Quantum Chemistry with GAMESS

Brett M. Bode

Scalable Computing Laboratory

Department of Electrical and Computer Engineering

Iowa State University

AMES LABORATORY

IOWA STATE UNIVERSITY

Outline

- Introduction to GAMESS
- GAMESS history
- GAMESS capabilities
- Novel capabilities
- Running GAMESS

GAMESS

- General Atomic and Molecular Electronic Structure System
- General purpose electronic structure code
- Primary focus is on *ab initio* quantum chemistry calculations
- Also can do
  - Density functional calculations
  - Other semi-empirical calculations (AM1, PM3)
  - QM/MM calculations
- Its free and in wide use on everything from laptops to supercomputers.

Obtaining GAMESS

- Its free, but not “Open source” in the normal sense.
- Group license: You get the source and can do anything your want with it, except distribute it.
- See http://wwwmsg.fi.ameslab.gov/GAMESS/ for more information and the registration page link.
- Distribution is source code, with pre-built binaries also available for Macintosh, Linux and Windows.
- Full manual also on web site. See section 2 for complete keyword description for the input file, section 4 for references for all of the methods.

GAMESS People

- GAMESS is a product of Dr. Mark Gordon's research group at Iowa State University.
- Dr. Mike Schmidt coordinates the development efforts and is the gatekeeper for the code.

GAMESS History

- The code base was began in 1980 from parts of other codes. Some code still goes back to that version!
- Currently stands at about 750000 lines of mostly Fortran 77 compatible code.
- Pretty much runs on any system with a working Fortran compiler.

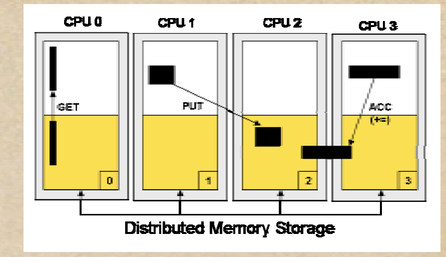
GAMESS Parallelization

- Began in 1991 with the parallelization of the SCF Energy and Gradient computations (almost trivially parallel).
- Initial parallel work done on the TouchStone Delta.
- In 1996 the Distributed Data Interface (DDI) was developed to support the new parallel MP2 energy and gradient code.

GAMESS Parallelization

- In 2004 DDI was rewritten and optimizations for SMP using SystemV shared memory were added. Focus remains distributed memory systems!
- Also added was subgroup support to enable the Fragment Molecular Orbital method.

DDI



- Modeled on the Global Array Framework.
- The Distributed Data Interface provides a pseudo global shared memory interface for a portion of a nodes memory.
- Normal MPI version uses 2 processes per processor, 1 compute, 1 data server.
    - Sockets are used for interrupts on data servers because MPI often polls in receive.
    - SHMEM and LAPI versions also available...
- Also provides processor subgroup support.

Program Capabilities

- Types of wavefunctions
  - Hartree-Fock (RHF, ROHF, UHF, GVB)
  - CASSCF
  - CI, MRCI
  - Coupled cluster methods (closed shells)
  - Second order perturbation theory
    - MP2 (closed shells)
    - ROMP2 (spin-correct open shells)
    - UMP2 (unrestricted open shells
    - MCQDPT(CASSCF - MRMP2)
  - Localized orbitals (SCF, MCSCF)

Program Capabilities

- Energy-related properties
  - Total energy as function of nuclear coordinates (PES): All wavefunction types
  - Analytic energy gradient
    - RHF, ROHF, UHF, MCSCF, CI, MP2, UMP2, DFT
    - ROMP2 in progress
  - Analytic hessian
    - RHF, ROHF, TCSCF/GVB
    - MCSCF just completed

