

Handout 1: Introduction to Molecular Integrals

1.1 Molecular Integrals and the HF Method

At the core of any molecular modeling computation based upon quantum mechanical methods (as opposed to classical mechanical methods such as those generally termed “molecular mechanics”) are the *molecular integrals*.

In computational molecular modeling using quantum chemical methods, the chemist seeks to determine the energy of a molecular system associated with some geometry, but without employing any empirical parameters. Quantum chemistry is an *ab initio* method, whereby energies of atomic and molecular systems are computed directly from first principles, rather than depending on any experimental measurements. In a sense this requires asking the computer to solve Schrödinger’s differential equation,

$$\hat{H}\Psi = E\Psi. \quad (1)$$

The wave function Ψ is difficult to define in tangible terms. Suffice it to say that it contains information about the system we are interested in, and that it has a mathematical form. This system could be a free electron, an atom, a molecule, or something hypothetical like a particle in an energy well. Physically meaningful information *is* extracted from the wave function when an operator is applied to it. The most important operator of quantum mechanics is the Hamiltonian \hat{H} ,

$$\hat{H} = \hat{T} + \hat{V}, \quad (2)$$

where \hat{T} is kinetic energy and \hat{V} is potential energy. When the wave function Ψ is subjected to the Hamiltonian, as in the Schrödinger equation (Eq. 1), the values of energy E are returned as multiplicative constants of an unaltered Ψ .

Whether one can *do* the complicated mathematics necessary to find the solutions E and Ψ is another matter. The undergraduate physical chemist learns integro-differential techniques for deriving the solutions – calculating the energies and wave functions – of a few systems, including the hypothetical particle in a box, the harmonic oscillator, and perhaps the hydrogen atom. In all three cases, solving Schrödinger’s equation is the *eigenvalue problem* of finding an *eigenfunction* (wave function) and *eigenvalue* (energy). These systems are three of only a handful for which mathematically exact, or *analytical*, solutions exist. Such analytically-soluble systems are precious few because of the many-body problem, whereby the exact solution of a system composed of more than two bodies is mathematically intractable. The particle in a box, whether in one dimension or three, is just one body. The hydrogen atom matches the many-body limit with one proton and one electron.

The many-body problem is particularly irksome to the chemist, who is interested in solving Schrödinger’s equation for systems composed of many bodies (nuclei and electrons) – not only the many bodies of atoms more complex than hydrogen, but of course those comprising molecules. Fortunately, certain approximations make the application of Schrödinger’s equation to chemically relevant many-body systems possible. The first of these is the Born–Oppenheimer approximation,¹ which professes that the nuclear motion of an atom or molecule can be decoupled from electronic motion, because nuclei are thousands of times heavier than electrons and move much slower. Hence, quantum chemistry comes down to solving the *electronic Schrödinger equation*,

$$\hat{H}_e\Psi_e = E_e\Psi_e, \quad (3)$$

which uses the molecule’s geometry as fixed parameters, but leaves out the nuclear operator terms of the Hamiltonian. The operator of the electronic Hamiltonian,

$$\hat{H}_e = \sum_i \underbrace{\left(-\frac{1}{2}\nabla_i^2 - \sum_C \frac{Z_C}{r_{iC}} \right)}_{h(i)} + \sum_{i < j} \underbrace{\left(\frac{1}{r_{ij}} \right)}_{v(i,j)}, \quad (4)$$

informs the structure of the molecular integrals. There are one-electron components, $h(i)$, as well as two-electron components, $v(i, j)$. The one-electron components (themselves operators) contain the *kinetic energy*,

which notably features the Laplace operator,

$$\nabla_i^2 = \frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2}. \quad (5)$$

Eq. 4 also has a term for *nuclear attraction*, where r_{iC} implies the distance between electron i and nucleus C . Z_C is the atomic number of nucleus C and is just a numerical constant. The two-electron component is the *electron repulsion* operator, and implies the distance between electrons i and j . Note that atomic units are employed throughout this paper, which substantially simplifies equations; see Table 1.

Physical quantity	Atomic units	SI units
Length	1 a_0	5.2918×10^{-11} m
Mass	1 m_e	9.1095×10^{-31} kg
Charge	1 e	1.6022×10^{-19} C
Energy	1 E_h (Hartree)	4.3598×10^{-18} J
Angular momentum	1 \hbar	1.0546×10^{-34} J s
Wave function	1 $a_0^{-3/2}$	2.5978×10^{15} m $^{-3/2}$

Table 1: The present paper employs atomic units (a.u.), the standard unit of quantum chemistry. Conversion to SI units is possible as shown.

Another important approximation in quantum chemistry is that of molecular orbital (MO) theory. This states that the wave function, Ψ , can be decomposed into a combination of molecular orbitals ψ_i :

$$\Psi = \psi_1 \psi_2 \dots \psi_n. \quad (6)$$

Note that the e subscript has been dropped. (In order to satisfy the principle of antisymmetry, this product of molecular orbitals is recast as a determinant, but that will not have direct consequence on the evaluation of molecular integrals, nor on implementing the HF method.) Each molecular orbital is constructed as a sum of weighted contribution from each atomic orbital:

$$\psi_i = \sum_{\mu=1}^N C_{\mu i} \phi_{\mu}. \quad (7)$$

In fact, Eqs. 6 and 7 are nothing more than the mathematical formulation of the *Linear Combination of Atomic Orbitals – Molecular Orbital* (LCAO–MO) theory, which may be introduced in freshman chemistry. The coefficient $C_{\mu i}$ is simply a number that indicates how much atomic orbital ϕ_{μ} contributes to molecular orbital ψ_i .

