EurJChem
An International Open Access Journal
European Journal of Chemistry
ATLANTA PUBLISHING HOUSE

View Journal Online
View Article Online

Unified formulae for augmented near orthonormalized STO-kG (1s, ..., 4f) basis sets via atomic orbital energy fit with graphical representations to use in molecular structure calculations

Sandor Kristyan *

Research Centre for Natural Sciences, (HUN-REN RCNS), H-1117 Budapest, Magyar Tudósok körútja 2, Hungary

* *Corresponding author at: Research Centre for Natural Sciences, (HUN-REN RCNS), H-1117 Budapest, Magyar Tudósok körútja 2, Hungary.*
e-mail: kristyan.sandor@ttk.hu (S. Kristyan).

RESEARCH ARTICLE

10.5155/eurjchem.16.4.382-394.2711

Received: 1 August 2025
Received in revised form: 31 October 2025
Accepted: 11 November 2025
Published online: 31 December 2025
Printed: 31 December 2025

KEYWORDS

Ab initio calculations
Computational chemistry
Unified STO-kG basis sets
Molecular structure calculations
STO-kG(3d and 4f) basis functions
Recursive formulae for overlap molecular integrals

ABSTRACT

STO-kG type (linear combination of k Gaussians) basis functions for 1s to 4f Hydrogen-like orbitals by 'energy fit' are reported as simple functions of the atomic number Z and quantum numbers to utilize in molecular electronic structure calculations. We analyze how they reproduce the one-electron atomic wave function shapes and energy values ($-Z^2/(2n^2)$) as an obvious primary claim along with reproducing the nodes exactly. The literature STO-kG yields a huge number of tables for different atoms in the periodic table; in contrast, our sets contain Z as a parameter yielding a compact list. A 'variation-like' property is also discussed for excited states (2s and up). The optimizations have been done by using least squares fits via the Lagrange multiplier method for energy with a constraint for normalization. All these STO-kG basis functions are normalized exactly to one in our tables. The general problem of STO-kG(3 and $4d_{u^2}$) and STO-kG($4f_{u^3}$ and $4f_{uv^2}$) basis functions among the six technical $3d_{uv}$ as well as the ten technical 4f used in practice is commented on. In relation and comparison, the Gaussian package STO-kG basis set is commented on.

Cite this: *Eur. J. Chem.* **2025**, *16(4)*, 382-394 **Journal website**: www.eurjchem.com

1. Introduction

The Schrödinger equation for the hydrogen-like atom placed its (M = 1) nucleus at the origin with the atomic number Z and N = 1 electron around it is $[h]\ \phi_{nlm} = E_n\ \phi_{nlm}$, with one-electron Hamilton operator (Equation 1).

$$[h] \equiv -½\nabla^2 - Z/r \tag{1}$$

The complete wave functions of the electron in hydrogen-like atoms have the form (Equation 2).

$$\phi_{nlm} = C\ s\ R_{nl}(\mathbf{r})Y_{lm}(\mathbf{r}) \tag{2}$$

In Equation 2, the quantum numbers have the values n = 1 (1s), 2 (2s, 2p), ..., l = 0, 1, ..., n-1, and m = -l, ..., 0, ..., l. Spin is present in physical reality, but since N = 1, it can be neglected mathematically here. To be clear, the spin is fully omitted from our formalism, as in the literature. The emblematic textbook solutions for energy in Equation 1 is well known as (Equation 3).

$$E_n = -Z^2/(2n^2) \tag{3}$$

A well-known fact is that in the case of multielectron atoms (N > 1), there are energy splits between 2s and 2p orbitals, as well as between 3s, 3p, and 3d, *etc.*; however, to create the STO-kG basis set, the one-electron atom is used for optimization (in the case of molecules, the ground state is notated as E_0). We define our computational energy error as (Equation 4).

$$ee\% = 100\ (E_n - E_{n,approx})/E_n \tag{4}$$

Equation 4 refers to how an STO-kG(orbital) basis function from an STO-kG basis set reproduces Equation 3.

The R_{nl} are related to the (real) associated Laguerre functions, and the Y_{lm} are the (in general, complex) spherical harmonics (expressed elegantly with polar coordinates). The $R_{nl}(\mathbf{r})Y_{lm}(\mathbf{r})$ is the spatial orbital of ϕ_{nlm} depending on the Cartesian (spatial) coordinates $\mathbf{r} = (x,y,z)$. In textbooks, Y_{lm} is expressed with polar coordinates ($r = (x^2+y^2+z^2)^{1/2}$, θ, φ); more specifically, it can be expressed with (θ, φ) only, but here we use Cartesian coordinates for the use of Equation 6 below: For example, the not-normalized $Y_{10} = \cos(\theta) = z/r$. The spin function $s = \alpha$ or β, and C (yet any constant) drops out of Equation 1. As in molecular structure calculations, the value of

ISSN 2153-2249 (Print) / ISSN 2153-2257 (Online)
https://dx.doi.org/10.5155/eurjchem.16.4.382-394.2711

C is fixed by normalization: $1 = \langle\phi_{nlm}|\phi_{nlm}\rangle = \int R_{nl}(\mathbf{r})^2|Y_{lm}(\mathbf{r})|^2 d\mathbf{r} = \int r^2 R_{nl}(r)^2|Y_{lm}(\theta,\varphi)|^2 \sin(\theta)\, dr d\theta d\varphi$. Important is how the polar coordinates separate in this M = N = 1 case. The $R_{nl}(r)$ is the hydrogen-like radial wave function, and the $r^2R_{nl}(r)^2$ is the radial distribution function beside the physical probability and one-electron density $(R_{nl}(\mathbf{r})Y_{lm}(\mathbf{r}))^2$. The $r^2R_{nl}(r)^2$ is a mathematical construction textbook slang, good for representative plots. Two other normalizations are used in textbooks: $\int|Y_{lm}(\theta,\varphi)|^2\sin(\theta)d\theta d\varphi = 1$ and $\int r^2R_{nl}(r)|^2dr = 1$, and their functional lists are calculated accordingly. The latter is possible because spherical symmetry is present in atoms. In relation to normalizations (Equation 5).

$$1 = \int R_{nl}^2|Y_{lm}|^2 d\mathbf{r} = \int r^2R_{nl}^2|Y_{lm}|^2 \sin(\theta)dr d\theta d\varphi = \int|Y_{lm}|^2\sin(\theta)d\theta d\varphi = 1 = \int r^2R_{nl}^2 dr \qquad (5)$$

The 2nd equality in Equation 5 comes from changing Cartesian coordinates to polar coordinates, while the 3rd - 5th ones are commonly used in textbooks [1], so R_{nl} and Y_{lm} have their own atomic normalization constants. If we take only the first two equality signs "=" for normalization in Equation 5, then Y_{ml} can be the raw ones in Table 1, and $\{c_i\}$ in R_{nl} takes care of the entire normalization; see Table 2. This is what we use in this work. The reason to use not-normalized Y_{lm} in Table 1 is that the products with accurate terms, $R_{nl}Y_{lm}$, and the approximate $f(\mathbf{r})Y_{lm}$ have the important form in Equation 6 next we need. All together, we consistently use below the R_{nl}, Y_{lm}, and f(r) functions as listed in Table 1, and the normalization for c_i is done by the 1st equality in Equation 5, as well as by the constraint in Equation 8 for A_i and B_i.

In relation to our rhetoric, we call Equation 2 (C: = 1 and neglecting s) as 'accurate solution Slater-type orbital' (ASTO), for a simpler discussion; in the literature STO term is used when exp(-a r) is the exponential part for a trial function.

A linear combination of k Gaussians to mimic the physically accurate ASTO solutions of one-electron atoms in Equation 2 is called the "STO-kG" basis functions with form (Equation 6).

$$R_{nl}(\mathbf{r})Y_{lm}(\mathbf{r}) \approx x^{L1}y^{L2}z^{L3}\exp(-Z\, r/n) \approx \text{STO-kG(orbital)} \equiv c_0\, x^{L1}y^{L2}z^{L3}\, \Sigma_{i=1}^{m} A_i \exp(-a_i\, r^2) \qquad (6)$$

In Equation 6 the first is a 'loose approximation', a 'mimicking ASTO with STO', but the accurate left-hand side can be achieved. The second approximation is a 'mimicking ASTO with GTO', a cruder one but vital for analytic molecular integration. The accurate $R_{nl}(\mathbf{r})Y_{lm}(\mathbf{r})$ are listed in Table 1. The c_0 in Equation 6 is the normalization constant, and the $L = L_1+L_2+L_3$ is a quasi-quantum number for atoms to mark functions as s-type (L = 0), p-type (L = 1), d-type (L = 2), f-type (L = 3). Although it is obvious, we emphasize rhetorically that STO-kG(orbital) is an 'STO-kG basis function' to mimic a single ASTO atomic orbital 1s, 2s, 2p, etc., while, for example, the {STO-kG(1s), STO-kG(2s), STO-kG(2p)} is a 'STO-kG basis set' to place this group on every atom individually in a molecular structure calculation. Especially, the latter is a minimal basis set for a carbon atom ($1s^22s^22p^2$), and a larger than minimal basis set for hydrogen ($1s^1$). For discussing STO-kG basis functions and basis sets, we fix the not-normalized Y_{lm} (Table 1), and we approximate the R_{nl} in Equation 6 with an $f(\mathbf{r})$, where f depends on (n,l) but is not marked, since it will be obvious in the discussion below. In this way, the spatial wave function approximation is $Y_{lm}f$.

If there were no difficulties in the analytical integration of ASTO functions (Equation 2), the 'accurate one-electron atomic solutions' in Tables 1 and 2 for Equation 2 below would be used as the best basis functions in HF-SCF/basis, DFT, CI, *etc.* calculations, and the atomic wave function approximations like STO-kG (right-hand side in Equation 6) would not be necessary at all (There are completely different basis sets, also, *e.g.*, plane-wave basis sets, which we do not discuss.) The first two initial requirements for an STO-kG(orbital) basis function are to approximate well the shape (Equation 2) and orbital energy (Equation 3) of the ASTO atomic orbital it represents.

Even if the ASTO basis functions do not have an analytical molecular integration problem (but they seriously do), a minimal basis set always needs improvement with the atomic excited states basis functions. We report analytical (k = 1) and numerical (k = 2, 3, 6) expressions for STO-kG (orbitals) as a function of the atomic number Z with respect to our "energy fit" device. In this way, the basis functions do not have to be listed separately for different atoms. We discuss some of their important properties in relation to the fit. This unified list (based on Z) is convenient to implement in programming molecular structures, and the same origin yields more consistent basis sets.

As a summary, our algorithm is not a (geometry or function) fit of the Gaussian part, $\Sigma_{i=1}^{m} A_i \exp(-a_i\, r^2)$, in Equation 6 to $\exp(-Z\, r/n)$, what we call a "radial fit" for this part of (the shape of the individual) wave functions of one-electron atoms. This is the basic practice in the literature for multi-electron atoms (followed by some energy optimizations as readjustment). Instead, we fit the entire $R_{nl}Y_{lm}$ in Equation 6 for full shape and one-electron energy, what we call an 'energy fit'. In the latter case, we reproduce Equations 2 and 3 as much as we can using low-order polynomials to fix the nodes exactly where they exist. Furthermore, we use $\exp(-a_i\,(Zr/n)^2)$ for unification (see discussion with Figure 1 below), and we keep the orthonormality of our basis functions as much as it is possible. These latter are not targeted and kept for STO-kG basis sets in the literature.