Program Capabilities

- Energy-related properties (cont'd)
    - Numerical hessians from finite differences of analytic gradients
    - Fully numerical derivatives for all methods
    - Saddle point (TS) search (requires hessian)
    - Minimum energy path=Intrinsic Reaction Coordinate
        - Several IRC options - GS2 is most effective
        - Requires frequency input, gradients along path
        - Follow reaction path from reactants through TS to products
        - Build reaction path Hamiltonian (RPH): dynamics

Program Capabilities

- Energy-related properties (cont'd)
  - Dynamic reaction coordinate (DRC)
    - Add kinetic energy to system at any geometry
    - Add photon(s) to any vibrational mode
    - Classical trajectory using QM-derived energies
    - Requires gradients
  - Monte Carlo sampling: find global minimum
  - Molecular dynamics (in progress)

Program Capabilities

- Other functionalities
  - Spin-orbit coupling
    - Any spin states, any number of states
    - Full two-electron Breit-Pauli
    - Partial two-electron (P2e)-very efficient, accurate
    - Semi-empirical one-electron Zeff
    - RESC
    - Averaging over vibrational states
  - Derivative (vibronic) coupling: planned

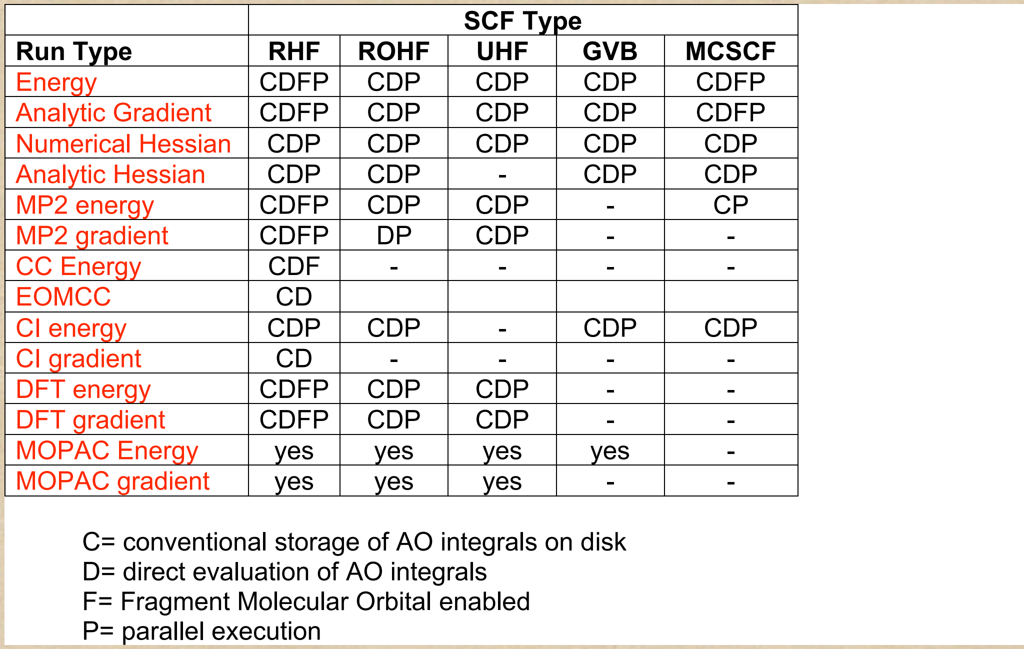
Program Capabilities

- Interpretive tools
  - Localized molecular orbitals (LMO)
  - Localized charge distributions (LCD)
- Nuclear and spectroscopic properties
  - Spin densities at nucleus (ESR)
  - NMR spin-spin couplings (in progress)
  - NMR chemical shifts
  - Polarizabilities, hyperpolarizabilities
  - IR and Raman intensities
  - Transition probabilities, Franck-Condon overlaps

Program Capabilities

- QM/MM Methods
  - Effective fragment potential (EFP) method for
    - Cluster studies of liquids
    - Cluster studies of solvent effects
    - Interfaced with continuum methods for study of liquids and solvation in bulk
    - Covalent link for study of enzymes, proteins, materials
  - SIMOMM: QM/MM method for surface chemistry
    - QM part can be any method in GAMESS
    - MM part from Tinker (Jay Ponder)

