Introduction to Gaussian program[1]

In this lab, we will use the Gaussian program in Windows environments. Gaussian is capable of predicting many properties of molecules and reactions, including the following:

- Molecular energies and structures.

- Energies and structures of transition sates

- Bond and reaction energies

- Molecular orbitals

- Multipole moments

- Atomic charges and electrostatic potential

- Vibrational frequencies

- NMR properties

- Reaction pathways

Computation can be carried out on systems in the gas phase or in solutions, and in their ground state or in an excited state.

Contents

1. <u>Gaussian input file</u>

Gaussian input file includes several different sections:

- *Link* 0 *Commands*: Locate and name scratch files. We will not use this option.
- *Route section* (# *lines*): Specify desired calculation type, the method and basis set and other options.
- *Title section:* Brief description of the calculation.
- *Molecule specification*: Specify molecular system to be studied in Cartesian coordinates or by the Z-matrix.
- *Optional additional sections:* Additional input needed for specific job types.

In general, *Gaussian* input is subject to the following syntax rules:

- Input is free-format and case-insensitive.
- Spaces, tabs, commas, or forward slashes can be used in any combination to separate items within a line. Multiple spaces are treated as a single delimiter.
- Options to keywords may be specified in any of the following forms:

  keyword = option
  keyword (*option*)
  keyword= (*option1, option2, ...*)
  keyword (*option1, option2, ...*)

- Multiple options are enclosed in parentheses and separated by any valid delimiter (commas are conventional and are shown above). The equals sign before the opening parenthesis may be omitted, or spaces may optionally be included before and/or after it.

- Note that some options also take values; in this case, the option name is followed by an equals sign: for example, **Opt(MaxCycle=99)**.

- All keywords and options may be shortened to their shortest unique abbreviation within the entire *Gaussian* system. Thus, the **Conventional** option to the **SCF** keyword may be abbreviated to **Conven**, but not to **Conv** (due to the presence of the **Convergence** option). This holds true whether or not both **Conventional** and **Convergence** happen to be valid options for any given keyword.

- Comments begin with an exclamation point (**!**), which may appear anywhere on a line. Separate comment lines may appear anywhere within the input file.

2. Job Types

The route section of a *Gaussian* input file specifies the type of calculation to be performed. There are three key components to this specification:

- The job type
- The method
- The basis set

The following table lists the job types available in *Gaussian*:

- **SP**   Single point energy.
- **Opt**   Geometry optimization.
- **Freq**   Frequency and thermochemical analysis.
- **IRC**   Reaction path following.
- **IRCMax**   Find the maximum energy along a specific reaction path.
- **Scan**   Potential energy surface scan.
- **Polar**   Polarizabilities and hyperpolarizabilities.

- **ADMP** and **BOMD**   Direct dynamics trajectory calculation.
- **Force**   Compute forces on the nuclei.
- **Stable**   Test wavefunction stability.
- **Volume**   Compute molecular volume.
- **Density**=**Checkpoint**   Recompute population analysis only.
- **Guess**=**Only**   Print initial guess only; recompute population analysis.
- **ReArchive**   Extract archive entry from checkpoint file only.

- ***In general, only one job type keyword should be specified. The exceptions to this***

- **Polar** and **Opt** may be combined with **Freq** (although **SCRF** may not be combined with **Opt Freq**). In the latter case, the geometry optimization is automatically followed by a frequency calculation at the optimized structure.
- **Opt** may be combined with **IRCMax** in order to specify options for the optimization portion   of the calculation.

When no job type keyword is specified within the route section, the default calculation type is usually a single point energy calculation (**SP**).

Predicting Molecular Properties The following table provides a mapping between commonly-desired predicted quantities and the *Gaussian 03* keywords that will produce them:

- Atomic charges: **Pop**
- Dipole moment: **Pop**
- Electron density: **cubegen**
- Electrostatic potential: **cubegen**, **Prop**
- Electrostatic-potential derived charges: **Pop**=**Chelp**, **ChelpG** or **MK**
- Hyperfine coupling constants (anisotropic): **Prop**
- Hyperfine spectra tensors (incl. *g* tensors): **NMR** and **Freq**=**(VibRot, Anharmonic)**