For unification, two expressions must be kept in mind: The norm $\int(A_1(Z/n)^{k+3/2}\, u^k \exp(-a_1(Zr/n)^2))^2 d\mathbf{r}$, wherein u = x, y and z does not depend on Z nor on n for k = 0, 1, 2, 3, ..., useful in modeling the s, p, d, f, ... orbitals, and as a consequence, the energy expressions $\langle Y_{lm}f|[h]|Y_{lm}f\rangle$ is proportional to $-(Z/n)^2$ as in Equation 3. On the other side, in the literature, the $\int((2a_i/\pi)^{3/4}\exp(-a_ir^2))^2 d\mathbf{r} = 1$ square normalized Gaussians are used in the linear combination to approximate exp(-Zr/n) individually for different particular (Z, n) values.

It is well known that minimal basis sets like STO-kG are hardly useful for practical electronic structure calculations on molecular systems, as they lack diffuse and polarization functions.

2. Necessary relations for fitting STO-kG basis sets

The characteristic radial, $R_{nl}(\mathbf{r})$, and auxiliary, $Y_{lm}(\mathbf{r})$, functions in Equation 2 are listed in Table 1. These achieve the accurate ASTO forms as P(x,y,z) exp(-Z r/n) in Equation 2 with the polynomial degree n-1. As indicated above, in molecular structure calculations (HF-SCF, post-HF-SCF, DFT, and various CI calculations), the exp(-Z r/n) part of it encounters a major difficulty in the evaluation of molecular multicenter integrals. Not only are analytic formulae not known, but the numerical integration is not easy, although possible [2,3]. The mathematical origin of this problem is that $\int\exp(-a|\mathbf{r}-\mathbf{A}|-b|\mathbf{r}-\mathbf{B}|)d\mathbf{r}$ has no analytical evaluation, while for Gaussians $\int\exp(-a|\mathbf{r}-\mathbf{A}|^2-b|\mathbf{r}-\mathbf{B}|^2)d\mathbf{r} = \pi^{3/2}(a+b)^{-3/2}\exp(-a\ b\ (a+b)^{-1}\ |\mathbf{A}-\mathbf{B}|^2)$ is simply available. For this reason, approximations with GTO in basis sets are used with $\exp(-a\, r^2)$ instead of exp(-Z r/n), which allows fast analytical integrations. In the literature, the exponential part, exp(-Zr/n) of R_{nl}, is approximated with $\Sigma_{i=1}^{m} A_i \exp(-a_i\, r^2)$ by least squares fit and called STO-kG atomic basis functions. This is what we call a 'radial fit'. However, this seemingly simple task is not as easy. There are many approximations providing different levels of accuracy, listed separately for the different regions of the periodic table.

Table 1. The form of accurate (ASTO) hydrogen-like radial wave functions, $R_{nl}(\mathbf{r})$, in the solution of Equation 1, and their approximation, $f(\mathbf{r})$, in STO-kG. The u, v, and w stand for simultaneous permutations of x, y, and z spatial coordinates; see restrictions in cases of d and f in the text.

Orbital name and quantum number (n, l, m) with $\vert m\vert$= 0, 1, ..., l for hydrogen-like ϕ_{nlm} = spin $R_{nl}(\mathbf{r})Y_{lm}(\mathbf{r})$		**$Y_{lm}(\mathbf{r})$, not-normalized spherical harmonics**	**P(r) of $R_{nl}(\mathbf{r})$ in Equation 9**	**$P_i(r)$ in STO-kG approximation f(r) in Equation 9 to mimic $R_{nl}(\mathbf{r})$ in the fit**
1s	(1,0,m)	1	$Z^{3/2} c_1$	$Z^{3/2} A_i$ (same as P)
2s	(2,0,m)	1	$Z^{3/2} (c_1 + c_2 Z r)$	$Z^{3/2} (A_i + B_i Z^2 r^2)$
2p (u)	(2,1,m)	u/r	$Z^{5/2} c_1 r$	$Z^{5/2} A_i r$ (same as P)
3s	(3,0,m)	1	$Z^{3/2} (c_1 + c_2 Z r + c_3 Z^2 r^2)$	$Z^{3/2} (A_i + B_i Z^2 r^2 + C_i Z^4 r^4)$
3p (u)	(3,1,m)	u/r	$Z^{5/2} (c_1 + c_2 Z r) r$	$Z^{5/2} (A_i + B_i Z^2 r^2) r$
3d (uv)	(3,2,m), u ≠ v	u v/r^2	$R_{32}(\mathbf{r}) = Z^{7/2} c_1 r^2$	$Z^{7/2} A_i r^2$ (same as P)
3d ($3u^2$-r^2)	(3,2,m)	($3u^2$-r^2)/r^2	$R_{32}(\mathbf{r})$	$Z^{7/2} A_i r^2$ (same as P)
3d (u^2-v^2)	(3,2,m), u ≠ v	(u^2-v^2) /r^2	$R_{32}(\mathbf{r})$	$Z^{7/2} A_i r^2$ (same as P)
3d (u^2)	(3,2,m)	u^2/r^2 (fake)	$R_{32}(\mathbf{r})$	$Z^{7/2} A_i r^2$ (same as P)
4s	(4,0,m)	1	$Z^{3/2}(c_1+c_2 Z r+c_3 Z^2r^2+c_4 Z^3r^3)$	$Z^{3/2}(A_i+B_i Z^2 r^2+C_i Z^4 r^4+D_i Z^6 r^6)$
4p (u)	(4,1,m)	u/r	$Z^{5/2} (c_1 + c_2 Z r + c_3 Z^2r^2) r$	$Z^{5/2}(A_i+B_i Z^2 r^2+C_i Z^4r^4) r$
4d (uv)	(4,2,m), u ≠ v	u v/r^2	$R_{42}(\mathbf{r}) = Z^{7/2} (c_1 + c_2 Z r) r^2$	$Z^{7/2} (A_i + B_i Z^2r^2) r^2$
4d ($3u^2$-r^2)	(4,2,m)	($3u^2$-r^2)/r^2	$R_{42}(\mathbf{r})$	$Z^{7/2} (A_i + B_i Z^2r^2) r^2$
4d (u^2-v^2)	(4,2,m), u ≠ v	(u^2-v^2)/r^2	$R_{42}(\mathbf{r})$	$Z^{7/2} (A_i + B_i Z^2r^2) r^2$
4d (u^2)	(4,2,m)	u^2/r^2 (fake)	$R_{42}(\mathbf{r})$	$Z^{7/2} (A_i + B_i Z^2r^2) r^2$
4f (u($3v^2$-u^2))	(4,3,m), u ≠ v	u($3v^2$-u^2)/r^3	$R_{43}(\mathbf{r}) = Z^{9/2} c_1 r^3$	$Z^{9/2} A_i r^3$ (same as P)
4f (uvw)	(4,3,m)	uvw/r^3	$R_{43}(\mathbf{r})$	$Z^{9/2} A_i r^3$ (same as P)
4f (uv^2)	(4,3,m), u ≠ v	uv^2/r^3 (fake)	$R_{43}(\mathbf{r})$	$Z^{9/2} A_i r^3$ (same as P)
4f (u^3)	(4,3,m)	u^3/r^3 (fake)	$R_{43}(\mathbf{r})$	$Z^{9/2} A_i r^3$ (same as P)
4f (u(v^2-w^2))	(4,3,m)	u(v^2-w^2)/r^3	$R_{43}(\mathbf{r})$	$Z^{9/2} A_i r^3$ (same as P)
4f (u($5u^2$-$3r^2$))	(4,3,m)	u($5u^2$-$3r^2$)/r^3	$R_{43}(\mathbf{r})$	$Z^{9/2} A_i r^3$ (same as P)
4f (u($5v^2$-r^2))	(4,3,m), u ≠ v	u($5v^2$-r^2)/r^3	$R_{43}(\mathbf{r})$	$Z^{9/2} A_i r^3$ (same as P)

Table 2. The accurate values of constants c_i of $R_{nl}(\mathbf{r})$ in Table 1 yielding $<\phi_{nlm}|\phi_{nlm}> = <R_{nl}Y_{lm}|R_{nl}Y_{lm}> = 1$ and $<\phi_{nlm}|[h]\phi_{nlm}> = -Z^2/(2n^2)$ via the corresponding second order equation system of c_i (Equation 15), or equivalently, from the ultimate $[h]\ \phi_{nlm} = -Z^2/(2n^2)\ \phi_{nlm}$. If c_1 is involved alone, the normalization fixes it, for example, for $3d_{uv}$ the $c_1 = (Z^7\int u^2v^2\exp(-2Zr/3)d\mathbf{r})^{-1/2} = (2/\pi)^{1/2}/81$. These values come from the normalization of the first two equalities in Equation 5, and are necessary for the fit in this work to create the STO-kG basis set. The huge ee% by $<\phi_{nlm}|[h]\phi_{nlm}>$ is emphasized when it differs from zero, e.g., for $3d_{u\wedge 2}$, which is not the solution of Equation 1 but a weak approximate wave function only. However, the $3d_{u\wedge 2}$ is widely used in STO-kG basis sets technically, so we list. The discrepancy simply comes from that $[h](3d_{x\wedge 2}) \neq \text{const}^*(3d_{x\wedge 2})$. In fact, it should not belong to this table; altogether, the (3 and 4)$d_{u\wedge 2}$, $4f_{uv\wedge 2}$ and $4f_{u\wedge 3}$ have this problem. The other values are 100% accurate, i.e., ee% = 0, not marked.

Orbital n	**c_1**	**c_2**	**c_3**	**E_n/Z^2**
1s	$1/\pi^{1/2}$	-	-	-1/2
2s	$1/(2\ (2\ \pi)^{1/2})$	$-1/(4\ (2\ \pi)^{1/2})$	-	-1/8
2p (u)	$1/(4\ (2\ \pi)^{1/2})$	-	-	-1/8
3s	$1/(3\ (3\ \pi)^{1/2})$	$-2/(9\ (3\ \pi)^{1/2})$	$2/(81\ (3\ \pi)^{1/2})$	-1/18
3p (u)	$2\ (2/\pi)^{1/2}/27$	$-(2/\pi)^{1/2}/81$	-	-1/18
3d (uv)	$(2/\pi)^{1/2}/81$	-	-	-1/18
3d ($3u^2$-r^2)	$1/(81\ (6\ \pi)^{1/2})$	-	-	-1/18
3d (u^2-v^2)	$1/(81\ (2\ \pi)^{1/2})$	-	-	-1/18 = -0.0555556
3d (u^2)	$(2/(3\ \pi))^{1/2}/81$	-	-	-13/162 = -0.0802469, ee% = -44.4
4s	$1/(8\ \pi^{1/2})$	$-3/(32\ \pi^{1/2})$	$1/(64\ \pi^{1/2})$, $c_4 = -1/(1536\ \pi^{1/2})$	-1/32
4p (u)	$(5/\pi)^{1/2}/32$	$-(5/\pi)^{1/2}/128$	$1/(512\ (5\ \pi)^{1/2})$	-1/32
4d (uv)	$(3/\pi)^{1/2}/128$	$-1/(512\ (3\ \pi)^{1/2})$	-	-1/32
4d ($3u^2$-r^2)	$1/(256\ \pi^{1/2})$	$-1/(3072\ \pi^{1/2})$	-	-1/32
4d (u^2-v^2)	$(3/\pi)^{1/2}/256$	$-1/(1024\ (3\ \pi)^{1/2})$	-	-1/32 = -0.03125
4d (u^2)	0.00334122	-0.000323037	-	-0.0393859, ee% = -26.0
4f (u($3v^2$-u^2))	$1/(3072\ (2\ \pi)^{1/2})$	-	-	-1/32
4f (uvw)	$1/(512\ (3\ \pi)^{1/2})$	-	-	-1/32
4f (uv^2)	$1/(1536\ \pi^{1/2})$	-	-	-1/24 = -0.0416667, ee% = -33.3
4f (u^3)	$1/(1536\ (5\ \pi)^{1/2})$	-	-	-1/20 = -0.05, ee% = -60.0 %
4f (u(v^2-w^2))	$1/(1024\ (3\ \pi)^{1/2})$	-	-	-1/32
4f (u($5u^2$-$3r^2$))	$1/(3072\ (5\ \pi)^{1/2})$	-	-	-1/32
4f (u($5v^2$-r^2))	$1/(1024\ (30\ \pi)^{1/2})$	-	-	-1/32

Although there are many Gaussian basis sets from different research groups, the accuracy of energy and the guide for choice are not totally controlled today. Mathematically, the backbone of the literature on STO-kG basis functions comes from minimizing (Equation 7).

$$<\exp(-Zr/n) - \Sigma_{i=1}^{m} A_i \exp(-a_i r^2) \mid \exp(-Zr/n) - \Sigma_{i=1}^{m} A_i \exp(-a_i r^2)> \qquad (7)$$

Equation 7 is minimized with a least squares fit (with certain modifications) for $\{a_i, A_i; i = 1, ..., m\}$. A primary test is how these basis functions reproduce Equation 2, and then these are tested in quantum chemical calculations for molecules where M and N >> 1.