Current Capabilities



	SCF Type				
Run Type	**RHF**	**ROHF**	**UHF**	**GVB**	**MCSCF**
Energy	CDFP	CDP	CDP	CDP	CDFP
Analytic Gradient	CDFP	CDP	CDP	CDP	CDFP
Numerical Hessian	CDP	CDP	CDP	CDP	CDP
Analytic Hessian	CDP	CDP	-	CDP	CDP
MP2 energy	CDFP	CDP	CDP	-	CP
MP2 gradient	CDFP	DP	CDP	-	-
CC Energy	CDF	-	-	-	-
EOMCC	CD				
CI energy	CDP	CDP	-	CDP	CDP
CI gradient	CD	-	-	-	-
DFT energy	CDFP	CDP	CDP	-	-
DFT gradient	CDFP	CDP	CDP	-	-
MOPAC Energy	yes	yes	yes	yes	-
MOPAC gradient	yes	yes	yes	-	-

C= conventional storage of AO integrals on disk
D= direct evaluation of AO integrals
F= Fragment Molecular Orbital enabled
P= parallel execution

Solvation

- Solvation Methods
    - Explicit vs. implicit methods
- Explicit Methods
    - TIP3P, TIP4P
    - SPC, SPC/E
- EFP Method for Solvation
    - Summary of EFP1 method for water
- Generalized EFP Method (EFP2)

General Effective Fragment Potential

- Discrete solvation method
- Fragment potential is one electron contribution to the *ab initio* Hamiltonian
- Potentials
  - are obtained by separate *ab initio* calculations
  - depend on properties of isolated molecules
  - can be systematically improved

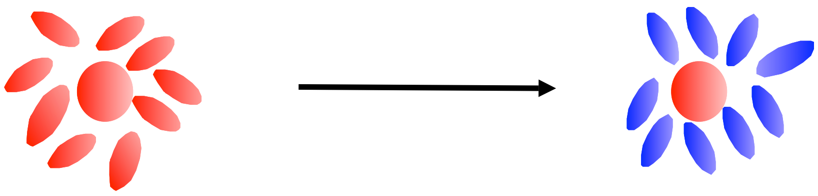
General Effective Fragment Potential

System is divided into

an *ab initio* region for the “solute” and

a fragment region for the solvent molecules.

$$E = E_{\textit{ab initio}} + E_{\text{interaction}}$$

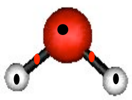


Hartree Fock based EFP

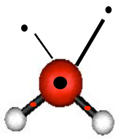
Interaction energy consists of : *electrostatic, polarization* and *exchange repulsion/charge transfer term*

$$E_{\text{interaction}} = E_{\text{coulomb}} + E_{\text{polarization}} + E_{\text{exchange repulsion/charge transfer}}$$

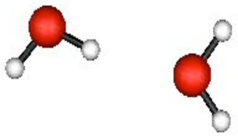
$$E_{\text{interaction}} = \sum_{k=1}^{K} V_k^{Elec}(\mu, s) + \sum_{l=1}^{L} V_l^{Pol}(\mu, s) + \sum_{m=1}^{M} V_m^{\mathrm{Re}\,p}(\mu, s)$$



