Computational Chemistry (Quantum Chemistry) Primer

This is a very succinct primer intended as supplementary material for an undergraduate course in physical chemistry.

<u>TABLE OF CONTENTS</u>

<u>METHODS</u>
Molecular mechanics

- Based on classical mechanics, hence not really a quantum chemistry method
- Uses a lot of empirical (i.e., experimental) data
- Used especially for macromolecules
- Can determine energies, conformations, vibrational frequencies
- Cannot determine molecular orbitals
- Accuracy level 1 (scale of 1-4)

Semi-empirical

- Variational (calculated energy is upper bound for true ground state energy)
- Based on Hartree-Fock method
- Replaces some integrals with empirical data
- Accuracy level 2

Hartree-Fock (HF)

- Ab initio method ("from first principles" – uses no empirical data)
- Variational
- A self-consistent field (SCF) method
- Omits much of the electron correlation
- Accuracy level 3

2nd order Moeller-Plesset (MP2)

- Ab initio method
- Variational
- Correlated method (includes some electron correlation omitted by Hartree-Fock)
- Begins with HF calculation and then corrects using perturbation theory
- Can extended for greater accuracy (MP3, MP4, …)
- There are many other correlated methods, e.g., CCSD, CI, …
- Accuracy level 4

Density Functional Theory (DFT)

- Not variational
- Includes some extra electron correlation (beyond HF) implicitly

- Based on the electron density, not the wavefunction
- A "functional" is a function of a function. DFT methods express the energy as a function of the electron density (ρ), which in turn is a function of position: $E(\rho(r))$
- Many different functionals are available. The most widely used is B3LYP, but others may be more accurate for specific cases.
- Uses small amount of empirical data, hence not strictly ab initio (but sometimes called that)
- Generally faster than MP2 for similar accuracy. But remember than MP2 energies are variational and can be systematically improved with MP3, MP4 etc.
- Accuracy level 4

BASIS SETS

Basis sets come in different sizes. Generally, small basis sets result in faster calculations but lower accuracy. The appropriate basis set needs to be determined on the basis of the accuracy needed and the computational resources available. There are several kinds of nomenclature used for basis sets. Below we describe one of the oldest and most widely used, the Pople basis sets (named for John Pople, a key developer and a Nobel prize winner). Most other nomenclatures are based on the same principles for building the basis set.

Pople basis sets are most easily understood by beginning with hydrogenic orbitals, and then replacing them with approximations in two steps. The principle is that that hydrogenic atomic orbitals constitute a good basis set, but are not computationally efficient to implement. The spatial part of the wavefunction (i.e., an MO) can be expanded as a linear combination of atomic orbitals (AOs), i.e, as an LCAO-MO:

$$\psi_{\mathrm{MO}} = \sum_i \mathrm{AO}_i \qquad \text{[the set of atomic orbital functions } \{\mathrm{AO}_i\} \text{ is a complete basis set]}$$

There is nothing magical about AOs. Any complete basis set could be used. There are two principles for choosing a good basis set:
1. The expression for ψ_{MO} converges to the needed accuracy after just a few terms
2. Computers can carry out arithmetic operations with the basis functions efficiently (i.e., fast)

The AOs satisfy criterion (1) but not criterion (2). The reason that they satisfy criterion (1) is simple. Although AOs do not solve the Schrödinger equation for a molecule (which has many nuclei and many electrons), AOs do satisfy the Schrödinger equation for a related, simpler problem (one nucleus and one electron). As a reminder, here is the equation for an AO:

$$\mathrm{AO}_{n,l,m_l} = R_{n,l}(r)Y_{l,m_l}(\theta,\phi) = Nr^l e^{-\left(\frac{Z}{n}\right)r} L_{n-l-1}^{2l+1} Y_{l,m_l}(\theta,\phi)$$

We have written r instead of ρ for transparency; some factors are incorporated into N, and it is assumed that r is expressed in units of the Bohr radius.

The first step toward modern basis sets was to replace AOs with other functions similar to them but simpler. This was proposed by Slater, and so the functions are called Slater type orbitals (STOs):

$$\text{STO}_{n,l,m_l} = N r^{n-1} e^{-\zeta r} Y_{l,m_l}(\theta, \phi)$$

Note that the STO does not include a Laguerre polynomial, so it does not have radial nodes. Also, ζ is not necessary equal to Z/n.