We introduce and analyze another alternative, what we call the 'energy fit' as a counterpart to the 'radial fit'. In it, the STO part is bypassed as follows: We approximate the entire $R_{nl}Y_{lm}$ in Table 1 (and not only the exp(-Zr/n) part) with the help of Equation 1 by minimizing Equation 8.

$$\text{Min.}(<Y_{lm}(\mathbf{r})f(\mathbf{r})|[h]Y_{lm}(\mathbf{r})f(\mathbf{r})> - (-Z^2/(2n^2)))^2$$
$$\text{with constraint} < Y_{lm}(\mathbf{r})f(\mathbf{r})|Y_{lm}(\mathbf{r})f(\mathbf{r})> = 1 \qquad (8)$$

Equation 8 provides simpler algebraic equations to solve, and a 'variation-like' property helps. For our STO-kG approximation, we mimic the $R_{nl}(\mathbf{r})$ with $f(\mathbf{r})$ in Equation 8 listed in Table 1. The "variational principle" guarantees a (global) minimum only for the ground state; furthermore, we use the term "variation-like property" for the obtained stationary points, which are local minima describing excited states. The more detailed approximation is (Equation 9).

$$R_{nl}(\mathbf{r}) \equiv P(r) \exp(-Z\ r/n) \approx f(\mathbf{r}) \equiv \Sigma_{i=1}^{m} P_i(r) \exp(-a_i (Z\ r/n)^2) \quad (9)$$

The wise choice for P_i, which is an 'augmentation' toward better quality orbital energy and shape, is that 1. It is very close or the same algebraically as P (see Table 1); 2. Sometimes it can be pulled out from the sum (see our Tables 3-8 for STO-kG below). In Equation 9, the P, P_i, and f depend on quantum numbers (n.l), but they are not marked. As mentioned, the perfect ASTO basis set would be the set containing functions $R_{nl}(\mathbf{r})Y_{lm}(\mathbf{r})$ from Tables 1 and 2 for molecular structure calculations, but for feasible analytical molecular integrals, a well-optimized good mimic basis set that is the STO-kG basis set contains the functions (Equation 10).

$$f(\mathbf{r})Y_{lm}(\mathbf{r}) \text{ for } (n,l,m) \text{ quantum numbers} \tag{10}$$

Using Equation 10 and Table 1, the approximate orbital energy value, the $<Y_{lm}(\mathbf{r})f(\mathbf{r})[h]Y_{lm}(\mathbf{r})f(\mathbf{r})>$ has to be optimized for parameters in $f(\mathbf{r})$ so that this energy value is as close as possible to Equation 3. Thus, for a set in Equation 10, these qualified functions can be selected. For Table 1 we note that, for example, with principal axis z, the five d type r^2Y_{2m} are the $(3z^2-r^2)$, x^2-y^2, xy, xz and yz, the seven f type r^3Y_{3m} are the $r^3Y_{30}= z(5z^2-3r^2)$, $r^3Y_{3,\pm1} = x(5z^2-r^2)$, $r^3Y_{3,\pm2} = z(x^2-y^2)$, $r^3Y_{3,\pm3} = x(x^2-3y^2)$. From the latter, the set of seven orthonormal and real ASTO(4f) basis functions (Table 2) with z as the principal axis can be accomplished using $z(5z^2-3r^2)$, $x(5z^2-r^2)$, $y(5z^2-r^2)$, $z(x^2-y^2)$, $x(x^2-3y^2)$, $y(y^2-3x^2)$ and xyz; analogously for near-orthonormal STO-kG(4f) basis functions (Tables 5-8). The use of these accurate auxiliary functions for five d instead of six technical ones and for seven f instead of 10 technical ones provide more effective basis functions in computations indicated by the ee% values in Table 2. Notice that u^2, uv^2, and u^3 are 'fake' forms alone, but these come up in linear combinations. For example, the $3d_{u\text{^}2}$ is not a solution of Equation 1 but a technical basis function in computations. The magnetic m values belong to complex forms of $Y_{lm}(\mathbf{r})$, see in [1]. When two or three of the variables u, v, and w appear, this automatically means that $u \neq v \neq w \neq u$, some are marked. The 'fake' means that, it is not a '$Y_{lm}(\mathbf{r})$ function', but is used in 'basis functions' in computations.

The complex versions of Y_{lm} in ϕ_{nlm} in Equation 2 have been transformed to their real versions [1] by linear combinations of degenerate states (allowed by the linear operator in Equation 1), listed in Table 1. The not-normalized Y_{lm} is used in this work, and the constants c_i in R_{nl} provide the entire normalization (see Equation 5), listed in Table 2. The $3d_{u\text{^}2}$ orbitals are not solutions of Equation 1, but are heavily used as technical basis functions in molecular structure calculations, so their properties are also shown in Table 1. The magnetic quantum number m of Y_{lm} in the right of Equation 10 is strictly used notations in literature. The magnetic quantum number is not used in this work but is very significant in Y_{lm} in quantum theory, so we cannot leave it out of our notation. As reasoned above, our $P_i(\mathbf{r})Y_{lm}(\mathbf{r})$ must not contain r^k with odd k; otherwise, the molecular integrals cannot be evaluated analytically when Equation 10 is used as the basis set. It is completely satisfied with this work, as Table 1 reveals. Mark "(same as P)" in Table 1 indicates when P_i is the same as P aside from constants, and the missing (same as P) indicates when $Zr \to (Zr)^2$ and $(Zr)^2 \to (Zr)^4$ transformations are necessary, along with P_i should be close to P as much as possible algebraically, e.g., a $Zr \to (Zr)^{20}$ would obviously distort. In this work, we consider the simpler but still powerful cases when P_i can be pulled out from the sum in Equation 9 (see Table 1) as $P_f\ \Sigma_{i=1}{}^m\ A_i \exp(-a_i\ (Z\ r/n)^2)$. See our tables (Tables 5-8) for STO-kG below.

It is important to note for our discussion that the three physical p orbitals (p_x, p_y, p_z) are represented by one algebraic orbital $p_u= u \exp(-Z\ r/n)$, so the five (seven) physical d (f) are represented by three (five) algebraic orbitals (Table 1).

As detailed above, we have chosen the algebraic form in Table 1 to be as close to R_{nl} as possible in Equation 9 by P_i vs. P (followed by an energy fit for the parameters in f). The restriction in choice for the algebraic form for fY_{lm} is to contain only positive powers in its r^k terms (see Table 1), which allows analytical evaluations for molecular integrals. In this way, the exponents a_i are supposed to counterbalance these powers (see, e.g., r *vs.* r^2 for 2s in Table 1) beside the STO → GTO function switch in Equations 6 and 9. To compare Equation 6 and Equation 9 in more detail, for example, the 2s is approximated with $c_0\ \Sigma_{i=1}{}^m\ A_i \exp(-a_i\ (Z\ r/n)^2)$ by Equation 6 while using Equation 9 and Table 1, we approximate it with the $Z^{3/2}\ (A + B\ Z^2\ r^2)\ \Sigma_{i=1}{}^m \exp(-a_i\ (Z\ r/n)^2)$. (We also note that functions $r^k \exp(\exp{-a}\ r^{(1 \text{ or } 2)})$ with a maximum at $r = k/a$ or $(k/(2a))^{1/2}$, resp., have the same global property for $k > 0$ in the range $r \geq 0$, but the initial slopes differ significantly, and there is a shift between maximums. Particularly, the k=1 curve is close to k = 2, and the k = 2 curve is close to k = 4 at a common value of "a", and the shift can be counterbalanced by the linear combinations in P_i. The powers of Z as used in f will be explained below; they pull out from the energy integrals, while ones in the literature that use Equation 6 do not have this good property. Our approximations are more powerful and more unified because the polynomial part of them is closer to the physically correct polynomials in R_{nl}. More specifically, the nodal property of 2s, for example, is fixed by the polynomial, unburdening the sum of Gaussians; the latter does it solely with sign variation (as mentioned in the Introduction, and see literature). In summary, the one-electron atomic orbitals, $R_{nl}Y_{lm}$, in Table 1 are approximated better by fY_{lm} in Table 1 than by Equation 6 for the STO-kG basis sets.

For 3d orbitals in Table 1, the $R_{32}(\mathbf{r})$ is a common polynomial aside from c_i, but their $Y_{2m}(\mathbf{r})$ are different. In the 3d case, the latter is distinguished by the 'energy fit'. The numerical values of these constants, c_i, are given in Table 2, among others. In the practice of computational quantum chemistry, 6 technical $3d_{uv}$ orbitals with full permutation for u = x, y, z are used instead of the theoretical 5 ones. Some discrepancies coming from the fact that the $3d_{u\text{^}2}$ is not a solution to Equation 1 will be detailed later. The analogous argument holds for 4f.

In Table 1, the parameters c_i of all listed orbitals depend on quantum numbers (n, l), so the parameters A_i, B_i, and C_i. With energy-optimized $f(\mathbf{r})$ in Equation 8, the $Y_{lm}(\mathbf{r})f(\mathbf{r})$ is the spatial part of the approximate one-electron atomic wave function, the $(Y_{lm}(\mathbf{r})f(\mathbf{r}))^2$ is the approximate one-electron density, and $r^2f(r)^2$ is the approximate radial distribution function of the electron around the one-electron atom. An important textbook fact is that the normalization of the approximation for ϕ_{nlm} (not only at N = 1 but for N >1 cases also) allows one to energy-minimize the $<S_0|H_{el}S_0>$ only instead of $<S_0|H_{el}S_0>/<S_0|S_0>$ in, *e.g.*, HF-SCF calculations with the Slater determinant S_0. It makes the Lagrange multiplier method below easier too. Small Gaussian expansions of Slater-type orbitals by the method of least squares have been presented in references [4,5], where the expansion lengths are from k = 1 to 6 for 1s through 5g Slater-type orbitals. We use the first (left) normalization in Equation 5 (and not the last) in Table 2, because we focus on energy optimization (called the 'energy' fit). Table 2 contains the accurate values for the c_i parameters in this way, as defined in Table 1.