Distributed Multipolar expansion



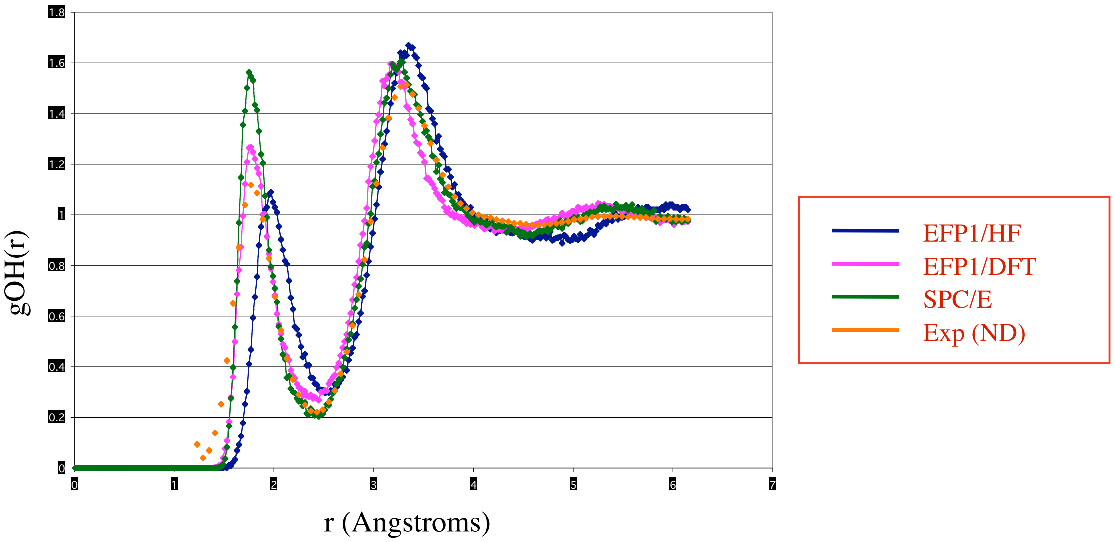
LMO polarizability expansion



Fit to Functional Form

EFP results

gOH(r):
EFP1/HF, EFP1/DFT, SPC/E
62 waters

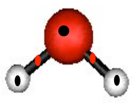


Exp (ND): Neutron Diffraction; Soper et. al.

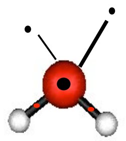
Generalized EFP2 Method

Interaction energy consists of : *electrostatic*, *polarization* and *exchange repulsion term*

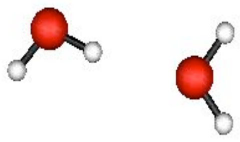
$$E_{\text{interaction}} = E_{\text{electostatic}} + E_{\text{polarization}} + E_{\text{exchange repulsion}}$$