- IR and Raman spectra: **Freq**
- Pre-resonance Raman spectra: **Freq**
- Molecular orbitals: **Pop**=**Regular**
- Multipole moments: **Pop**
- NMR shielding and chemical shifts: **NMR**
- NMR spin-spin coupling constants: **NMR**=**SpinSpin**
- Optical rotations: **Polar**=**OptRot CPHF**=**RdFreq**
- Polarizabilities: **Freq**, **Polar**
- Thermochemical analysis: **Freq**
- UV/Visible spectra: **CIS**, **Zindo**, **TD**

3. **Basis sets**[2]

Most methods require a basis set be specified; if no basis set keyword is included in the route section, then the STO-3G basis will be used. The following basis sets are stored internally in the *Gaussian* program (see references cited for full descriptions), listed below by their corresponding *Gaussian* keyword (with two exceptions):

- **STO-3G**
- **3-21G**
- **6-31G**

- **6-311G**

Adding Polarization and Diffuse Functions

Single first polarization functions can also be requested using the usual ***** or ****** notation. Note that (**d, p**) and ****** are synonymous-**6-31G**** is equivalent to **6-31G(d, p)**, for example-and that the 3-21G* basis set has polarization functions on second row atoms only. The + and ++ diffuse functions are available with some basis sets, as are multiple polarization functions. The keyword syntax is best illustrated by example: **6-31+G(3df,2p)** designates the 6-31G basis set supplemented by diffuse functions, 3 sets of d functions and one set of f functions on heavy atoms, and supplemented by 2 sets of p functions on hydrogens.

---

[2] For more information about the basis sets, read the theoretical background.

Adding a single polarization function to **6-311G** (i.e. **6-311G(d)**) will result in one d function for first and second row atoms and one f function for first transition row atoms, since d functions are already present for the valence electrons in the latter. Similarly, adding a diffuse function to the **6-311G** basis set will produce one s, one p, and one d diffuse functions for third-row atoms.

The following table lists polarization and diffuse function availability and the range of applicability for each built-in basis set in *Gaussian*:

Basis Set	Applies to	Polarization Functions
STO-3G	H-Xe	*
3-21G	H-Xe	* *or* **
6-31G	H-Kr	**(3df,3pd)**
6-311G	H-Kr	**(3df,3pd)**

4. Gaussian keywords

The following is are some Gaussian keywords that can be used in the route section of the Gaussian input file. For more keywords, use the http://www.gaussian.com/g_ur/keywords.htm web site.

# #

The route section of a *Gaussian* job is initiated by a pound sign (#) as the first non-blank character of a line. The remainder of the section is in free-field format. For most jobs, all of the information can be placed on this first line, but overflow to other lines (which may but need not begin with a # symbol) is permissible.

Alternative forms:

#N

Normal print level; this is the default.

Additional output is generated. This includes messages at the beginning and end of each link giving assorted machine-dependent information (including execution timing data), as well as convergence information in the SCF.

#T

Terse output: output is reduced to essential information and results.

HF

This method keyword requests a Hartree-Fock calculation. Unless explicitly specified, **RHF** is used for singlets and **UHF** for higher multiplicities. In the latter case, separate α and β orbitals will be computed[3]. **RHF**, **ROHF** or **UHF** can also be specified explicitly.

SCF single point energy calculations involving basis sets which include diffuse functions should use the **SCF**=**Tight** keyword to request tight SCF convergence criteria.

The Hartree-Fock energy appears in the output as follows:

```
SCF Done: E(RHF) =  -74.9646569691      A.U. after   4 cycles
         Convg  =     .6164D-03                -V/T =  2.0063
         S**2   =     .0000
```

The second and third lines give the SCF convergence limit and the expectation value of S^2.

[3] C. C. J. Roothan, Rev. Mod. Phys. **23**, 69 (1951).
J. A. Pople and R. K. Nesbet, *J. Chem. Phys.* **22**, 571 (1954).
R. McWeeny and G. Dierksen, *J. Chem. Phys.* **49**, 4852 (1968).

B3LYP[4]

B3LYP is one of the energy functionals of the density functional methods. The energy is reported in DFT calculations in a form similar to that of Hartree-Fock calculations. Here is the energy output from a **B3LYP** calculation:

```
SCF Done:  E(RB+HF-LYP) =  -75.3197099428     A.U. after    5 cycles
```

One can use the UB3LYP energy functional in the case of open shell systems.

SP

This calculation type keyword requests a single-point energy calculation. *It is the default when no calculation type keyword is specified.*

If no keywords are present in the route section, the calculation defaults to **HF/STO-3G SP**.