Orbitals such as $3d_{u\text{^}2}$ or $4f_{u\text{^}3}$, which are not eigenfunctions of the hydrogenic Hamiltonian (in Equation 1 because of their polynomial part), are still frequently treated along with physical orbitals. Their purely technical nature as auxiliary basis functions should be emphasized, and a note on them follows for $n \geq 3$. Other basis functions of the basis sets correct the problem of their weak quality. Their problem is indicated by their huge ee% values in Table 2. However, one can do a better job using the following device. The coordinate system is always fixed for a molecular frame in any particular calculation. Then, instead of the six technical d basis functions centered on atoms in the molecular frame, the more plausible five basis functions should be used. That is, on all atoms, besides the (correct) STO-kG($3d_{xy}$), STO-kG($3d_{xz}$), and STO-kG($3d_{yz}$), the (also correct and z axis fixed) STO-kG($3d3_{z\text{^}2-r\text{^}2}$) and STO-kG($3d_{x\text{^}2-y\text{^}2}$) should be used instead of the (problematic) STO-kG($3d_{x\text{^}2}$), STO-kG($3d_{y\text{^}2}$),

Table 3. Cross table to exhibit the orthonormality property, ie the overlap integrals <STO-3G (orbital-a)| STO-3G(orbital-b)> for the emblematic Gaussian94 basis set and of this work for comparison. This matrix is symmetric; only the upper diagonal part is filled up. The <STO-3G(1S or 2S)|STO-3G(2P> = 0 comes out trivially by the even or odd parity of the functions. Generally, an {STO-3G(1S), STO-3G(2S), STO-3G(2P)} basis set should yield a unit matrix (for the sub-sets that are placed on the same atom, of course); also, any STO-kG basis set should own this because Equation 2 is perfectly orthonormal and the STO-kG basis set is supposed to mimic Equation 2. The functions in this table are placed at the origin. The Gaussian94 STO-3G basis set is surprisingly far from orthonormal. Our basis set is much closer to orthonormality. The Gaussian94 basis functions are based on Equation 7 with reoptimization for molecular energies, while our functions are from the optimization of Equation 13. The S and P are Gaussian94 notations; these mean s and p orbital mimics, respectively, are used in the last four columns here. (Here, we use the same notation, STO-3G, for both basis sets.)

Orbitals	**STO-3G Gaussian 94**				**STO-3G, Table 7**			
	H, 1S	**C, 1S**	**C, 2S**	**C, 2P**	**H, 1S**	**C, 1S**	**C, 2S**	**C, 2P**
H, 1S	1.5376	3.12026	2.06328	0	1	0.328099	-0.779766	0
C, 1S		32.1489	2.42213	0		1	-0.0792236	0
C, 2S			2.9584	0			1	0
C, 2P				1.61243				1

Table 4. The individual one-electron atomic orbital energies, En,approx. = <STO-3G |[h]| STO-3G>/<STO-3G|STO-3G> for diagonal elements in Table 3 to compare with the accurate $-Z^2/(2n^2)$ value in Equation 3 (the variation principle holds for 1S). The energy error ee% in Equation 4 is in the bracket. (We use the ratio for $E_{n,\mathrm{approx.}}$, because the Gaussian94 functions are not normalized. Energy values 18, 4.5, and 4.5 hartree refer to the energy of basis functions, not the accurate correlated orbital energy for a carbon atom).

Creator of STO-kG	$E_{n,\mathrm{calc}}$**(H, 1S)** ***vs.*** **-1/2**	$E_{n,\mathrm{calc}}$**(C, 1S)** ***vs.*** **-18**	$E_{n,\mathrm{calc}}$**(C, 2S)** ***vs.*** **-4.5**	$E_{n,\mathrm{calc}}$**C, 2P)** ***vs.*** **-4.5**
Gaussian94, STO-3G	-0.466578 (6.7%)	-17.7615 (1.3%)	-4.71752 (-4.8%)	-3.29131 (26.9%)
Table 5, STO-1G	$-4/(3\pi) = -0.424413$ (15.1%)	$-48/\pi = -15.2789$ (15.1%)	-3.9492 (12.2%)	$-64/(5\pi) = -4.07437$ (9.5%)
Table 6, STO-2G	-0.478896 (4.2%)	-17.2402 (4.2%)	-4.22222 (6.2%)	-4.37785 (2.7%)
Table 7, STO-3G	-0.491739 (1.7%)	-17.7026 (1.7%)	-4.3051 (4.3%)	(-4.47322) (0.6%)
Table 8, STO-6G	-0.498513 (0.3%)	-17.9465 (0.3%)	-4.39848 (2.3%)	-4.49262 (0.2%)

and STO-kG($3d_{z^2}$), opposite to today's practice. (Cyclic permutation of x, y, and z applies for designating the z axis, of course.) The situation is analogous for higher excited than d atomic orbitals as basis functions: nf, ng, *etc.*

3. How about numerical molecular integrals

The computational importance of Tables 1 and 2 follows: If the Slater-type ASTO atom-centered orthonormalized basis set is used in quantum chemical calculations for molecular structures, as the best basis set, only numerical integrations are available [2], and the parameters of these accurate basis functions, the $R_{nl}(\mathbf{r})Y_{lm}(\mathbf{r})$, are listed in Table 2 for immediate use (Statement 1). The STO-kG basis functions are to mimic these ASTO in Tables 1 and 2, so if we are satisfied with numerical integrations for molecular integrals instead of analytical ones, the STO-kG fit is not needed at all. However, analytical molecular integration is in the fashion in literature, so we are continuing our work here.

4. The 'Radial Fit' *vs* 'Energy Fit' in production of STO-kG basis sets

The functional relationship between the shape of STO type $\exp(-Zr/n)$ vs. GTO type $\exp(-a_i (Zr/n)^2)$ in the range $0 \le r < \infty$ is that both decrease strictly monotonically from 1 to 0, but the GTO one is steeper. Their derivatives at r= 0 are -Z/n vs. 0, and these properties are what $f(\mathbf{r})$ has to counterbalance primarily in Equation 9, the other is the nodal property, if any, for example, no nodes in 1s but there are in 2s, *etc.* The 'radial fit' for Equation 9 can be done with the simple Equation 7 or the more sophisticated Equation 11 below. For this, the first normalization for $R_{nl}(\mathbf{r})$ in Equation 5 is supposed to be used as in Tables 1 and 2; however, the last normalization in Equation 5 could also be used, but it makes unwanted weightings at large r. Even more, e.g., the $\int R_{nl}(\mathbf{r})^2 d\mathbf{r} = 1$ could also be used; the key is to reproduce the graph of $R_{nl}(\mathbf{r})$ with $f(\mathbf{r})$, and for this, one needs particular values for c_i in Tables 1 and 2. Having these values in hand, for an optimal STO-kG via 'radial fit', one has to minimize, using least squares fit, for shape (S) the

$$S \equiv \langle f(\mathbf{r}) - R_{nl}(\mathbf{r}) \mid f(\mathbf{r}) - R_{nl}(\mathbf{r})\rangle = \int (f(\mathbf{r}) - R_{nl}(\mathbf{r}))^2 d\mathbf{r} \tag{11}$$

Equation 11 provides a better quality, or more flexible $f(\mathbf{r})$ than Equation 7, because Equation 7 focuses only on the exponential part, while Equation 11 contains the radial polynomial part also; however, the algebraic difficulties are the same, as commented on next. As mentioned, the nodes (polynomials) in the radial part are taken into account via the altering sign of Gaussians with different exponents in the literature ones; Equation 11 is an alternative and better improvement. Equation 7 and Equation 11 can be evaluated only numerically for the parameters in Equation 9. The difficulty of numerical optimization is demonstrated by an example. The simplest case for Equation 11 is to mimic 1s with STO-1G as $S= Z^3 \int_0^\infty (A_1 \exp(-a_1 (Z r)^2) - \pi^{-1/2} \exp(-Z r))^2 d\mathbf{r}$. An advantage is that Z drops from S after integral evaluation, which is not detailed, but this advantage is not taken in the literature because $\exp(-a_1 r^2)$ is used. Mathematically, it is necessary to minimize S, i.e., the $\{\partial S/\partial A_1=0, \partial S \partial a_1=0\}$ system has to be solved. By the square, the integral for the cross term $\int_0^\infty \exp(-b r - a r^2)dr = ½ \pi^{1/2} a^{-1/2} \exp(b^2/(4 a)) \,\mathrm{Erfc}(b/(2 a^{1/2}))$ comes up. The system is second order in "A_1" and far non-linear in "a_1", and because the Erfc function enters, the numerical treatment of this system is not easy for higher excited states in Table 1, mainly when one needs to create STO-kG basis functions with m>1. Basically, these systems are solved numerically (using Equation 7) in the literature to produce the STO-kG basis set with some improvements (see Introduction); see [5-7] or the STO-3G by Gaussian94 exhibited in Tables 3 and 4. The complexity of numerical evaluation is, of course, a "thing of the past" when one has the STO-kG basis set in hand to use in molecular structure calculations. However, we will show that the 'energy fit' of this work is not only easier, but also produces better characteristics in relation to Equations 1-3.

The problem with Erfc coming out from Equation 11 can be bypassed by fitting the approximate functions, $f(\mathbf{r})$, in Table 1 to the energies in Equation 3 instead of the ASTO solutions in Equation 2. Its algebraic background is that the cross term $\int \exp(-br -ar^2)d\mathbf{r}$ in "radial fit" (for shape) produces Erfc, but the corresponding $\int \exp(-(a+b)r^2)d\mathbf{r}$ in "energy fit" does not. For example, for the ground state 1s orbital of a one-electron atom, the STO-1G basis function is STO-1G(1s) = $Y_{0m}f = Z^{3/2} A_1 \exp(-a_1(Z r)^2)$, see Table 1. To optimize its parameters, one can use the variation principle (because it is a ground state) and minimize L (a Lagrangian, because it is an energy) by solving the system $\{\partial L \partial A_1=0, \partial L \partial a_1=0\}$, wherein in this simple case, A_1 drops, yielding final Equation 12 reproduces the 1s orbital in Equations 1-3 on the STO-1G basis level with a 15.1% error, as listed in Table 5.

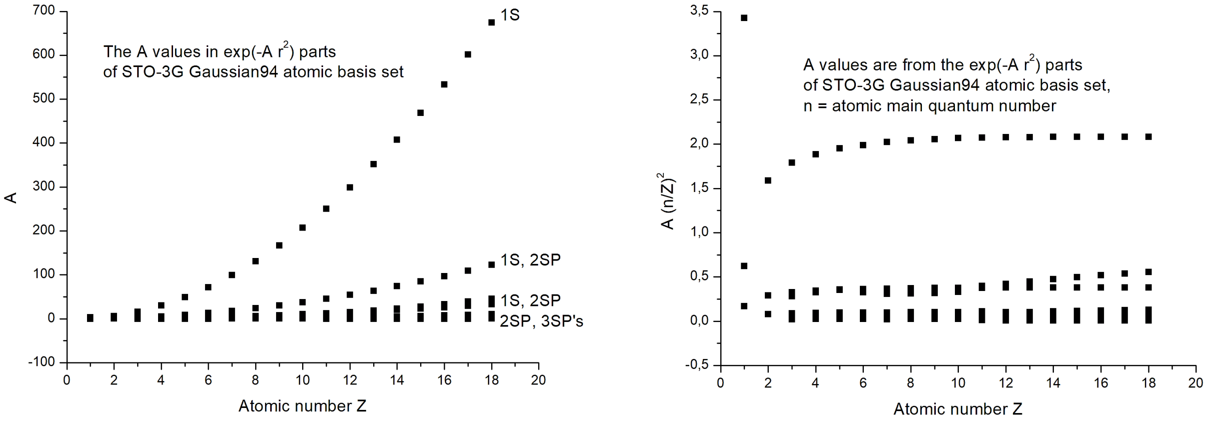

Figure 1. Exhibiting that unification of the STO-3G atomic basis set is possible in relation to atomic number Z, so in the cases of all STO-kG basis sets. For example, the three exponents of the STO-3G Gaussian94 atomic basis set [8] for the 2SP orbital of Ar (Z = 18, n = 2) yield this correspondence as $45.164244*(2/18)^2 = 0.557583$, $10.495199*(2/18)^2 = 0.12957$, and $3.4133644*(2/18)^2 = 0.0421403$. These values are plotted for Z = 1 (H) - 18 (Ar), including all their 1S, 2SP, and 3SP orbitals (if any, i.e., 1S for H-He, 1S and 2SP for Li-Ne, and 1S (~n = 1), 2SP (~n=2), and 3SP (~n = 3) for Na-Ar). The exponent A in the left figure is replotted as $A*(n/Z)^2$ in the right, exhibiting that the quasi-parabolic trends transform to quasi-constants in this statistic. The order of curves is the same in both plots; the two upper ones correspond to each other, etc.