Distributed Multipolar expansion

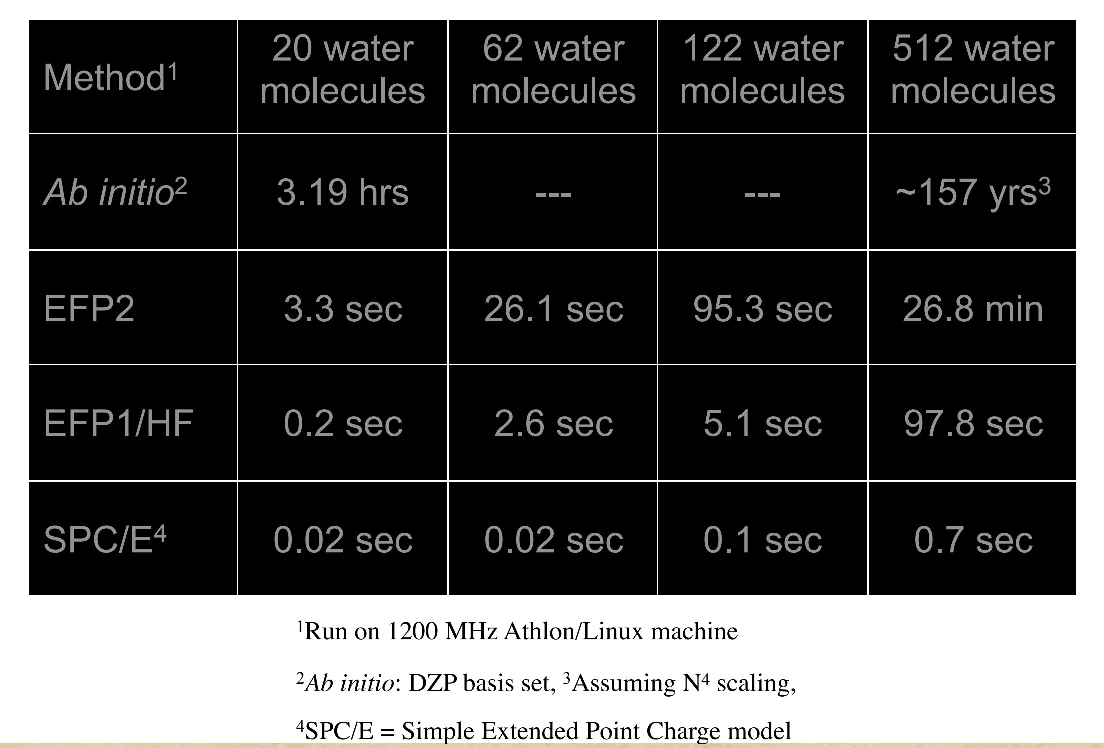


LMO polarizability expansion



From first principles using LMO overlaps

EFP Performance
Energy + Gradient Calculation



Method[1]	20 water molecules	62 water molecules	122 water molecules	512 water molecules
Ab initio[2]	3.19 hrs	---	---	~157 yrs[3]
EFP2	3.3 sec	26.1 sec	95.3 sec	26.8 min
EFP1/HF	0.2 sec	2.6 sec	5.1 sec	97.8 sec
SPC/E[4]	0.02 sec	0.02 sec	0.1 sec	0.7 sec

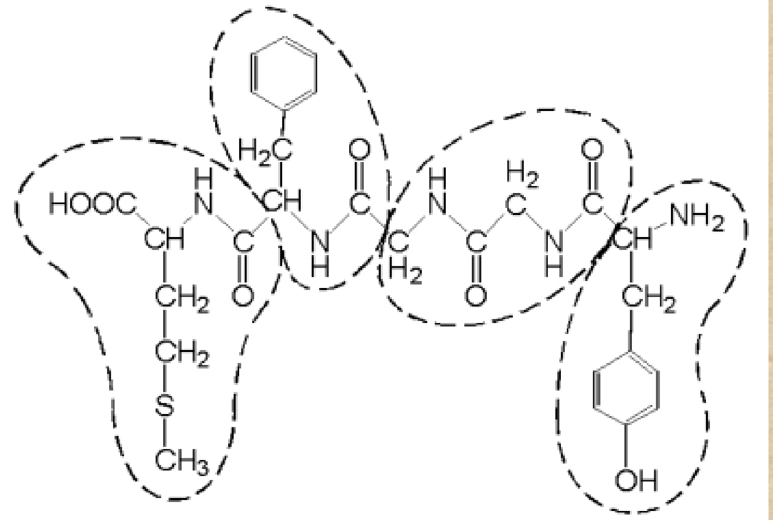
[1] Run on 1200 MHz Athlon/Linux machine

[2] *Ab initio*: DZP basis set, [3] Assuming N^4 scaling,

[4] SPC/E = Simple Extended Point Charge model

Fragment Molecular Orbital

- Divide up the system into fragments
  - Ignore exchange and self-consistency due to other fragments
  - Do *ab initio* calculations of fragments in the Coulomb field due to the whole system.
  - Likewise, compute pairs and triples of fragments.



work by Kitaura, Ishida and Federov at AIST

FMO Features

- No hydrogen caps.
- All n-mer calculations are *ab initio*.
- Interfragment charge transfer, dispersion and exchange are included.
- Systematic many-body effects.
- Total properties closely reproduce *ab initio* values.
- No fitted parameters.

FMO

- Can also add in electron correlation.
  - MP2
  - Coupled Cluster
  - DFT
  - MCSCF
- Can be multilayer - ie MCSCF for active site, RHF everywhere else.