Population

This properties keyword controls printing of molecular orbitals and several types of population analysis and atomic charge assignments. The default is to print just the total atomic charges and orbital energies. Populations are done once for single-point calculations and at the first and last points of geometry optimizations.

Population analysis results are given in the standard orientation.

Output controlled by the **Pop** keyword includes:

- Molecular orbitals and orbital energies
- Atomic charge distribution
- Multipole moments: dipole through hexadecapole

---

[4] For more information about the B3LYP energy functional and DFT (density functional theory) method, read appendix A in the theoretical background of the course lab.

By default, *Gaussian* prints molecular orbitals and performs population analyses regarding the MO coefficients from a semi-empirical calculation as coefficients of orthogonalized atomic orbitals (OAO's).

None

No orbitals are printed, and no population analysis is done.

Minimal

Total atomic charges and orbital energies are printed.

Regular

The five highest occupied and five lowest virtual orbitals are printed, along with the density matrices and a full (orbital by orbital and atom by atom) Mulliken population analysis. Since the size of the output depends on the square of the size of the molecule, it can become quite substantial for larger molecules.

Full

Same as the **Regular** population analysis, except that *all* orbitals are printed.

Bonding

Do a bonding population analysis in addition to the standard analysis. This is a Mulliken population analysis in which only density terms involving pairs of basis functions on different centers are retained. The other options control how much is printed.

SCF

This keyword controls the functioning of the SCF procedure. Options are used to specify the desired behavior, alternate algorithms, and so on. Single point direct SCF calculations are run with modest convergence criteria automatically in the interest of speed. The default for this case is sufficient for 0.1 kcal mole^{-1} accuracy in the SCF energy and 3 decimal places in the density matrix. **SCF=Tight** requests full convergence for this case.

Opt

- This keyword requests that a geometry optimization be performed. The geometry will be adjusted until a stationary point on the potential surface is found. Gradients will be used if available. For the Hartree-Fock, CIS, MP2, MP3, MP4(SDQ), CID, CISD, CCD, CCSD, QCISD, CASSCF, and all DFT and semi-empirical methods, the default algorithm for both minimizations (optimizations to a local minimum) and optimizations to transition states and higher-order saddle points is the Berny algorithm using redundant internal coordinates[5] (specified by the **Redundant** option).

Basic information as well as techniques and pitfalls related to geometry optimizations are discussed in detail in chapter 3 of *Exploring Chemistry with Electronic Structure Methods*[6].

MaxCycle=N

Sets the maximum number of optimization steps to N. The default is the maximum of 20 and twice the number of redundant internal coordinates in use (for the default procedure) or twice the number of variables to be optimized (for other procedures).

MaxStep=N

Sets the maximum size for an optimization step (the initial trust radius) to $0.01N$ Bohr or radians. The default value for N is 30.

TS

Requests optimization to a transition state rather than a local minimum.

QST2

Search for a transition structure using the STQN method. This option requires the

[5] C. Peng, P. Y. Ayala, H. B. Schlegel, and M. J. Frisch, *J. Comp. Chem.* **17**, 49 (1996).
A. E. Reed and F. Weinhold, *J. Chem. Phys.* **78**, 4066 (1983).

[6] J. B. Foresman and Æ. Frisch, *Exploring Chemistry with Electronic Structure Methods*, 2nd ed. (Gaussian, Inc., Pittsburgh, PA, 1996).

reactant and product structures as input, specified in two consecutive groups of title and molecule specification sections. Note that the atoms must be specified in the same order in the two structures. **TS** should *not* be specified with **QST2**.

QST3

Search for a transition structure using the STQN method. This option requires the reactant, product, and initial TS structures as input, specified in three consecutive groups of title and molecule specification sections. Note that the atoms must be specified in the same order within the three structures. **TS** should *not* be specified with **QST3**.

COORDINATE SYSTEM SELECTION OPTIONS

Z-matrix

Perform the optimization in internal coordinates.

CalcFC

Specifies that the force constants be computed at the first point using the current method (available for the HF, MP2, CASSCF, DFT, and semi-empirical methods only).

CalcAll

Specifies that the force constants are to be computed at every point using the current method (available for the HF, MP2, CASSCF, DFT, and semi-empirical methods only). Note that vibrational frequency analysis is automatically done at the converged structure and the results of the calculation are archived as a frequency job.