$$L = \langle Y_{0m}f|[h]Y_{0m}f\rangle / \langle Y_{0m}f|Y_{0m}f\rangle = \left(\frac{3a1}{2} - 2\sqrt{\frac{2}{\pi}}\sqrt{a1}\right)Z^2 \Rightarrow$$
$$a1 = \frac{8}{9\pi}, E_1 = -Z^2/2 \approx L = -\frac{4Z^2}{3\pi} = -0.424413\,Z^2 \tag{12}$$

On the STO-1G level, like in Equation 12, analytical expressions can be obtained, but at higher k values, only numerical evaluations are possible; see below. To give a taste of the increasing algebraic complexity when, e.g., STO-1G(1s) in Equation 12 is extended to STO-2G(1s) level to decrease the ee% value, see Equation 19 at the end of this work (discussed among some other integrals) for the energy integral in the Lagrange multiplier method (Equation 13), L. The L is second order in A_i with i= 1, 2 in STO-2G, but reciprocal powers of a_i come up; see Equation 19. One faces even more complicated expressions for STO-kG with k>2 if k increases. Analytical solutions cannot be obtained for $\{A_i, a_i\}$ in these higher-level (m>1) cases; compare Tables 5 *vs.* 6-8. We have developed the numerical algorithm for energy optimization as detailed below, and toward this, the first step was to pull out certain powers of Z in f(**r**) as unification; see Table 1.

We provide some of our results in advance before going on to describe our work. Table 3 exhibits that our STO-3G basis functions have better orthonormality properties than the Gaussian94 ones (see how numbers 0 and 1 are reproduced), and Table 4 exhibits that our STO-kG basis functions yield better atomic orbital energy (N=1) as k increases (see how ee% values decrease).

The most convincing statistic for a possible unification of STO-kG atomic basis sets in relation to the atomic number Z follows. In our STO-kG atomic basis functions, the exponential parts are $\exp(-a(Z\,r/n)^2)$, see Equation 9, while in the STO-kG Gaussian94 atomic basis functions, the exponential parts are $\exp(-A\,r^2)$, see [8]. It simply follows that $a = A\,(n/Z)^2$, or at least it is a correspondence (~ instead of =) relation. The pair of plots in Figure 1 is extremely impressive because the quasi-parabolic A as a function of (n,Z) in the left transforms to the quasi-constant a $\sim A*(n/Z)^2$ function in the right, which validates our procedure of unification. Notice that our "a" values and the Gaussian94 basis set [8] "A" values are from completely different energy fits, as well as our a_1, a_2, and a_3 values in Tables 5-7 below for our STO-3G basis set show very strong correspondence to the quasi-constant values of branches in the right side of Figure 1.

5. Algorithm of the "Energy Fit" for STO-kG Basis Sets

The 'energy fit' method optimizes the parameters of f(**r**) listed in Table 1 by minimizing the energy with normalization constraints by solving the system $\{\partial L/\partial A_i = 0, \partial L/\partial a_i = 0, \partial L/\partial \lambda = 0\}$ using Equation 13. This 'energy' fit yields analytical expressions for parameters in Equation 9 for k = 1 and numerical parameters for k>1. Importantly, it fits not only for the full shape of the wave function ($Y_{lm}R_{nl}$), but also for the orbital energy (Equation 3). In the literature, only part of the wave function shape, the exp(-Zr/n) part of R_{nl}, is targeted in the fit, essentially via Equation 7, followed by adjustment from test calculations for molecular energies. Algebraically, the second derivatives of the operator in Equation 1 are not taken into account in Equations 7 or 11 primarily. We optimize, more precisely minimize the Lagrangian for the STO-kG basis functions with any of the two equivalent functiontionals (Equations 13 and 14).

$$L = \langle Y_{lm}f|[h]Y_{lm}f\rangle + \lambda(\langle Y_{lm}f|Y_{lm}f\rangle - 1) \geq -Z^2/(2n^2) \tag{13}$$

$$L = \langle Y_{lm}f|[h]Y_{lm}f\rangle / \langle Y_{lm}f|Y_{lm}f\rangle \geq -Z^2/(2n^2) \tag{14}$$

Another key point in Equations 9, 11, and 13,14 is the algebraic correspondence $-(Z\,r/n) \rightarrow -a_i\,(Z\,r/n)^2$, the literature uses the poorer $-(Z\,r/n) \rightarrow -a_i\,r^2$ only. As mentioned, in this way the Z eliminates from normalization $\langle Y_{lm}f|Y_{lm}f\rangle$ and Z^2 pulls out from energy as $\langle Y_{lm}f|[h]Y_{lm}f\rangle \rightarrow \langle Y_{lm}f|[h]Y_{lm}f\rangle\,Z^2$ in accordance with Equation 3, and finally, only the parameters $\{A_i, a_i, \ldots\}$ remain in the two brakets of Equations 13 and 14. As a consequence, the optimizations belong to general Z and not particular Z values, a convenient unification. Equations 13 and 14 yield the same optimization result, but Equation 13 provides analytic expressions for k=1 also, as well as Equation 14 can provide numerical results more conveniently if k is larger.

The mimic of the ground state 1s orbital has come out from Equation 12 above, called STO-1G(1s); generally, the minimum via Equations 13 and 14 for 1s is in agreement with the variation principle. For excited states (2s, 2p, 3s, *etc.*), the good algebraic forms for f(**r**) in Table 1 offer an important "variation-like" principle. That is, for example, the $2p_x$ mimic $Y_{1m}f = Z^{5/2}$ x $A_1 \exp(-a_1(Z\,r/2)^2)$ cannot fall as an energy minimum to the energy of the 1s mimic $Y_{0m}f = Z^{3/2}\,A_1 \exp(-a_1(Z\,r/2)^2)$ via Equations 13 and 14; the physically accurate polynomial (now $P_f = x$) prevents this.

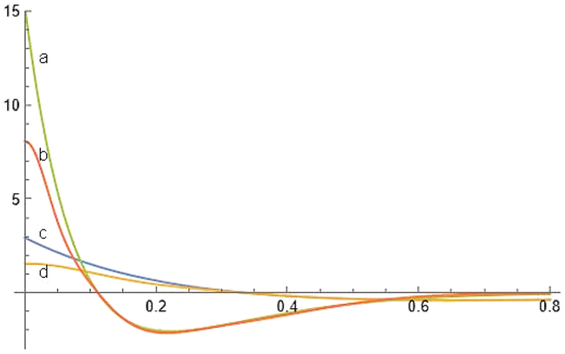

Figure 2. The $R_{20}(r) = Z^{3/2}$ $(c_1 + c_2 Z r) \exp(Z r/2)$ part of the ASTO(2s) = $Y_{0m}R_{20}$ solution of Equation 1 from Tables 1 and 2, and the $f(r) = Z^{3/2}$ $(1 - 0.25(Z r)^2)$ $g_6(\mathbf{r})$ part of the STO-6G(2s) basis function from Table 8. The $<Y_{0m}R_{20}(\mathbf{r})|Y_{0m}R_{20}(\mathbf{r})> = <Y_{0m}f(\mathbf{r})|Y_{0m}f(\mathbf{r})> = 1$ for norms. It represents that the accurate ASTO (orbital) and approximate STO-kG(orbital) of this work intersect the horizontal r axis at exactly the same values, and the algebraic form for Z makes unification for the periodic table, as well as the odd power of r in R_{20} *vs.* the even power of r in f, which does not prevent it from well tracking the graph it mimics. Intersections of the curves on the vertical axis in decreasing order: a: ASTO(2s, Argon, Z = 18), b: STO-6G(2s, Argon, Z = 18), c: ASTO(2s, Carbon, Z = 6), d: STO-6G(2s, Carbon, Z = 6). General rule: $R_{nl}(Z,r = 0) > f(Z,r = 0)$.

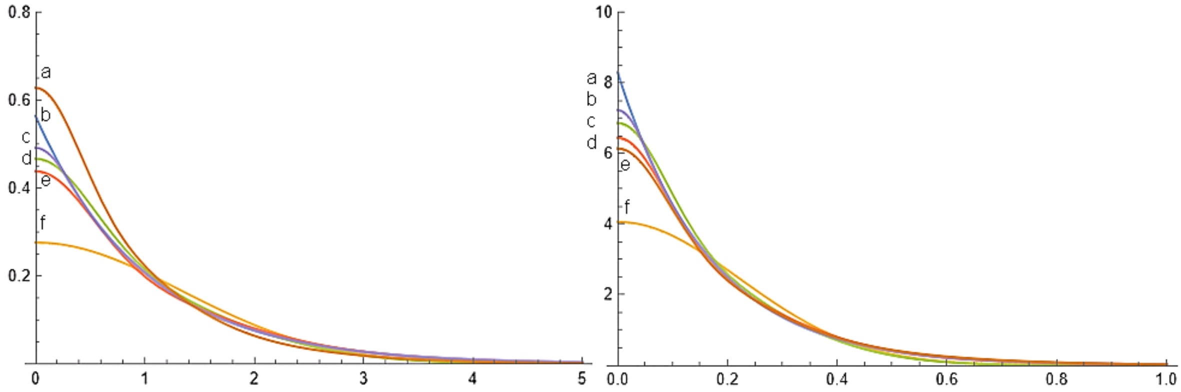

Figure 3. The $R_{nl}(r)$ part of ASTO(1s) from Tables 1 and 2 and the f(r) part of STO-kG(1s) with k = 1, 2, 3, 6 from Tables 5-8, as well as the Gaussian94 f(r) part of STO-3G(1s) form [8] basis functions, r is the horizontal axis. Vertical axis intersection in decreasing order: a: STO-3G(Gaussian94, 1S), b: ASTO(accurate), c: STO-6G, d: STO-2G, e: STO-3G, and f: STO-1G for hydrogen (left, Z=1) and a: ASTO(accurate), b: STO-6G, c: STO-2G, d: STO-3G, e: STO-3G(Gaussian94, 1S), and f: STO-1G for carbon (right, Z=6). The $<Y_{lm}R_{nl}(\mathbf{r})|Y_{lm}R_{nl}(\mathbf{r})> = <Y_{lm}f(\mathbf{r})|Y_{lm}f(\mathbf{r})> = 1$ holds for the norms.

This variation-like principle holds if one knows the accurate algebraic forms of wave functions (Table 1), or at least very plausible approximate functions are in hand. Fortunately, it is available for one-electron atoms (Table 1), but the case of molecules with many electrons is more complicated.

Statement 2: $E_n(c_1,c_2,\ldots) = L$ in Equations 13 and 14, for ground and excited states, it has energy minimum with algebraic forms $R_{nl}(c_1,c_2,\ldots)$ in Table 1, and the minimums are at the parameter values in Table 2 with energy values in Equation 3 (Ground state 1s is global minimum, the excited states are local minimums.) Actually, it recovers the exact solutions in this way as $[h](Y_{lm}R_{nl}) = -(Z^2/(2n^2))Y_{lm}R_{nl}$. It is in agreement with the known variation principle for ground states and a 'variation-like' property for excited states used in this work. If the algebraic form of $Y_{lm}R_{nl}$ and $Y_{lm}f$ is 'very close' to each other, as in Table 1, the 'variation-like' property holds for these approximate ground and excited states $Y_{lm}f$ as minimum in Equations 13 and 14. This provides the 'energy fit' by minimization for the STO-kG basis sets in this work.