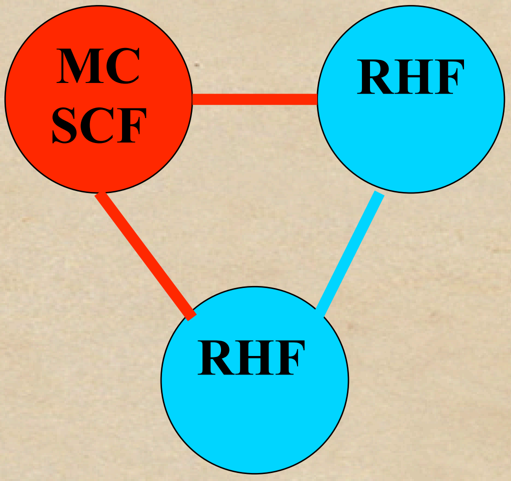
FMO-MCSCF

One monomer is defined as MCSCF.

Other monomers are RHF.

Dimers including the MCSCF monomer are MCSCF (red lines).

Other dimers are RHF (blue lines).



Applications of FMO

Geometry optimisation (≃2,000 atoms).

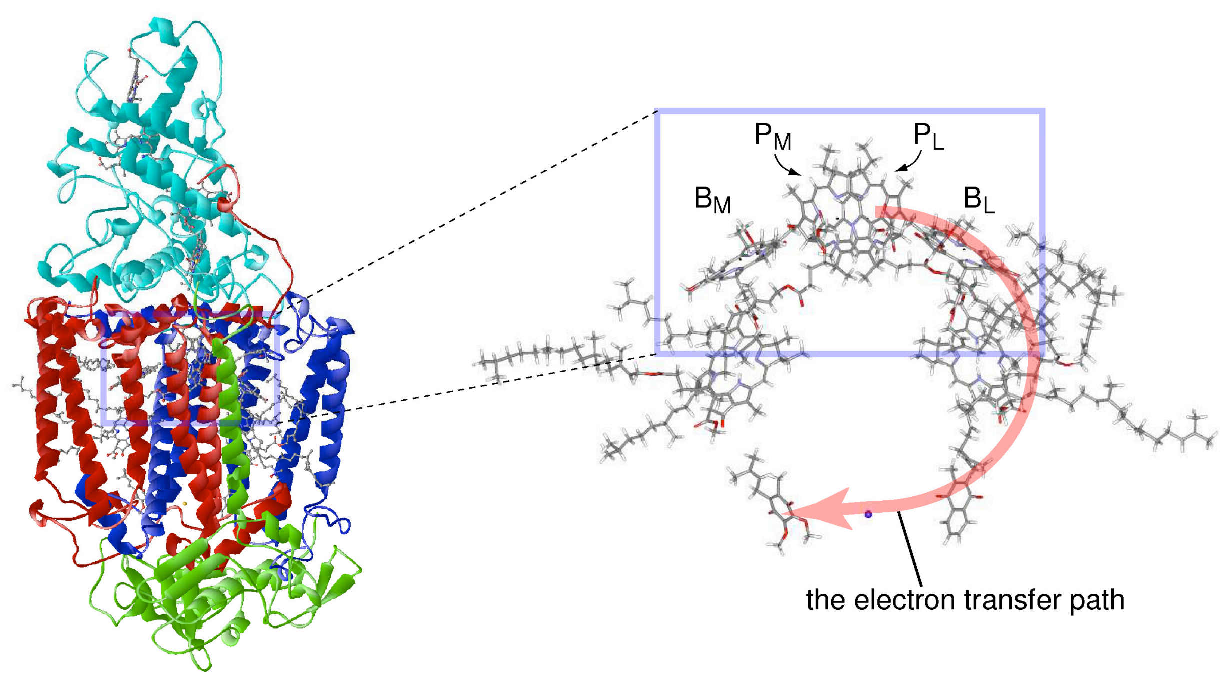
Ab initio level **single point energies** (≃20,000 atoms).

