步过渡(LST)的过渡态寻找	
L108	势能曲面扫描	
L109	Newton-Raphson 优化	
L110	能量的二阶数值微分产生频率	
L111	能量的二阶数值微分，用以计算极化率和超极化率	
L113	使用解析梯度进行EF 优化	
L114	EF 数值优化(仅使用能量)	
L115	使用内反应坐标(IRC) 追踪反应路径	
L116	数值的自恰反应场(SCRF)	
L117	后-SCF SCRF	
L118	轨迹计算	
L120	控制ONIOM 计算	
L121	ADMP 计算	
L122	平衡计算	
L202	重新定位坐标，计算对称性，检查变量	
L301	产生基组信息	
L302	计算重叠积分，动能和势能积分	并行
L303	计算多极积分	
L308	计算偶极速率和 $R_x \nabla$ 积分	
L310	用简单方式计算spdf 双电子积分	
L311	计算sp 双电子积分	
L314	计算spdf 双电子积分	
L316	打印双电子积分	
L319	对近似的自旋-轨道耦合计算单电子积分	
L401	形成初始MO 猜测	

L402	进行半经验和分子力学计算	
L405	初始化MCSCF 计算	
L502	迭代求解SCF 方程(用于通常的UHF和ROHF , 所有的直接方法, 以及SCRF)	并行
L503	使用直接最小化迭代求解SCF 方程	
L506	进行ROHF 或GVB-PP 计算	并行
L508	二次收敛SCF 程序	并行
L510	MC-SCF	并行
L601	布居数和相关的分析(包括多极矩)	
L602	单电子特性(势, 场, 和场梯度)	并行
L604	求解MO, 或一个网格点的密度	
L607	进行NBO 分析	
L608	非迭代DFT 能量	
L609	分子中的原子特性	
L701	单电子积分一阶或二阶导数	并行
L702	双电子积分一阶或二阶导数(sp)	
L703	双电子积分一阶或二阶导数(spdf)	并行
L716	处理优化和频率的信息	
L801	双电子积分变换的初始化	
L802	进行积分变换(对in-core 方法是N3)	
L804	积分变换	
L811	变换积分导数, 并计算其对MP2 二阶导数的贡献	
L901	反对称化双电子积分	
L902	确定Hartree-Fock 波函的稳定性	
L903	旧的in-core MP2	
L905	复合的MP2	
L906	半直接的MP2	并行
L908	OVGF (闭壳层)	
L909	OVGF (开壳层)	
L913	计算后-SCF 能量和梯度项	并行
L914	CIS, RPA 和Zindo 激发能; SCF 稳定性	并行
L915	计算五阶的量(用于MP5, QCISD(TQ) 和BD(TQ))	
L916	旧的MP4 和CCSD	
L918	重新优化波函	
L1002	迭代求解CPHF 方程; 计算各种特性(包括NMR)	并行
L1003	迭代求解CP-MCSCF 方程	
L1014	计算解析的CIS 二阶导数	并行
L1101	计算单电子积分导数	并行
L1102	计算偶极导数积分	
L1110	双电子积分导数对Fock 矩阵的贡献	并行
L1111	2 PDM 和后-SCF 导数	
L1112	MP2 二阶导数	并行
L9999	完成计算和输出	

## 总结

Gaussian03是一个当前计算化学领域最流行、应用范围最广的量子化学程序包，可以实现很多计算化学领域内的计算要求，具有相当的准确性和计算速度。但是它对高精度的计算方法实现不足，例如

没有实现CCSD ( T )、CCSDTQ以及激发态的EOM-CCSD等方法；此外它的并行是通过OPENMP来实现的，对CLUSTER型的高性能计算机的使用有一定的限制。

## 参考文献

- [1] Gaussian 03 , Revision A.1, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople, Gaussian, Inc., Pittsburgh PA, 2003.
- [2] Gaussian 03 Online Manual [http://www.gaussian.com/g\\_ur/g03mantop.htm](http://www.gaussian.com/g_ur/g03mantop.htm)
- [3] <http://www.quantumchemistry.net/Programs/G/200512/304.html>