In the application of Equations 13 and 14, the augmentation in Equation 9 is very powerful if one requires P_i to fix the nodes, that is, to fix the intersections on the r-axis (if any). For example, see the $(1\text{-}0.25(Zr)^2)$ part of the STO-kG(2s) basis functions in Tables 5-8, *etc*., and Figures 2-8. The minimization of Equations 13 and 14 is to solve the system (Equation 15).

$$\{\partial L/\partial A_i = 0, \partial L/\partial B_i = 0, \text{etc.}, \partial L/\partial a_i = 0, \partial L/\partial \lambda = 0\} \qquad (15)$$

As mentioned, Equation 13 provides multi-variable second-order polynomials for A_i, B_i, etc. in the 'energy fit', while reciprocal polynomials for a_i. The full details of the fits for STO-kG basis functions with k = 1, 2, 3, and 6 are straightforward in this way. In this task, we have used symbolic algebra from the Wolfram package [9]. Tables 5-8 list the results for $Y_{lm}f$ for Equation 10, as the main results of this work; see the next sections.

An important additional step for Equation 15 in our task is that after calculating A_1 and a_1 (Table 5) for STO-1G, for STO-2G, a_1 is taken from the previous STO-1G (Table 6), and only A_1 (recalculation), A_2, and a_2 are calculated; similarly, for STO-3G and STO-6G (Tables 7 and 8); it makes this nonlinear fit much easier, but still, ee% can be decreased drastically.

An important example of this work for Statement 2 follows. In Table 2, the 3d(u^2) orbital energy is much deeper (ee% = -44.4 %) than that of the other true 3d orbitals. It is so because 3d(u^2) is a false function for atomic orbitals, i.e., not a solution of Equation 1. However, in Tables 5-8 and Figure 6, the variation-like property holds for plausible approximate functions for excited states, $Y_{lm}f$, as indicated with their ee% > 0 and close to 0 values along with their plausible graphs.

To show that our algebraic forms follow the zero places (nodes) exactly, as well as the global properties (maximums, minimums, and asymptotes), see, e.g., Figure 7.

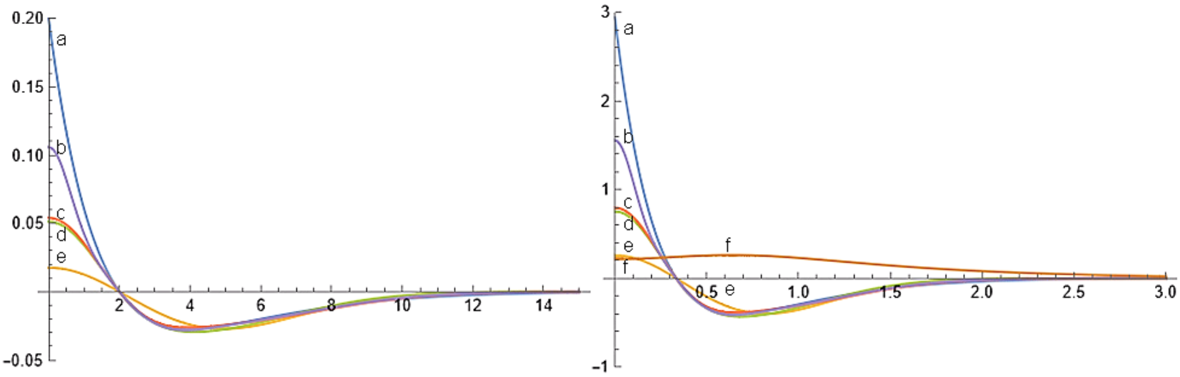

Figure 4. Same as Figure 3, but for orbital = 2s: Hydrogen (left, Z = 1) and carbon (right, Z = 6), r is the horizontal axis. Vertical axis intersection in decreasing order: a: ASTO(accurate), b: STO-6G, c: STO-3G, d: STO-2G, and e: STO-1G for Hydrogen, as well as a: ASTO(accurate), b: STO-6G, c: STO-3G, d: STO-2G, e: STO-1G, and f: STO-3G (Gaussian94, 2S) for carbon. The Gaussian 94 for Carbon, in this case, has relatively good energy (see also Table 4), but is in very poor shape.

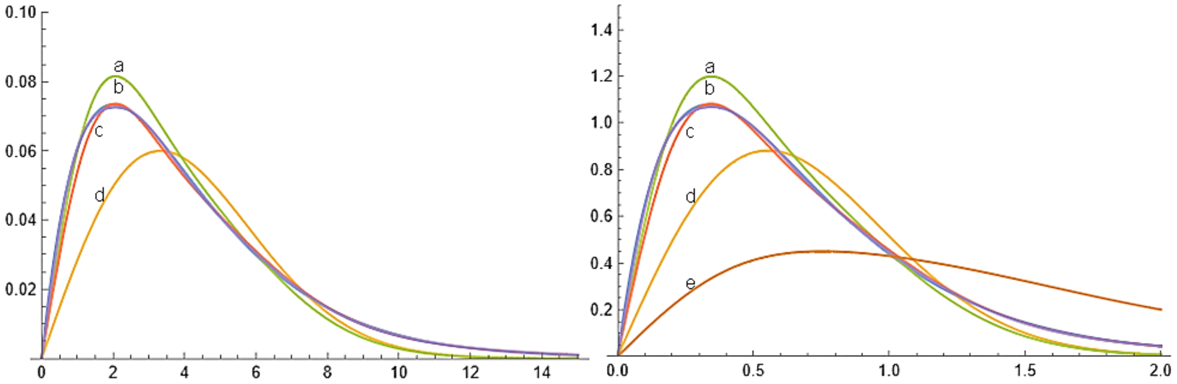

Figure 5. Same as Figure 3, but for orbital = $2p_u$, hydrogen (left, Z = 1) and carbon (right, Z = 6), r is the horizontal axis. Maximum values in decreasing order: a: STO-2G, the {b: ASTO (accurate, blue), c: STO-3G, c: STO-6G} run almost together, and d: STO-1G for hydrogen and carbon. The e: STO-3G (Gaussian94, 2P) for Carbon is also plotted; it has the smallest maximum value (weak shape).

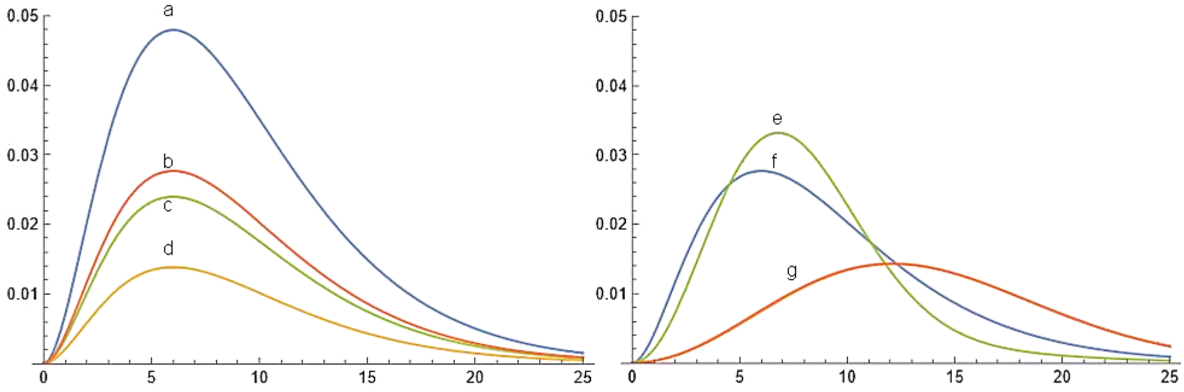

Figure 6. Demonstration of the problem of $3d_{u^2}$ "orbital" for Hydrogen (Z = 1) as a basis function: The common $R_{32}(r) = Z^{7/2} c_i r^2 \exp(-Z r/3)$ part of ASTO(3d) from Tables 1 and 2 on the left and the f(r) part of STO-kG($3d_{u^2}$) with k = 1 and 2 from Tables 5 and 6 on the right, r is the horizontal axis. The $<Y_{lm}R_{nl}(\mathbf{r})|Y_{lm}R_{nl}(\mathbf{r})> = <Y_{lm}f(\mathbf{r})|Y_{lm}f(\mathbf{r})> = 1$ holds for all norms. The $R_{32}(r)$ part of STO($3d_{u^2}$) is plotted on the left and on the right as well; see these two maximums at ≈0.03. Maximums in decreasing order on the left are a: uv, b: u^2, c: u^2-v^2, and d: $3u^2$-r^2. These are the R_{32} parts of accurate wave functions, except the one belonging to $3d_{u^2}$, which is a technical basis function only, although its graph is well located among the others. On the right, the largest maximum (curve e) is the f part of STO-2G($3d_{u^2}$) with good shape (i.e., f follows the R_{32} (middle curve (curve f), blue)), but with inaccurate orbital energy of -0.0792301 hartree (ee% = -42.6), see Table 6. On the right, the smallest maximum (curve g) is the f part of STO-1G($3d_{u^2}$) and STO-2G($3d_{u^2}$) running together (indistinguishable by eye) with weak shape (*i.e.*, these f do not follow the R_{32}), but excellent orbital energy of -0.0553608 and -0.0555554 hartree (ee% = 0.4 and 0.00028, respectively), see Tables 5 and 6.

To exhibit that not only the atomic number (Z) unification is fulfilled by our algebraic forms but the degeneracy also, see, e.g., Figure 8 for how the 4f orbitals are simultaneously recovered.

6. STO-kG Basis Functions from 'Fake' Orbitals Have Good Shape, Bad Energy or Bad Shape, Good Energy

For 3d orbitals, the use of $Y_{lm} = 2z^2$-x^2-y^2, x^2-y^2 and xy (or the cyclic permutations) leads to Equation 3, but the single x^2, y^2, and z^2 do not; (fake in Tables 1 and 2). The $3d_{u^2}$ cannot produce accurate energy, *i.e.*, $<3d_{u^2}|[h]|3d_{u^2}>/<3d_{u^2}|3d_{u^2}> \neq -Z^2/18$ in Table 2, wherein the constant c_1 drops anyway, and the difference is huge: $(-1/18 + 13/162)Z^2$ hartree = 15.49 Z^2 kcal/mol. It comes from the fact that it is not an accurate wave function, i.e., $[h](3d_{x^2}) \neq 3d_{x^2}$, but only a technical basis func-

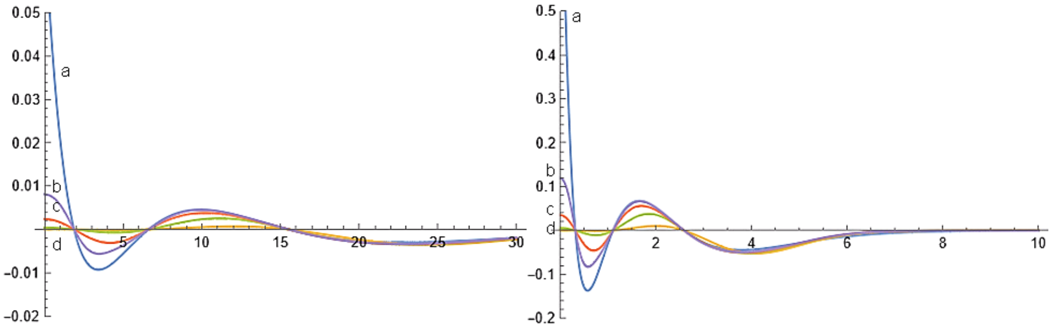

Figure 7. Same as in Figure 3, but for orbital = 4s: Hydrogen (left, Z = 1) and carbon (right, Z = 6), r is the horizontal axis. Intersections of the curves on the vertical axis in decreasing order: a: ASTO (accurate), b: STO-6G, c: STO-3G, d: STO-2G ≈ STO-1G in both plots.