OPTIMIZING TO A TRANSITION STATE OR HIGHER-ORDER SADDLE POINT

Transition State Optimizations Using Synchronous Transit-Guided Quasi-Newton (STQN) Methods. *Gaussian* includes the STQN method for locating transition structures. This method, implemented by H. B. Schlegel and coworkers[7], uses a quadratic synchronous transit approach to get closer to the quadratic region of

---

[7] C. Peng, P. Y. Ayala, H. B. Schlegel, and M. J. Frisch, *J. Comp. Chem.* **17**, 49 (1996).
C. Peng and H. B. Schlegel, *Israel J. Chem* . **33**, 449 (1994).

the transition state and then uses a quasi-Newton or eigenvector-following algorithm to complete the optimization. Like the default algorithm for minimizations, it performs optimizations by default in redundant internal coordinates. This method will converge efficiently when provided with an empirical estimate of the Hessian and suitable starting structures.

This method is requested with the **QST2** and **QST3** options. **QST2** requires two molecule specifications, for the reactants and products, as its input, while **QST3** requires three molecule specifications: the reactants, the products, and an initial structure for the transition state, in that order. *The order of the atoms must be identical within all molecule specifications*

Despite the superficial similarity, this method is very different from the Linear Synchronous Transit method for locating transition structures requested with the now-deprecated **LST** keyword. **Opt=QST2** generates a guess for the transition structure that is midway between the reactants and products in terms of redundant internal coordinates, and it then goes on to optimize that starting structure to a first-order saddle point automatically.

Traditional Transition State Optimizations Using the Berny Algorithm. The Berny optimization program can also optimize to a saddle point using internal coordinates, if it is coaxed along properly. The options to request this procedure are **Opt=TS** for a transition state (saddle point of order 1) or **Opt(Saddle=**N**)** for a saddle point which is a maximum in N directions.

When searching for a local minimum, the Berny algorithm uses a combination of rational function optimization (RFO) and linear search steps to achieve speed and reliability (as described below). This linear search step cannot be applied when searching for a transition state. Consequently, transition state optimizations are much more sensitive to the curvature of the surface. A transition state optimization should always be started using one of the options described above for specifying curvature information. Without a full second derivative matrix the initial step is dependent on the choice of coordinate system, so it is best to try to make the reaction coordinate (direction of negative curvature) correspond to one or two redundant internal coordinates or Z-matrix variables (see the examples below).

IRC

This method keyword requests that a reaction path be followed[8]. The initial geometry (given in the molecule specification section) is that of the transition state, and the path can be followed in one or both directions from that point.

Forward

Follow the path only in the forward direction.

Reverse

Follow the path only in the reverse direction.

MaxPoints=*N*

Number of points along the reaction path to examine (in each direction if both are being considered). The default is 6.

StepSize=*N*

Step size along the reaction path, in units of 0.01 $amu^{1/2}$-Bohr. The default is 10.

MaxCyc=*N*

Sets the maximum number of steps in each geometry optimization. The default is 20.

CalcFC

Specifies that the force constants be computed at the first point.

CalcAll

Specifies that the force constants be computed at every point.

Frequency

This calculation type keyword computes force constants and the resulting vibrational frequencies. Intensities are also computed. By default, the force constants are determined analytically if possible (for RHF, UHF, MP2, CIS, all DFT methods, and CASSCF), by single numerical differentiation for methods for which only first derivatives are available (MP3, MP4(SDQ), CID, CISD, CCD, QCISD, and all semi-empirical methods), and by double numerical differentiation for those methods for which only energies are available.

---

[8] C. Gonzalez and H. B. Schlegel, *J. Chem. Phys.* **90**, 2154 (1989).
C. Gonzalez and H. B. Schlegel, *J. Phys. Chem.* **94**, 5523 (1990).

Vibrational frequencies are computed by determining the second derivatives of the energy with respect to the Cartesian nuclear coordinates and then transforming to mass-weighted coordinates. *This transformation is only valid at a stationary point!* Thus, it is *meaningless* to compute frequencies at any geometry other than a stationary point *for the method used for frequency determination*.

NMR

This properties keyword predicts NMR shielding tensors and magnetic susceptibilities using the Hartree-Fock method, all DFT methods and the MP2 method[9]. Structures used for NMR calculations should have been optimized at a good level of theory.

5. The title section

This section is required in the input, but is not interpreted in any way by the *Gaussian* program. It appears in the output for purposes of identification and description. Typically, this section might contain the compound name, its symmetry, the electronic state, and any other relevant information. The title section cannot exceed five lines and must be followed by a terminating blank line. The following characters should be avoided in the title section: @ # ! - _ \ *all control characters*, *and especially* ^G.