But how is $C_{\mu i}$ determined, and what does ϕ_{μ} look like? We address these questions by presenting the HF equations (more accurately, the HF–Roothan–Hall^{2,3} equations). Only the essential HF equations necessary for motivating the molecular integrals are presented; their full derivation is provided in the classic text by Szabo and Ostlund,⁴ and discussed in other *J. Chem. Ed.* articles.^{5,6} The central HF equation is

$$\mathbf{FC} = \epsilon \mathbf{SC}. \quad (8)$$

Each term is either a one- or two-dimensional matrix (see below) whose size is determined by the number of atomic orbitals present.

The HF equation, Eq. 8, is the computational chemist’s approximation to the Schrödinger equation. It even bears a similar form: The Hamiltonian operator \hat{H} has been substituted by \mathbf{F} , the “Fock operator” matrix. (The two-dimensional matrices of interest in this article, including \mathbf{F} , are also known as rank-2 tensors.) The wave function Ψ has been substituted by \mathbf{C} , which is the matrix (rank-2 tensor) of $C_{\mu i}$ coefficients. Each column of the matrix \mathbf{C} represents one molecular orbital; in the language of linear algebra, each column of \mathbf{C} is an *eigenvector*. Analogous to the eigenfunctions in a differential eigenvalue equation, eigenvectors *are* the molecular orbitals. More specifically, we say that molecular orbitals are eigenvectors “in the atomic orbital basis,” because of Eq. 7. The vector ϵ holds the eigenvalues (analogous to E), and so it

is a rank-1 tensor (or “one-dimensional matrix”) of energies. The first eigenvalue, or element of the vector ϵ , is the energy of the the first eigenvector (column) in \mathbf{C} ; the second eigenvalue is the energy of the second eigenvector in \mathbf{C} ; and so on.

Each eigenvector of \mathbf{C} is a linear combination of the atomic orbitals. This means that each element of this column vector is a number, the coefficient $C_{\mu i}$, indicating how much of a particular atomic orbital ϕ_μ contributes to the molecular orbital ψ_i .

One thing clearly distinguishing the Schrödinger equation (Eq. 1) from the HF equation (Eq. 8) is the presence of \mathbf{S} , which is a matrix of overlap integrals. Every element of this matrix quantifies to what extent one atomic orbital of the molecule overlaps with another atomic orbital; the maximal value is 1, because orbitals are normalized. The value of 1, indicating perfect orbital overlap, only occurs when describing the overlap of an orbital with itself. \mathbf{S} is the first of the molecular integral matrices we need to generate to solve the HF equation.

The Fock matrix \mathbf{F} is a sum of three two-dimensional matrices,

$$\mathbf{F} = \mathbf{T} + \mathbf{V} + \mathbf{P}, \quad (9)$$

or, element-by-element,

$$\begin{aligned} F_{\mu\nu} &= T_{\mu\nu} + V_{\mu\nu} + P_{\mu\nu} \\ &= T_{\mu\nu} + V_{\mu\nu} + \sum_{\lambda\sigma} D_{\lambda\sigma} \left(G_{\mu\nu\sigma\lambda} - \frac{1}{2} G_{\mu\lambda\sigma\nu} \right). \end{aligned} \quad (10)$$

This is where the remaining three molecular integrals appear. \mathbf{T} represents the kinetic energy integrals matrix; \mathbf{V} is the nuclear attraction integrals matrix. The rank-4 tensor \mathbf{G} (note the four subscript terms) makes up part of \mathbf{P} , and is the electron repulsion integrals rank-4 tensor. That \mathbf{G} is a rank-4 tensor simply means it is a “four-dimensional matrix.” All of the usual matrix algebra operations that apply to two-dimensional matrices, such as addition and taking dot products, also apply to rank-4 tensors such as \mathbf{G} ; these operations are all that we need. Figs. 7 and 8 of Handout 2 offer a way of visualizing a rank-4 tensor. (From now on, specific reference to rank-1 tensors will be termed vectors, rank-2 tensors will be termed matrices, and rank-4 tensors will simply be tensors; when considered together they will be the “molecular integral matrices,” in keeping with the literature.)

The density matrix \mathbf{D} , in \mathbf{P} of Eq. 10, is built from the molecular orbitals:

$$D_{\lambda\sigma} = \sum_i^{N/2} C_{\lambda i} C_{\sigma i}. \quad (11)$$

It is formed by multiplying together the lower $N \div 2$ molecular orbitals of the wave function, where N is the number of electrons, and each orbital holds two electrons. These lower $N \div 2$ orbitals are the *occupied* orbitals, whereas the upper $N \div 2$ orbitals are unoccupied or, in the parlance of quantum chemistry, *virtual*.