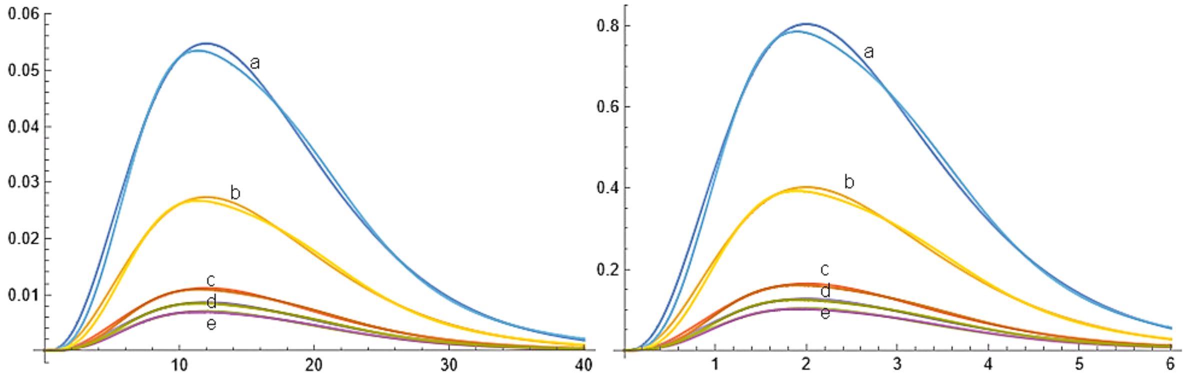

Figure 8. Same as in Figure 3, but for the five nonfake algebraic 4f orbitals listed in Table 7; the ASTO and its STO-3G approximation are plotted for the five cases for Hydrogen (left, Z = 1) and carbon (right, Z = 6), r is the horizontal axis. The curves in both plots: a: 4f(uvw), b: 4f(u(v^2-w^2)), c: 4f(u(3v^2-u^2)), d: 4f(u(5v^2-r^2)), e: 4f(u(5u^2-3r^2)); see c_1 in Table 2. The seven physical 4f orbitals are mixed of these; see the caption of Table 1. In all case pairs, the maximum of STO-3G is slightly under (in maximum) the ASTO in both plots. The unification, such as the Z is parameter, is manifesting, so the excellent shape quality, that is, how the ASTO and STO-3G run together, as well as their very low ee% = 0.2 in Table 7.

tion in calculations for molecules with N >> 1 in practice. We also emphasize that it can still be used as a basis function, even though it is not perfect. Other basis functions can help (correct) to it.

Equations 13 and14 can be applied to exited states because the 'variation-like' property holds if 'plausible' approximate wave functions, $Y_{lm}f$, are used (Table 1 except $3d_{u^2}$). For $3d_{u^2}$ the "variation-like" property does not hold; it yields too deep energy (<< $-Z^2/18$); in fact, in this case, the variation principle works by tending to the ground state 1s orbital with a bad trial function for 1s. Yet, to create basis functions using algebraic forms that are far from the true wave functions (as they are also in some cases in Equation 6), for an optimal STO-kG, one can fit it to the accurate energy with the Lagrange multiplier as (Equation 16)

$$L = (\langle Y_{lm}f\,|[h]Y_{lm}f\rangle - (-Z^2/(2n^2)))^2 + \lambda(\langle Y_{lm}f\,|\,Y_{lm}f\rangle - 1) \qquad (16)$$

The energy term <|[h]> is second order in A_i in Equation 13, while it is fourth order in Equation 16 making the fit more complicated, but one can manage. Equation 16 has benefits for n > 1, for n = 1 (1s ground state orbital), it falls into the variation principle. If plausible approximate excited wave functions, $Y_{lm}f$, are used in Equation 16, *e.g.*, STO-kG($3d_{u^2-v^2}$) with parameters to optimize, it yields the same optimum as Equations 13 and 14, however, it is more wasteful algebraically to solve a fourth-order problem for A_i than the simpler second-order one in Equations 13 and14.

If no 'plausible' approximate wave function is used, Equation 16 produces a very good energy but a weak shape; see e.g., the STO-1G($3d_{u^2}$) with its ee%= 0.4 in Table 5 and its weak shape (that is, it does not follow exp(-Zr/3)) in Figure 6. However, when STO-kG basis functions have plausible form, both their shapes and energies values are very good; see Tables 5-8 and Figure 6. Notice that, all P of R_{32} and all P_i of f in the case of 3d orbitals are exactly the same type of polynomials, $r^2\exp(-Zr/3)$ and $r^2\exp(-a_i(Zr/3)^2)$, respectively (see Table 1); the unproper $Y_{lm} = u^2/r^2$ makes the discrepancy in $3d_{u^2}$. Equation 13 for STO-1G($3d_{u^2}$) yields $A_1 = 0.00552615$ and $a_1 = 0.347165$, $E_{3,\mathrm{approx}} = -0.0835767\ Z^2$ (minimum), and ee% = -50.4 % very weak shape (not plotted) and very weak energy value, i.e., for this fake orbital form, Equations 13-14 are false algorithms. For STO-2G($3d_{u^2}$), the a_2 value was scanned between 0.001-5, and Equation 16 was solved using Equation 15 listed in Table 6. In summary, as Table 6 and Figure 6 reveal, one can obtain good-shape bad orbital energy and weak-shape accurate orbital energy in these cases, like $3d_{u^2}$, wherein the algebraic form for STO-kG is not plausible. As mentioned above, these latter basis functions are not good wave function approximations for Equation 1, but they can still be useful (even if they are not perfect) basis functions. The case of $3d_{u^2}$ is a manifestation of the possible problem in the 'radial fit' generally, when one fits to the shape of exp(-Zr/n) only, as the experience in literature suggests, as opposed to what we do here, i.e., to fit to the full shape of $R_{nl}Y_{ml}$ and accurate orbital energy. A further property of Equation 16 when one uses a not plausible approximate wave function is that, for any given a_i value (see ee%= 0.4 and 0.00028 values in Tables 5 and 6 for $3d_{u^2}$), Equation 15 always fixes the A_i coefficients, such as yielding totally accurate orbital energy. In the case of STO-2G($3d_{u^2}$), the "radial fit" has yielded a_2= 0.2 (that is, during the a_2 scan with Equation 16 this a_2 has yielded the best graph), and this a_2 value was used to recover (or fix) the orbital energy in the fit for A_1 and A_2 listed in Table 6. For these reasons, the STO-kG($3d_{u^2}$) with k=2 in Table 6 is

Table 5. STO-1G basis functions mimic Equations 2 and 3, calculated by the 'energy fit' in Equations 13 and 14, $E_{n,\text{approx}}$ = <STO-1G|[h]|STO-1G> with normalization <STO-1G|STO-1G> = 1. For u and v, see Table 1, the $g_1(\mathbf{r}) \equiv A_1 \exp(-a_1(Z\,r/n)^2)$, analytic values are listed for A_1 and a_1. The ee% (Equation 4) values are direct and immediate measures of the quality of fit here and in Tables 6-8.

Orbital n	**$R_{nl}(\mathbf{r})Y_{lm}(\mathbf{r})$ by STO-1G(orbital) = $Y_{lm}f$**	**A_1**	**a_1**	**$E_{n,\text{approx.}}/Z^2$**	**ee%**
1s	$Z^{3/2} g_1(\mathbf{r})$	$8/(3\sqrt{3}\pi^{3/2}) = 0.276492$	$8/(9\pi) = 0.282942$	$-4/(3\pi) = -0.424413$	15.1
2s	$Z^{3/2}(1-0.25(Zr)^2)\, g_1(\mathbf{r})$	0.0175679	0.19571	-0.1097	12.2
2p (u)	$u\, Z^{5/2} g_1(\mathbf{r})$	$256/(225\sqrt{15}\,\pi^2) = 0.0297654$	$128/(225\pi) = 0.181083$	$-16/(45\pi) = -0.113177$	9.5
3s	$Z^{3/2}(1 - (8/27)(Z\,r)^2 + (4/729)(Z\,r)^4)\, g_1(\mathbf{r})$	0.000905736	0.132232	-0.0508649	8.4
3p (u)	$u\, Z^{5/2}(1 - (1/36)(Z\,r)^2)\, g_1(\mathbf{r})$	0.00300656	0.123232	-0.051777	6.8
3d (uv)	$u\, v\, Z^{7/2} g_1(\mathbf{r})$	$262144(2/105)^{1/2}/(1157625\,\pi^{5/2}) = 0.00178656$	$512/(1225\pi) = 0.133041$	$-256/(1575\pi) = -0.051738$	6.9
3d ($3u^2-r^2$)	$(3u^2-r^2)\, Z^{7/2} g_1(\mathbf{r})$	$131072\,(2/35)^{1/2}/(3472875\,\pi^{5/2}) = 0.000515736$	0.133041	-0.051738	6.9
3d (u^2-v^2)	$(u^2-v^2)\, Z^{7/2} g_1(\mathbf{r})$	$131072\,(2/105)^{1/2}/(1157625\,\pi^{5/2}) = 0.00089328$	0.133041	-0.051738	6.9
3d (u^2)	$u^2 Z^{7/2} g_1(\mathbf{r})$	0.000263047	0.0609375	-0.0553608	0.4
4s	$Z^{3/2}(36863.6 - 11519.9\,(Z\,r)^2 + 287.999\,(Z\,r)^4 - (Z\,r)^6)\, g_1(\mathbf{r})$	1.27236×10^{-9}	0.100019	-0.0293057	6.2
4p (u)	$u\, Z^{5/2}(6400 - 240\,(Z\,r)^2 + (Z\,r)^4)\, g_1(\mathbf{r})$	3.84582×10^{-8}	0.0937500	-0.0296129	5.2
4d (uv)	$u\, v\, Z^{7/2}(1 - Z^2r^2/144)\, g_1(\mathbf{r})$	0.000223125	0.09296875	-0.0297499	4.8
4d ($3u^2-r^2$)	$(3u^2-r^2)\, Z^{7/2}(1 - Z^2r^2/144)\, g_1(\mathbf{r})$	0.0000675809	0.0943750	-0.0297449	4.8
4d (u^2-v^2)	$(u^2-v^2)\, Z^{7/2}(1 - Z^2r^2/144)\, g_1(\mathbf{r})$	0.000109775	0.09250	-0.0297504	4.8
4f ($u(3v^2-u^2)$)	$u(3v^2-u^2)\, Z^{9/2} g_1(\mathbf{r})$	$134217728/(29536801875\,(105^{1/2})\,\pi^3) = 0.0000143022$	$32768/(99225\pi) = 0.105118$	$-1024/(11025\pi) = -0.0295646$	5.4
4f (uvw)	$uvw\, Z^{9/2} g_1(\mathbf{r})$	$268435456\,(2/35)^{1/2}/(29536801875\,\pi^3) = 0.0000700661$	0.105118	-0.0295646	5.4
4f ($u(v^2-w^2)$)	$u(v^2-w^2)\, Z^{9/2} g_1(\mathbf{r})$	$134217728\,(2/35)^{1/2}/(29536801875\,\pi^3) = 0.000035033$	0.105118	-0.0295646	5.4
4f ($u(5u^2-3r^2)$)	$u(5u^2-3r^2)\, Z^{9/2} g_1(\mathbf{r})$	$134217728\,(2/21)^{1/2}/(147684009375\,\pi^3) = 9.04549\times10^{-6}$	0.105118	-0.0295646	5.4
4f ($u(5v^2-r^2)$)	$u(5v^2-r^2)\, Z^{9/2} g_1(\mathbf{r})$	$134217728/(147684009375\times7^{1/2}\,\pi^3) = 0.0000110784$	0.105118	-0.0295646	5.4

suggested for k= 3 and 6 as well. Similar arguments hold for 4d and 4f 'fake' orbitals listed in Table 1; that is, good shape and good energy cannot be modeled at the same time.