6. Molecule specification

This input section specifies the nuclear positions and the number of electrons of α- and β-spin. There are several ways in which the nuclear configuration can be

[9] J. Gauss, *J. Chem. Phys.* **99**, 3629 (1993).
J. Gauss, PCCP Phys. Chem. *Chem. Phys.* **99**, 1001 (1995).
J. R. Cheeseman, G. W. Trucks, and M. J. Frisch, in prep. (2003).

specified: as a Z-matrix, as Cartesian coordinates, or as a mixture of the two (note that Cartesian coordinates are just a special case of the Z-matrix).

The first line of the molecule specification section specifies the net electric charge (a signed integer) and the spin multiplicity (a positive integer). Thus, for a neutral molecule in a singlet state, the entry **0 1** is appropriate. For a radical anion, **-1 2** would be used.

The remainder of the molecule specification gives the element type and nuclear position for each atom in the molecular system.

The following are the basic formats for specifying atoms within the molecule specification (omitting all of the optional items):

Element-label x y z
Element-label [*n*] *atom1 bond-length atom2 bond-angle atom3 dihedral-angle* [*format-code*]

Although these examples use spaces to separate items within a line, any valid separator may be used. The first form specifies the atom in Cartesian coordinates, while the second uses internal coordinates. Lines of both types may appear within the same molecular specification.

Element-label is a character string consisting of either the chemical symbol for the atom or its atomic number. If the elemental symbol is used, it may be optionally followed by other alphanumeric characters to create an identifying label for that atom. A common practice is to follow the element name with a secondary identifying integer: C1, C2, C3, and so on; this technique is useful in following conventional chemical numbering.

In the first form, the remaining items on each line are Cartesian coordinates specifying the position of that nucleus. In the second form, *atom*1, *atom*2, *atom*3 are the labels for previously-specified atoms which will be used to define the current atoms position (alternatively, the other atoms line numbers within the molecule specification section may be used for the values of variables, where the charge and spin multiplicity line is line 0).

15

The position of the current atom is then specified by giving the length of the bond joining it to atom1, the angle formed by this bond and the bond joining *atom*1 and *atom*2, and the dihedral (torsion) angle formed by the bond joining *atom*2 and *atom*3 with the plane containing the current atom, *atom*1 and *atom*2.

Here are two molecule specification sections for ethane:

```
0    1
C    0.00    0.00    0.00
C    0.00    0.00    1.52
H    1.02    0.00   -0.39
H   -0.51   -0.88   -0.39
H   -0.51    0.88   -0.39
H   -1.02    0.00    1.92
H    0.51   -0.88    1.92
H    0.51    0.88    1.92


0,1
C1
C2,C1,1.5
H3,C1,1.1,C2,111.2
H4,C1,1.1,C2,111.2,H3,120.
H5,C1,1.1,C2,111.2,H3,-120.
H6,C2,1.1,C1,111.2,H3,180.
H7,C2,1.1,C1,111.2,H6,120.
H8,C2,1.1,C1,111.2,H6,-120.
```

The version on the first uses Cartesian coordinates while the one on the second represents a sample Z-matrix (illustrating element labels). Note that the first three atoms within the Z-matrix do not use the full number of parameters; only at the fourth atom are there enough previously-defined atoms for all of the parameters to be specified.

Here is another Z-matrix form for this same molecule:

```
0    1
C1
```

```
C2   C1   RCC
H3   C1   RCH   C2   ACCH
H4   C1   RCH   C2   ACCH   H3   120.
H5   C1   RCH   C2   ACCH   H3  -120.
H6   C2   RCH   C1   ACCH   H3   180.
H7   C2   RCH   C1   ACCH   H6   120.
H8   C2   RCH   C1   ACCH   H6  -120.
     Variables:
RCH = 1.5
RCC = 1.1
ACCH = 111.2
```

In this Z-matrix, the literal bond lengths and angle values have been replaced with variables. The values of the variables are given in a separate section following the specification of the final atom. Variable definitions are separated from the atom position definitions by a blank line or a line like the following:

Variables:

More keywords can be found at
http://www.gaussian.com/g_ur/keywords.htm
(**www.gaussian.com** → **Tech Support** → **Gaussian 03 online manual** → **Gaussian 03 keywords).**