The four *molecular integral matrices* we are interested in generating, then, are:

1. \mathbf{S} : overlap integrals matrix
2. \mathbf{T} : kinetic energy integrals matrix
3. \mathbf{V} : electron-nuclear attraction integrals matrix
4. \mathbf{G} : electron-electron repulsion integrals tensor

Each integral matrix is formed only once, and remains unchanged during the HF procedure. By contrast, the molecular orbitals \mathbf{C} (and hence the density \mathbf{D}) change with each iteration, until there is no change between cycles. It is a successive process of energy minimization, each step tweaking the orbital coefficients until a set of molecular orbitals that gives the *lowest* total electronic energy for the system is produced.⁷

The determination of a molecular orbital ψ_i additionally yields its energy ϵ_i . If we have every ψ_i , then we have the wave function Ψ (Eq. 6). *The computational equivalent of the Schrödinger equation has been solved.*

The total electronic energy, or the HF energy, of the system does not come directly out of Eq. 8. But it is an *expectation value* that is easily evaluated, the equation for which appears in the HF programming project, Handout 5.

Of course, before we can actually execute the HF iterative procedure, we must populate the matrices with numbers. The question remains as to how to calculate the matrix *elements* of **S**, **T**, **V**, and **G**. Individual elements are where integrals are actually solved, which when solved yield numbers in the energy units Hartree. This is because each element is an expectation value. The form of the expectation value for a generic operator \hat{O} is

$$\langle O \rangle = \int \phi_A \hat{O} \phi_B d\tau. \quad (12)$$

Because we form molecular orbitals ψ_i out of atomic orbitals ϕ_μ , the operators act on ϕ_μ , as Eq. 12 shows. The linear combination of these expectation values over ϕ_μ is equal to having the expectation value of ψ_i .

Based on our discussion of the matrices necessary to solve the HF equation (Eq. 8), the operators required for calculating the matrix elements are kinetic energy, electron-nuclear attraction energy, and electron-electron repulsion energy. These are found in the electronic Hamiltonian, Eq. 4.

Consider the matrix element T_{AB} , which corresponds to the first one-electron operator of Eq. 4:

$$T_{AB} = \langle T \rangle = \int \phi_A \hat{T} \phi_B d\tau = \int \phi_A \left(-\frac{1}{2} \nabla^2 \right) \phi_B d\tau. \quad (13)$$

\hat{T} acts on orbital ϕ_B . The integrand is formed by multiplying the result of this operation by orbital ϕ_A ; integration of the integrand is over all space, $d\tau$. (A typical set of coordinates in physical chemistry are the spherical coordinates (r, θ, ϕ) , where $d\tau = r^2 \sin \theta d\theta d\phi dr$; as will be addressed in Handout 2, Cartesian coordinates are employed in the present treatment, such that $d\tau = dx dy dz$.) We will also use the notation $(A | -\frac{1}{2} \nabla^2 | B)$ to express this single matrix element, where A represents orbital ϕ_A , and B represents orbital ϕ_B .

The matrix elements V_{AB} of the electron-nuclear attraction integral matrix are

$$V_{AB} = \langle V \rangle = (A | \frac{-Z_C}{r_{iC}} | B) = \int \phi_A \hat{V} \phi_B d\tau = \int \phi_A \left(\frac{-Z_C}{r_{iC}} \right) \phi_B d\tau. \quad (14)$$

The operator features the atomic number of atom C in the numerator; r_{iC} indicates the distance between electron i and nucleus C .

For the matrix elements of **G** – perhaps “tensor elements” would be more apt – we have two electrons i and j :

$$G_{ABCD} = \langle G \rangle = (AB | CD) = \int \phi_A \phi_B \hat{G} \phi_C \phi_D d\tau = \int \phi_A \phi_B \left(\frac{1}{r_{ij}} \right) \phi_C \phi_D d\tau. \quad (15)$$

It is a peculiar convention that the notation $(AB | CD)$ does not show the operator, but it is implied.

Of course, matrix elements of the overlap integral matrix are needed. For these, there is no operator, and the integrand is simply the product of two orbitals,

$$S_{AB} = (A | B) = \int \phi_A \phi_B d\tau. \quad (16)$$

It is worth noting that the elements of matrices **S**, **T**, and **V** are *two-center, one-electron integrals* (centers \vec{A} and \vec{B} corresponding to orbitals ϕ_A and ϕ_B , for electron i), whereas elements of the tensor **G** are *four-center, two-electron integrals* (centers \vec{A} , \vec{B} , \vec{C} , and \vec{D} , for electrons i and j).

Before these integrals can be evaluated, we must consider the mathematical form the orbitals ϕ_μ take.

1.2 Atomic Orbitals and Gaussian Products

Molecular orbitals are linear combinations (weighted sums) of atomic orbitals. The expectation value for each operator in the electronic Hamiltonian acts on atomic orbitals. We say that atomic orbitals form our *basis*, and that each energy integral matrix is in the *atomic orbital basis*.