Pair interaction analysis (≃20,000 atoms):

drug design,
ligand docking,
polymer chemistry,
molecular clusters.

FMO results

Quantum chemistry *e silico in vivum*



FMO-RHF/6-31G*, 72.5 hours on 600 Opterons
20,581 atoms, 77,754 electrons, 164,442 basis functions.

T. Ikegami *et al.*, Supercomputing 2005 (Best technical paper award).

Running GAMESS

- GAMESS runs on
  - Any Unix-based system available in the U.S.
  - Any Linux based system
  - Any Macintosh
  - Windows based system using WinGAMESS or PCGAMESS
- Obtained from www.msg.ameslab.gov

GAMESS

- GAMESS is a back-end program, ie no GUI.
- Typically it is run via a script
- Input is taken from a file (usually .inp)
- Output appears in .log file (stdout)
    - This is intended to be human readable
- MO Vectors, coordinates, hessians, etc appear in .dat file. Can be used for restarts.
- IRC and DRP data and numerical hessian restart information appear in .irc file.
- These are all ASCII text files.

GAMESS Input file

- Input files are modular, arranged in $groups
- Most common input groups
  - $SYSTEM: specifies memory, time limit
  - $CONTRL: specifies basics of calculation
  - $BASIS: specifies basis set if standard
  - $DATA: specifies nuclear coordinates, basis set if non-standard
- Other important groups:
  - $GUESS, $SCF, $FORCE, $HESS, $VEC, $IRC, $VIB

GAMESS Input file

- The input file is mostly free-format (ie flexible spacing) except:
  - '$' sign specifying group must be in column 2!
  - All groups must terminate with a $END (this '$' can be anywhere except column 1).
  - anything in column 1 indicates a comment line

Some key groups

$SYSTEM group:

- TIMLIM=(default=525600 min = 1 yr)
- MWORDS=(default=1=8MB)
- MEMDDI=
  - relevant for parallel run
  - Total required memory (divide by number of processors to get memory requested/node)

Some key groups

$CONTRL

- ICHARG= (specifies charge on system)
- MULT= (specifies spin multiplcity)
    - 1 for singlet, 2 for doublet, ...
- EXETYP=
    - Check: checks input for errors
    - Run: actual run
- UNITS=
    - angs (default)
    - bohr

Some key groups

$CONTRL

- Runtyp= (type of run)
- Energy (single point energy run)
- Gradient (energy 1st derivative wrt coordinates)
- Optimize (optimize geometry)
- Hessian (energy second derivative, vibrational frequencies, thermodynamic properties):
  generates $HESS group in .dat file)
- Sadpoint (saddle point search:requires hessian in $HESS group)
- IRC (performs IRC calculation: usually requires $IRC group, $HESS group)

Some key groups

$CONTRL

- scftyp= (type of wavefunction)
  - RHF
  - ROHF
  - UHF
  - MCSCF
  - GVB
- mplevl=
  - 0 (default, no perturbation theory)
  - 2 (MP2: valid for RHF, ROHF, MCSCF)

Some key groups

$CONTRL

- Cctyp=
  - NONE (no coupled cluster, default)
  - LCCD (linearized doubles CC)
  - CCD (doubles CC)
  - CCSD (singles+doubles)
  - CCSD(T) adds perturbative triples to CCSD
    - Most popular method
    - Triples essential for accurate calculations
  - R-CC, CR-CC
    - Specialized methods to approximate bond-breaking

Some key groups

$BASIS - Used to select among the built-in basis sets

- GBASIS=
  - STO
  - N21
  - N31
  - TZV...
- NGAUSS=(# gaussians for STO, N21, N31)
- NDFUNC=(# sets of d's on heavy atoms
- NPFUNC=(# sets of p's on hydrogens)
- NFFUNC=(# sets of f's on TM's)
- DIFFSP=.T. (diffuse sp functions on heavy atoms)
- DIFFS=.T. (diffuse s functions on hydrogens)