The second step toward modern basis functions was to replace STOs with Gaussian-type functions (or sums of Gaussian-type functions). A Gaussian function has the form:

$$e^{-\alpha r^2}$$

where α is a number. The advantage of Gaussian functions is that they are computationally efficient for quantum chemistry calculations. The calculations can take a long time, and the bottleneck is often the calculation of many integrals. The most time consuming of these are so-called "4-center" integrals. They are integrals over four basis functions, each centered on a different atom. Consequently, the integral is over 12 coordinates, which makes it very time consuming. Gaussians, it so happens, have a very useful property: the product of two Gaussians, each centered on a different nucleus, is a different Gaussian centered at the midpoint of the first two. Consequently, the number of coordinates over which integration must be done is reduced, and the calculations are faster. This property is called the "Gaussian Product Theorem."

Therefore, STOs were replaced with Gaussian type orbitals (GTOs) having the form:

$$\text{GTO}_{n,l,m_l} = N r^{n-1} e^{-\alpha r^2} Y_{l,m_l}(\theta, \phi)$$

It was found that one cannot approximate an STO well with a single GTO, but if one uses a sum of GTOs, each having a different value of α, a good fit can be achieved. Even though the number of basis functions needed is increased by using sums of GTOs, the calculations are much faster than using STOs.
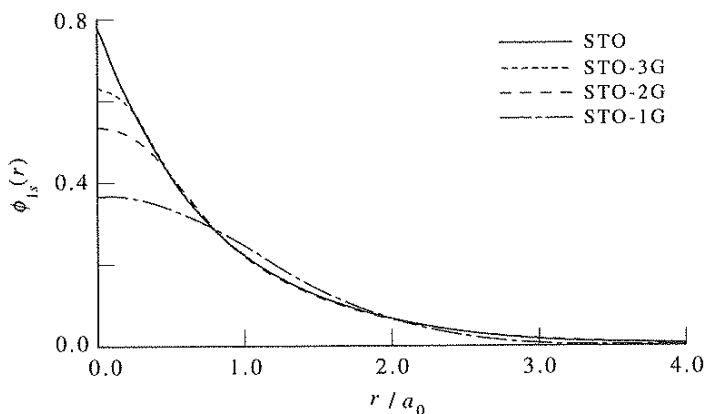
FIGURE 11.2
The Slater ϕ_{1s}^{STO} orbital is compared with $\phi_{1s}(r)$ represented by sums of different numbers of Gaussian functions.

From: *Physical Chemistry, A Molecular Approach*, D. McQuarrie and J. Simon, p. 415 (University Science Books, ISBN 0-93507-99-7).

The final major point that must be made is of a chemical nature. Consider the molecule HCN. It has one σ orbital, made primarily from $2p_z$ orbitals on the C and the N, and two π orbitals, each made primarily from $2p_x$ or $2p_y$ orbitals on the C and the N. It is known that π orbitals are more diffuse than σ orbitals; that is, they extend farther in space from the nuclei. That means that the $2p_x$ and $2p_y$ functions needed to represent the π orbitals well are, ideally, larger than the $2p_z$ functions needed to represent the σ orbital.

How do we deal with this when constructing a basis set? We cannot use different types of functions for the different 2p orbitals, because the basis set has to be used for many different molecules, not just HCN. The decision was a compromise. Choosing the $2p_x$ as an example, instead of representing the $2p_x$ with one STO, the $2p_x$ is represented with 2 STOs, each having a different value of ζ:

$$AO_{2p_x} \rightarrow d_1 STO(\zeta_1) + d_2 STO(\zeta_2)$$

Note that a large value of ζ leads to a compact STO, and a small value of ζ leads to a diffuse (more spread out) STO. The d's are just coefficients that get adjusted during the calculation. The STOs are not used directly, of course. Each one is expanded as a sum of GTOs:

$$STO(\zeta_j) = \sum_i GTO_i(\alpha_i)$$

The extra basis functions make the calculation a little slower, of course, but are needed for accuracy.

Now we are ready to put it all together. Core orbitals generally need less flexibility than valence orbitals. That is because core orbitals are held near the nucleus and are not much affected by the

nuclei of other atoms. Therefore, for core orbitals, usually just one STO is used for each AO. For the valence orbitals, though, several STOs are generally used for each AO. (The STOs are, of course, replaced with sums of GTOs.) There is special nomenclature. When different numbers of STOs are used for core and valence AOs, we speak of a "split-valence" basis set. In particular, when a sum of two STOs is used for the valence AOs, we speak of a "double-zeta" basis set (or to be completely specific, a "double-zeta, split-valence" basis set, but nobody says that).