In addition to being useful qualitative devices, atomic orbitals (the “objects” that electrons occupy) are mathematical constructions. Their mathematical form, which cannot be exactly known but only approximated, is based on the assumption that the orbitals for any atom can be approximated by the *known* mathematical form of the orbitals of the hydrogen atom. (Erwin Schrödinger derived the equations of hydrogen’s orbitals. Due to the many-body problem, the mathematics is intractable for determining exact orbital

equations for any other atom.)

The American physicist John C. Slater suggested this approximation in 1930.⁸ Like the orbitals Schrödinger derived for the hydrogen atom, Slater-type orbitals (STO) have an angular and a radial component,

$$\phi = N \underbrace{x^l y^m z^n}_{\text{angular}} \underbrace{e^{-\alpha r}}_{\text{radial}}. \quad (17)$$

In Eq. 17, N is a normalization constant, and α is a number found in tables for basis sets unique to every atom and orbital (s, p, d , etc.). The variables x , y , and z are the three-dimensional Cartesian coordinates describing the position of the electron (the nucleus being at the origin), and the powers l , m , and n associated with each coordinate are integers describing the directional contributions to the electron's orbital angular momentum. The sum $l + m + n$ indicates the type of orbital to which the electron belongs: $l + m + n = 0$ is an s orbital, $l + m + n = 1$ is a p orbital, $l + m + n = 2$ is a d orbital, and so on. If $l + m + n = 1 + 0 + 0 = 1$, then the electron belongs to a p_x orbital; $l + m + n = 0 + 2 + 0 = 2$ is a d_{y^2} orbital; and so on. The term r is the distance of the electron from the center of the atom:

$$r = \sqrt{x^2 + y^2 + z^2}. \quad (18)$$

It turns out that Slater-type orbitals form effectively intractable (*i.e.*, not analytically solvable) integrals (Eqs. 13 through 16) for anything but atoms, some diatomics, and some linear polyatomic molecules.⁹ The two-electron integrals (Eq. 15) are particularly problematic; while approximate, numerical solutions involving expansions in series do exist, these are computationally very expensive, requiring many terms in a summation before adequate convergence is achieved.

In 1950, Samuel F. Boys proposed¹⁰ that Slater-type orbitals could be approximated by Gaussian functions:

$$\phi \approx \sum_{p=1}^{K_p} d_p N_p x^l y^m z^n e^{-\alpha_p r^2}. \quad (19)$$

Therein, p is a generic index substituted with a , b , c , or d when referring to orbitals ϕ_A , ϕ_B , ϕ_C , or ϕ_D , respectively. Notably, Eq. 19 is distinguished from Eq. 17 by the squaring of r . This has, perhaps surprisingly, enormous significance for the taking of derivatives and the calculation of integrals over multiple centers. This is because of the *Gaussian Product Theorem*: The product of two Gaussian functions on two different centers yields one Gaussian function located on a center *between* the original two Gaussians. As a simple demonstration, consider two s orbitals, ϕ_A and ϕ_B , each approximated by Eq. 19. Each orbital has $l = m = n = 0$. For the moment let's disregard the constants d_p and N_p . Then,

$$\begin{aligned} \phi_A \times \phi_B &= \exp(-\alpha_a r_A^2) \times \exp(-\alpha_b r_B^2) \\ &= \exp \left[- \left(\frac{\alpha_a \alpha_b}{\alpha_a + \alpha_b} \right) |\vec{A}\vec{B}|^2 \right] \times \exp [-(\alpha_a + \alpha_b) r_P^2] \\ &= \tilde{K} \times \exp [-(\alpha_a + \alpha_b) r_P^2]. \end{aligned} \quad (20)$$

Clearly, the product of two Gaussian functions yields a third, single Gaussian function where \tilde{K} is a constant (not a function of the variable r), $|\vec{A}\vec{B}|^2$ is the squared distance between \vec{A} and \vec{B} ,

$$|\vec{A}\vec{B}|^2 = (\vec{A}_x - \vec{B}_x)^2 + (\vec{A}_y - \vec{B}_y)^2 + (\vec{A}_z - \vec{B}_z)^2, \quad (21)$$

and r_P describes a position located between r_A and r_B . Two centers, A and B , are reduced to one, P ; see Figs. 1 and 2. It is not hard to imagine that an integral over a single center is simpler to evaluate than an integral over two centers. This vital property is especially valuable when analyzing the four-center, two-electron integrals of **G**: The four centers A, B, C and D are reduced to two centers, P and Q .

In addition to the substitution of an exponential radial function for a Gaussian radial function, Eq. 19 features summation and a new constant, d_p . To highlight their significance, consider the hydrogen $1s$ orbital approximated by the STO, versus that estimated with a Gaussian function. This Gaussian approximation is also known as a Slater-Type Orbital – 1 Gaussian (STO-1G). The nomenclature STO-NG is a little unfortunate, as it would seem to imply that these functions are Slater-type orbitals. But the terms after the