Some key groups

$Data - Gives the molecular geometry

- Title line (will be printed in output)
- Symmetry group
  - C1
  - CS
  - CNV 2 (C2V), ...
- Blank line except C1
- Symbol Z xcoord ycoord zcoord
  - Symbol = atomic symbol
  - Z = atomic number
  - xcoord,ycoord, zcoord = Cartesian coordinates
  - Internal coordinates is another option

Some key groups

$Data - continued

- Repeat this line for each symmetry unique atom (see below)
- Need to specify basis set after each coordinate line if $BASIS is not present
- symmetry unique atoms
  - H2O: O and 1 H
  - NH3: N and 1 H
- saves CPU time (e.g., numerical hessians only displace symmetry unique atoms)
- Need to follow conventions in GAMESS manual
- Cs, Cnh: plane is XY
- Cnv: axis is Z
- For Cinfv, use C4v
- For Dinfh, use D4h

Some key groups

$GUESS - Initial MO guess

- Built-in guess (default) works much of the time
- $GUESS=MOREAD NORB=xx $END
  - Requires $VEC group (usually from .dat file)
  - NORB=# MO's to be read in
  - Useful when SCF convergence is difficult
  - Necessary for MCSCF, CI

GAMESS output

The log file output is intended to be human readable:

```
     ----------------------------
         RHF SCF CALCULATION
     ----------------------------


  NUCLEAR ENERGY =       8.9064898741
  MAXIT =  30    NPUNCH=   2
  EXTRAP=T  DAMP=F  SHIFT=F  RSTRCT=F  DIIS=F  DEM=F  SOSCF=F
  DENSITY MATRIX CONV=  1.00E-05
  MEMORY REQUIRED FOR RHF STEP=     30441 WORDS.

ITER EX DEM     TOTAL ENERGY       E CHANGE  DENSITY CHANGE    DIIS ERROR
 1  0  0    -74.7936151096  -74.7936151096   .595010038   .000000000
 2  1  0    -74.9519661838    -.1583510742   .180249713   .000000000
...
11  6  0    -74.9659012167    -.0000000014   .000018538   .000000000
12  7  0    -74.9659012170    -.0000000003   .000008228   .000000000
13  8  0    -74.9659012171    -.0000000001   .000003650   .000000000

     ------------------
     DENSITY CONVERGED
     ------------------
  TIME TO FORM FOCK OPERATORS=       .0 SECONDS (       .0 SEC/ITER)
  TIME TO SOLVE SCF EQUATIONS=       .0 SECONDS (       .0 SEC/ITER)

FINAL RHF ENERGY IS     -74.9659012171 AFTER  13 ITERATIONS
```

The Dat file

- The dat file contains formatted numerical data.
- Useful, sometimes required for restarts.
- Contains items such as:
  - MO Vectors ($VEC)
  - Gradient ($GRAD) and Hessian ($HESS)
- When copying a group make sure you copy everything from the beginning $ sign through the corresponding $END.

GAMESS output

- You will need to look at the log file to verify the results.
  - Did the run finish correctly?
  - Was the input specified correctly?
  - Were there errors in the computation?

Running GAMESS

- You frequently need the results from one run as input to another run.
  - restarting incomplete runs
  - Multi step problems
    - A Saddle point search might take several optimization and hessian computations followed by IRC computations.
  - Multi-reference computations often multiple runs to get the orbital guess correct.

Visualization

- A number of programs can visualize GAMESS results to varying degrees.
- MacMolPlt is one such program that has been specifically designed for visualizing GAMESS output.

Demo

- This afternoon I will present a demo of running GAMESS and using MacMolPlt.

Acknowledgments

- Mark Gordon
- Dmitri Federov
- the rest of the Gordon group in Ames

Financial Support

- Air Force Office of Scientific Research
- National Science Foundation
- DoD CHSSI Software Development
- DOE SciDAC Program
- Ames Laboratory
- DoD HPC Grand Challenge Program