Consider, for example, the 3-21G basis set. The nomenclature means:
- 1 STO for core orbitals (the STO is sum of 3 GTOs)
- 2 STOs for valence orbitals (one STO is a sum of 2 GTOs, the other just 1 GTO)
- this is both a split-valence and a double-zeta basis set
- the G just stands for "Gaussians"

Self-Test 1: describe what a 6-31G basis set means (answer at end of handout)

An exception to the nomenclature is STO-3G. This is the simplest basis set used, and frankly, it is rarely used except for qualitative calculations. The STO-3G uses a sum of three GTOs for every STO. The STO-3G is often referred to as a "minimal basis set."

Nothing is ever simple, and so there is one more set of things to mention. You will see an asterisk (*) or plus sign (+) in some basis sets. They refer to extra additions to the basis set. The nomenclature is:

        * "polarization function"
        + "diffuse function"

These basis functions are present for specific, chemical reasons. A polarization function is a GTO of higher angular momentum than STOs (each a sum of GTOs) in the basis set. Polarization functions allow for distortions needed by some molecules. A single * means that 2nd row elements will receive one set of 3d GTOs in addition to the 2p functions present in the basis set (3rd row elements would receive one set of 4f GTOs, etc.). A second * means that hydrogens will receive one set of 2p GTOs in addition to the 1s function. A diffuse function is commonly used with anions. The negative charge results in one electron being in a very diffuse orbital that extends far into space. The diffuse function is an extra s-type GTO with a very small exponent α. A single plus (+) means diffuse functions will be placed on all atoms except hydrogens. A double plus (++) means hydrogens will also receive diffuse functions.

Self-Test 2: describe what a 6-31G* basis set means (answer at end of handout)

Selt-Test 3: describe what a 6-311+G** basis set means (..)

<u>TYPES OF CALCULATIONS</u>
There are three basic types of calculations you are likely to do. The names used by Spartan are not the standard names, so here we give both:

- <u>Single-point energy</u> (Spartan: "energy"): The energy is calculated for a specific nuclear geometry. The nuclear coordinates are not changed.
- <u>Geometry optimization</u> (sometimes called "energy minimization") (Spartan: "equilibrium geometry"): Nuclei are moved until the energy is at a minimum. This is really a cyclical process. The energy is calculated for one nuclear geometry; the nuclei are moved; the energy is recalculated; and so on until an energy minimum is reached.
- <u>Transition state optimization</u> (Spartan: "transition state"): Nuclei are moved until a transition state is reached. This type of calculation is very sensitive to the starting geometry and may fail if the starting geometry is not near that of the transition state. Thus it requires a good initial guess for the transition state geometry.

<u>COMPLETE SPECIFICATION OF COMPUTATIONAL APPROACH</u>
The method and basis set are usually combined into one name, as "method"/"basis set". This name should be given whenever you report a calculation. Examples:

HF/3-21G   B3LYP/6-311+G**   MP2/6-31G*

Sometimes, one does a geometry optimization at a lower level of theory, followed by a single-point energy calculated at a higher level of theory. This is written using a double forward-slash, as "single-point"//"geometry optimization". Example:

MP2/6-311+G*//HF/6-31G   (geometry optimization is at HF/6-21G)

WHAT YOU ARE RESPONSIBLE FOR KNOWING (e.g., for exams)
Names of all the methods
Relative accuracy of methods
The meaning of "semi-empirical"
What a self-consistent field (SCF) calculation is (covered in class and in the text, not here)
Which methods are variational
Which methods are ab initio
Which methods include electron correlation

What a Gaussian is ($e^{-\alpha r^2}$)
What STO and GTO stand for (Slater type orbital, Gaussian type orbital)
The fact that an STO is like an AO but does not have nodes
The fact that STOs are replaced by sums of GTOs in quantum chemistry programs
Why STOs are replaced by GTOs (GTOs are more computationally efficient)
The meaning of:
- split-valence
- double-zeta
- diffuse function
- polarization function

How to interpret a Pople-type basis set (e.g., 6-31G)

The three basic types of calculations (both standard and Spartan nomenclature)
How to specify the entire computation approach (e.g., HF/6-31G)

<u>SELF-TEST ANSWERS</u>
1. meaning of 6-31G
- one STO for core orbitals (STO is sum of 6 GTOs)
- two STOs for valence orbitals (one STO is sum of 3 GTOs, other STO consists of 1 GTO)
- this both a split-valence and a double-zeta basis set

2. meaning of 6-31G*
- like (1) plus one set of polarization functions on atoms from 2nd row and higher

3. meaning of 6-311+G**
- one STO for core orbitals (STO is sum of 6 GTOs)
- three STOS for valence orbitals (one STO is sum of 3 GTOs, the other two STOs consist of 1 GTO each)
- polarization functions are added to all atoms, including hydrogen
- diffuse functions are added to atoms from 2nd row and higher
- this is both a split-valence and a triple-zeta basis set