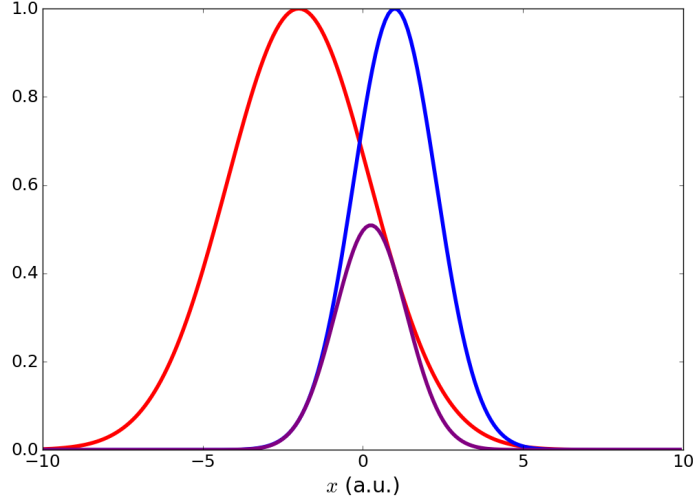


Figure 1: A demonstration of the Gaussian Product Theorem with two s -type orbitals approximated as Gaussian functions. Ignoring y - and z -components, the x -component of orbital ϕ_A (red) is centered at $x = -2.0$, and the x -component of orbital ϕ_B (blue) is centered at $x = 1.0$. The width of each individual orbital is determined by the exponential factor α . The thicker orbital ϕ_A has the smaller exponential factor, $\alpha_a = 0.10$; the thinner orbital ϕ_B has the larger exponential factor, $\alpha_b = 0.30$. The product of ϕ_A and ϕ_B is the purple line; in accordance with Eq. 20, it is located between the centers of ϕ_A and ϕ_B .

hyphen indicate that they are *approximations* to STOs by way of a linear combination of Gaussian functions; in the case of STO-1G, just one Gaussian function. (Recall that STO-1G is an approximation to the STO, which is an approximation to the *actual* orbital, the exact form of which we cannot know.)

$$\phi_{A,\text{STO}} = N_a x^l y^m z^n e^{-\alpha_a r_A} = \frac{1}{\sqrt{\pi}} \times x^0 y^0 z^0 \times e^{-1.0 r_A}$$

$$\phi_{A,\text{STO-1G}} = \sum_{a=1}^{K_a=1} d_a N_a x^l y^m z^n e^{-\alpha_a r_A^2} = 1.0 \times 0.26766 \times x^0 y^0 z^0 \times e^{-0.28294 r_A^2}$$

In the STO-1G basis of ϕ_A , a hydrogen $1s$ orbital, there is only one coefficient d_a and one exponential factor α_a ($K_a = 1$).¹¹ As Fig. 3 reveals, a single Gaussian function does a poor job of emulating a Slater-type orbital. It grossly underestimates the amplitude of the orbital at small r near the nucleus, overestimates it at intermediate distances between about 1 and $2.5 a_0$, and then underestimates it again at distances beyond $2.5 a_0$. The one d_a value (in this case, 1.0), the α_a constant in the exponent, and $K_a (= 1)$ make up the STO-1G *basis set* parameters for the $1s$ orbital of the hydrogen atom.

If we add a second Gaussian function to our approximation of STO, which would be the Slater-Type Orbital – 2 Gaussian Functions (STO-2G) basis set, the approximation is better. In this case, we have a linear combination of Gaussian functions. Each Gaussian function is called a *Gaussian primitive*, and is weighted by a coefficient. This weighting factor, d_p , is known as a *contraction coefficient*. This contraction of Gaussian primitives gives us the *contracted Gaussian-type orbital* (CGTO), also known as a *contracted Gaussian-type function* (CGTF). Thus, with a minimal basis set like STO-3G, Eq. 19 is known as a *contracted Gaussian-type orbital*. (In Handout 2, we will discuss why it is better to refer to Eq. 19 as a CGTF when using extended basis sets.) Applied to the hydrogen $1s$ orbital,

$$\begin{aligned} \phi_{A,\text{STO-2G}} &= \sum_{a=1}^{K_a=2} d_a N_a x^l y^m z^n e^{-\alpha_a r_A^2} \\ &= d_1 N_1 x^l y^m z^n e^{-\alpha_1 r_A^2} + d_2 N_2 x^l y^m z^n e^{-\alpha_2 r_A^2} \\ &= 0.67891 \times 0.17359 \times x^0 y^0 z^0 \times e^{-0.233136 r_A^2} + 0.43013 \times 0.63193 \times x^0 y^0 z^0 \times e^{-1.309756 r_A^2} . \end{aligned}$$

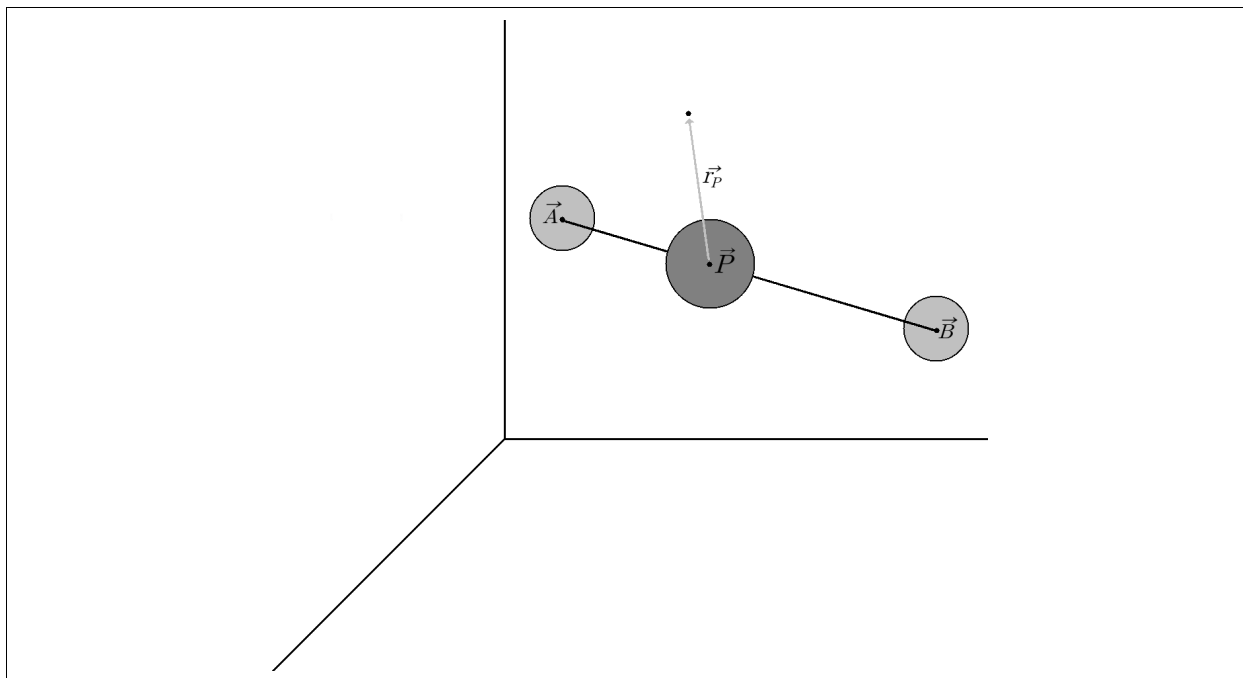


Figure 2: In taking the product of orbitals ϕ_A and ϕ_B , each represented by Cartesian Gaussian functions, a third center \vec{P} located between \vec{A} and \vec{B} is formed. \vec{r}_P is the distance between the electron and the third center \vec{P} .

Once again, the basis set parameters α_p and d_p are found from a table. Fig. 4 provides a visualization of this contraction of two Gaussian primitives, and shows that it is a better approximation to a true STO.

As we would expect, the contraction of three Gaussian primitives in the STO-3G basis set improves on the STO-2G approximation; a linear combination of six primitives, STO-6G, does a better job still (see Fig. 5). Generally, the more Gaussian functions that are included, the better the approximation to Slater-type orbitals. The price that is paid is computational expense. This is because an integral must be evaluated for each *primitive*, not just each orbital. In a STO-1G basis, only one integral is evaluated per matrix element. In a STO-3G basis, three integrals are evaluated and summed to form the final value of the matrix element. Higher quality comes at a price. Fig. 5 reveals that even STO-6G fails to capture Slater-type orbital behavior very near the nucleus – the *cusp condition*¹² – and underestimates the function at longer distances. Depending on the molecular system, increasing basis set size beyond a certain point yields diminishing returns. In the programming projects, a balance will be struck between expense and accuracy by employing the STO-3G basis set. (It is worth noting that the evaluation of integrals depending on several Gaussian primitives per orbital entails far fewer terms than would be required to evaluate integrals depending on a single exponential Slater-type orbital; see Hoggan *et al.*⁹ for further discussion.)

These basis sets and their tables are curated on the EMSL (Environmental Molecular Sciences Laboratory) Basis Set Exchange (<https://bse.pnl.gov/bse/portal>). An example of how to read these tables is shown in Fig. 6.

There are many excellent discussions of how to select basis sets in computations,^{13,14} and so these considerations are not addressed in this paper. The present focus is on the essential content of a basis set, such as the fact that they are composed of contraction coefficients d_p and exponent factors α_p , as well as their direct relation to the molecular integrals. In addition to the earlier discussion, further details on the basic structure of basis sets is provided in Handout 2.

1.3 Moving Forward

When a computer “calculates the molecular integrals,” it is not as if it is solving new calculus or differential equation problems with every computation. The solutions to the molecular integrals (Eqs. 13 – 16) have

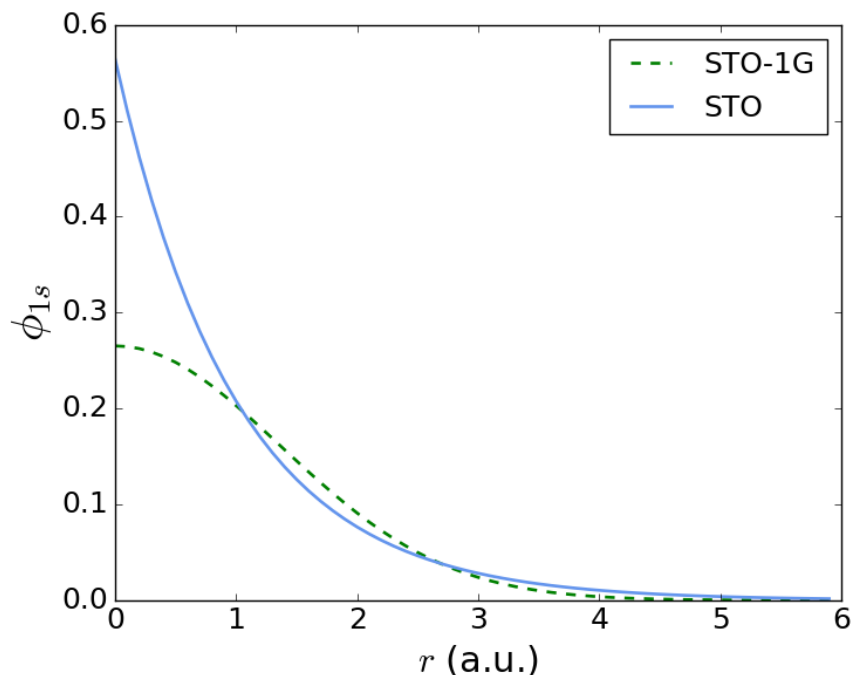


Figure 3: The approximation of a hydrogen 1s Slater-type orbital by a single Gaussian function with the STO-1G basis set. One Gaussian function (STO-1G) does a rather poor job of emulating an STO.

already been derived,¹⁵ just as the reader can anticipate the derivation in Handout 2. “Calculating the molecular integrals” is a matter of having the computer substitute numbers in to variables in the known solutions, which are implemented in computer code (as the reader will also implement them in Handout 4). These numbers include the basis set parameters (α_p and d_p), and the geometric information about the orbitals in the molecule. Because orbitals are typically centered on atoms, orbital coordinates are taken directly from the molecule coordinates `.xyz` file. Knowing what orbitals to use depends on both the input file and the basis set: if carbon is an atom of the input file, then there must be at least carbon 1s, 2s, and 2p orbitals present in the basis. Some basis sets will, however, include orbitals of higher principal number and angular momentum (3s, 3p, 3d, ...) whether they are occupied or not; doing so often improves the quality of the molecular orbitals **C** and accuracy of the HF energy.¹³ Because each of these orbital types (s, p, d, ...) has angular momentum associated with it, the corresponding integer values of l , m , and n are also (Eq. 19) “plugged in” to the solutions.

In Handout 2, which follows from the foregoing discussion, the readers participate in the derivation of the molecular integrals. In Handout 4, they will implement the solutions in Python code. The reader with little or no programming experience should work through Handout 3 before Handout 4, because it is an introduction to the Python programming language, tailored to the integrals and HF programming projects.

Once the material in Handout 4 and Handout 5 (HF implementation) is complete, the reader will have fully developed the beginnings of an independent quantum chemistry code.

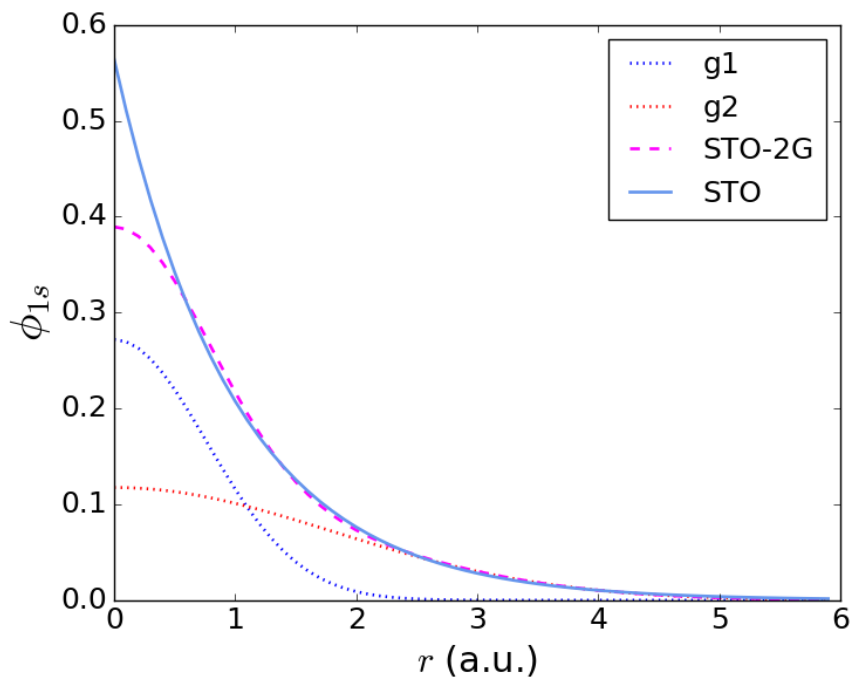


Figure 4: The approximation of a Slater-type orbital (hydrogen $1s$) by Gaussian-type functions with the STO-2G basis set, which is a *contraction* (linear combination) of two weighted (with contraction coefficients) Gaussian *primitives* (Gaussian functions “g1” and “g2”). Hence, the hydrogen $1s$ orbital constructed with the STO-2G basis set is a *contracted Gaussian-type orbital* (CGTO).

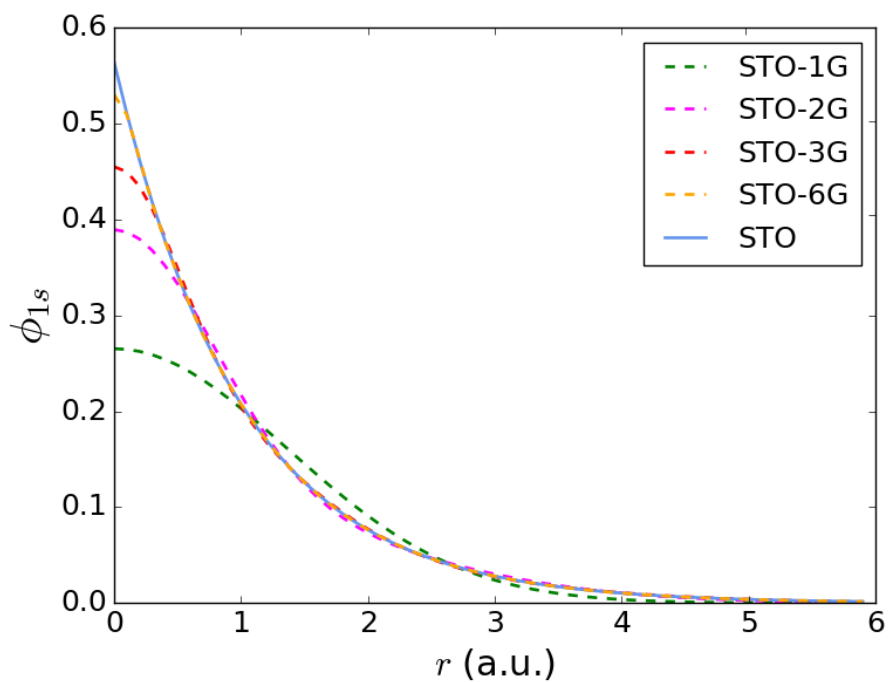


Figure 5: Increased numbers of Gaussian primitives in an orbital leads to an improvement in approximating the true Slater-type orbital. However, larger basis sets (those including more primitives) are also more computationally expensive.

\$basis			
*			
		basis set	
atom	C	STO-3G	
*			
		α	
3		s exponent factors	
		71.6168370	
		13.0450960	
		3.5305122	
# primitives per orbital	3	s	
		2.9412494	
		0.6834831	
		0.2222899	
3		p orbital type	
		2.9412494	
		0.6834831	
		0.2222899	
*			
\$end			

		d	
		contraction coefficients	
		0.15432897	$n=1$
		0.53532814	
		0.44463454	
		-0.09996723	$n=2$
		0.39951283	
		0.70011547	
		0.15591627	
		0.60768372	
		0.39195739	

Figure 6: An example of the structure of the tabulated STO-3G basis set for the carbon atom. Many table formats are found; this is “Turbomole” selected from the drop-down menu (with “Optimized General Contractions”) of the EMSL Basis Set Exchange, with “carbon” highlighted in the web site’s periodic table. The red text was added in this figure to highlight salient features of the table. n is the principal quantum number.

References

- [1] Born, M.; Robert, O. J. *Annalen der Physik* **1927**, *389*, 457–484.
- [2] Roothaan, C. C. J. *Rev. Mod. Phys.* **1951**, *23*, 69–89.
- [3] Hall, G. G. *Proc. Roy. Soc. A* **1951**, *205*, 541–552.
- [4] Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*; Dover: New York, 1996.
- [5] Page, T. R.; Boots, C. A.; Freitag, M. A. *J. Chem. Ed.* **2008**, *85*, 159–160.
- [6] Messina, M. *J. Chem. Ed.* **1999**, *76*, 1439–1443.
- [7] Grubbs, W. T. *J. Chem. Ed.* **2001**, *78*, 1557.
- [8] Slater, J. C. *Phys. Rev.* **1930**, *30*, 57–64.
- [9] Hoggan, P.; Ruiz Ruiz, M. B.; Ozdogan, T. In *Quantum Frontiers of Atoms and Molecules in Physics, Chemistry, and Biology*, 1st ed.; Putz, M. V., Ed.; Nova Science Publishers, Inc.: New York, 2010; pp 63–90.
- [10] Boys, S. F. *Proc. Roy. Soc. A* **1950**, *200*, 542–554.
- [11] Quinn, C. M. *Computational Quantum Chemistry*, 1st ed.; Elsevier: Cambridge, MA, 2002.
- [12] Mayer, I. *Simple Theorems, Proofs, and Derivations in Quantum Chemistry*; Springer Science & Business Media: New York, 2013; Chapter 9.2, pp 283–290.
- [13] Jensen, F. *Introduction to Computational Chemistry*, 3rd ed.; Wiley: Hoboken, New Jersey, 2017; Chapter 5, pp 188–228.
- [14] Cramer, C. *Essentials of Computational Chemistry*, 2nd ed.; Wiley: Hoboken, New Jersey, 2004; Chapter 6.2, pp 166–180.
- [15] Taketa, H.; Huzinaga, S.; Oohata, K. *J. Phys. Soc. Jap.* **1966**, *21*, 2313–2314.