7. The Coefficients of STO-kG Basis Functions from Energy Fits

Normalized $Y_{lm}f$ with energy $E_{n,\text{approx.}} \equiv L$ from Equations 13 and 14 approximates the accurate orbital energy (Equation 3) and mimics the shape (Equation 2) very well; the parameters are listed in Tables 5-8. These are the suggested STO-kG basis functions of this work to use in molecular structure calculations for basis sets. The energy error, ee%, in Equation 4 judges the orbital energy quality of the fit. As expected, the ee% decreases as k increases in the STO-kG basis functions. The radial distributions are plotted on Figures 2-8 for Z = 1 (Hydrogen) and Z = 6 (Carbon) to see how the STO-kG develops in following the graph of the accurate ASTO in Equation 2.

The minimal STO-1G basis functions are obviously poor for molecular calculations, as their ee% values reveal in Table 5. This manifests itself in the cases of 1s and 2s with large ee% = 15.1 and 12.2, respectively. However, for example, for LiH, the {STO-1G(H,1s), STO-1G(Li,1s), STO-1G(Li,2s)} minimal basis set can still be improved if all or at least a few excited state basis functions from Table 5 are used, extending it to a larger, augmented STO-1G set. Using the STO-kG basis set with a smaller k needs at least less computation time. Also recall that in computational quantum chemistry, the energy differences are most important between configurations or conformations, wherein the errors from two calculations at the same level can fortunately cancel each other out to some degree. From the experience in the literature, based on the strategy used for basis set creation, STO-3G or STO-6G level basis functions have to be used for basis sets at least (*i.e.*, k ≥ 3) for acceptable accuracy. Tables 7-8 reveal that the ee% values of our STO-3G and STO-6G basis functions are remarkably low.

As we have emphasized above, the ee% values drastically decrease in Tables 5-8 as k increases in STO-kG, an expected trend. In comparison to the literature, Table 4 reveals that our STO-kG basis functions from the m>2 level exhibit much better ee% values than Gaussian94. We call attention to the fact that, for example, the STO-6G(3d) does not improve much over the STO-3G(3d), as indicated by their ee% values in Tables 7 and 8; their graphs (not exhibited) are almost indistinguishable. Particularly, the fact that the $3d_{u^{\wedge}2}$ "wave function" is a fake wave function, only a technical basis function, is indicated by the problems in its fit shown in Table 6, detailed above.

The fit for 2s and 3s is good, but not as good as the others; their ee% values are always significantly larger than the others; see Tables 6-8. Comparing STO-kG(1s or 2s) to STO-kG(2p) in Figures 2-5, the fit is better if $R_{nl}(r = 0) = 0$ along with $f(r = 0) = 0$ than the cases when $R_{nl}(r = 0) > 0$ along with $f(r = 0) > 0$ (actually when $R_{nl}(r=0) > f(r=0) > 0$), which holds for other orbitals also in our work and literature cases as well; see ee% in Tables 5-8. That is, when nodes are at r=0. It simply comes from that exp(-r) and exp(-a r^2) asymptotes only at infinity at a general "a" value, while r*exp(-r) and r*exp(-a r^2) intersect at r=0 as a constraint at any "a" value besides their asymptotic property.

8. Virial Theorem, Orthonormality, Molecular Integrals, and Recurrence Formulae for STO-kG Basis Functions

The virial theorem well holds approximately in Tables 5-8; we comment with an example only: For STO-1G($4f_{uvw}$), the <potential>/<kinetic> = -2.00001, and for STO-6G($4f_{uvw}$), it is -2.01132 in comparison to the theoretical -2 value.

Two parametric analytic evaluations of the overlap integrals, $\langle Y_{lm}f|Y_{lm}f\rangle$, in Equations 13 and 14 are exhibited as examples. These kinds of equations have come up in the way of filling up Tables 3-8 for orthonormality. Notice that the atomic number, Z, cancels, making a unification for our STO-kG basis functions: (Equations 17 and 18).

Table 6. STO-2G basis functions, a_1 is from Table 5, and the other parameters are calculated by the ‘energy fit’ in Equations 13 and 14, $E_{n,\mathrm{approx}}$ = <STO-2G |[h]| STO-2G> and <STO-2G | STO-2G> = 1. For u and v, see Table 1; numerical values are listed for A_i and a_i, here and in Tables 7 and 8 also, the $g_2(\mathbf{r}) \equiv A_1\exp(-a_1(Z\,r/n)^2) + A_2\exp(-a_2(Z\,r/n)^2)$.

Orbital n	**$R_{nl}(\mathbf{r})Y_{lm}(\mathbf{r})$ by STO-2G(orbital) = $Y_{lm}f$**	**A_1**	**A_2**	**a_2**	**$E_{n,\mathrm{approx.}}/Z^2$**	**ee%**
1s	$Z^{3/2}\, g_2(\mathbf{r})$	0.245808	0.221228	2.0	-0.478896	4.2
2s	$Z^{3/2}\,(1 - 0.25\,(Z\,r)^2)\, g_2(\mathbf{r})$	0.0156455	0.035623	0.65	-0.117284	6.2
2p (u)	$u\, Z^{5/2}\, g_2(\mathbf{r})$	0.0256471	0.043082	0.8	-0.121607	2.7
3s	$Z^{3/2}\,(1 - (8/27)(Z\,r)^2 + (4/729)(Z\,r)^4)\, g_2(\mathbf{r})$	0.000822294	0.00462247	0.367531	-0.0527318	5.1
3p (u)	$u\, Z^{5/2}\,(1 - (1/36)(Z\,r)^2)\, g_2(\mathbf{r})$	0.00267813	0.00735288	0.332031	-0.0533047	4.1
3d (uv)	$u\, v\, Z^{7/2}\, g_2(\mathbf{r})$	0.00157092	0.00374698	0.495	-0.0544496	2.0
3d ($3u^2$-r^2)	$(3u^2-r^2)\, Z^{7/2}\, g_2(\mathbf{r})$	0.000453483	0.00108166	0.495	-0.0544496	2.0
3d (u^2-v^2)	$(u^2-v^2)\, Z^{7/2}\, g_2(\mathbf{r})$	0.000785461	0.00187349	0.495	-0.0544497	2.0
3d (u^2)	$u^2\, Z^{7/2}\, g_2(\mathbf{r})$, good shape, inaccurate energy	0.000035238	0.00193495	0.2	-0.0792301	-42.6
3d (u^2)	$u^2\, Z^{7/2}\, g_2(\mathbf{r})$ weak shape, accurate energy	0.000262193	0.0000122121	0.2	-0.0555554	0.00028
4s	$Z^{3/2}\,(36863.6 - 11519.9\,(Z\,r)^2 + 287.999\,(Z\,r)^4 - (Z\,r)^6)\, g_2(\mathbf{r})$	1.13777×10^{-9}	1.00751×10^{-8}	0.225	-0.0300204	3.9
4p (u)	$u\, Z^{5/2}\,(6400 - 240\,(Z\,r)^2 + (Z\,r)^4)\, g_2(\mathbf{r})$	3.47114×10^{-8}	2.88726×10^{-7}	0.240625	-0.0302237	3.3
4d (uv)	$u\, v\, Z^{7/2}(1 - Z^2r^2/144)\, g_2(\mathbf{r})$	0.000202482	0.000896882	0.2515625	-0.0302615	3.2
4d ($3u^2$-r^2)	$(3u^2-r^2)Z^{7/2}(1 - Z^2r^2/144)\, g_2(\mathbf{r})$	0.0000612838	0.000505374	0.321875	-0.0302334	3.3
4d (u^2-v^2)	$(u^2-v^2)\, Z^{7/2}(1 - Z^2r^2/144)\, g_2(\mathbf{r})$	0.0000992811	0.000417873	0.24420	-0.0302746	3.1
4f ($u(3v^2$-$u^2)$)	$u(3v^2-u^2)\, Z^{9/2}\, g_2(\mathbf{r})$	0.0000125143	0.0000415555	0.3328	-0.0307449	1.6
4f (uvw)	$uvw\, Z^{9/2}\, g_2(\mathbf{r})$	0.0000613169	0.000203755	0.3330	-0.0307449	1.6
4f ($u(v^2$- $w^2)$)	$u(v^2-w^2)\, Z^{9/2}\, g_2(\mathbf{r})$	0.0000306584	0.000101878	0.3330	-0.0307448	1.6
4f ($u(5u^2$-$3r^2)$)	$u(5u^2-3r^2)\, Z^{9/2}\, g_2(\mathbf{r})$	7.91597×10^{-6}	0.0000263047	0.3330	-0.0307449	1.6
4f ($u(5v^2$-$r^2)$)	$u(5v^2-r^2)\, Z^{9/2}\, g_2(\mathbf{r})$	9.69513×10^{-6}	0.000032215	0.3330	-0.0307449	1.6

Table 7. STO-3G basis functions, the a_1 and the a_2 are from Tables 5 and 6, and the other parameters are calculated by the ‘energy fit’ in Equations 13 and 14, $E_{n,\mathrm{approx}}$ = <STO-3G |[h]| STO-3G> and <STO-3G | STO-3G> = 1. For u and v, see Table 1, the $g_3(\mathbf{r}) \equiv A_1\exp(-a_1(Z\,r/n)^2) + A_2\exp(-a_2(Z\,r/n)^2) + A_3\exp(-a_3(Z\,r/n)^2)$.

Orbital n	**$R_{nl}(\mathbf{r})Y_{lm}(\mathbf{r})$ by STO-3G(orbital) = $Y_{lm}f$**	**A_1**	**A_2**	**A_3**	**a_3 and ($E_{n,\mathrm{approx.}}/Z^2$)**	**ee%**
1s	$Z^{3/2}\, g_3(\mathbf{r})$	0.19124	0.220093	0.0272301	0.08 (-0.491739)	1.7
2s	$Z^{3/2}\,(1 - 0.25(Z\,r)^2)\, g_3(\mathbf{r})$	0.00198175	0.0435985	0.00852152	0.153 (-0.119586)	4.3
2p (u)	$u\, Z^{5/2}\, g_3(\mathbf{r})$	0.0208266	0.0394248	0.00177026	0.058 (-0.124256)	0.6
3s	$Z^{3/2}\,(1 - (8/27)(Z\,r)^2 + (4/729)(Z\,r)^4)\, g_3(\mathbf{r})$	0.000817618	0.00415247	0.0038832	0.626232 (-0.0527932)	4.9
3p (u)	$u\, Z^{5/2}\,(1 - (1/36)(Z\,r)^2) g_3(\mathbf{r})$	0.00249612	0.00956062	0.0218383	1.700 (-0.0543716)	2.1
3d (uv)	$u\, v\, Z^{7/2}\, g_3(\mathbf{r})$	0.00127526	0.00362288	0.0000784273	0.05 (-0.0553459)	0.4
3d ($3u^2$-r^2)	$(3u^2-r^2)\, Z^{7/2}\, g_3(\mathbf{r})$	0.000368135	0.00104584	0.00002264	0.05 (-0.0553458)	0.4
3d (u^2-v^2)	$(u^2-v^2)\, Z^{7/2}\, g_3(\mathbf{r})$	0.000637628	0.00181144	0.0000392136	0.05 (-0.0553457)	0.4
4s	$Z^{3/2}\,(36863.6 - 11519.9\,(Z\,r)^2 + 287.999\,(Z\,r)^4 - (Z\,r)^6)\, g_3(\mathbf{r})$	1.08537×10^{-9}	1.16265×10^{-8}	5.23634×10^{-8}	0.5625 (-0.0303107)	3.0
4p (u)	$u\, Z^{5/2}\,(6400 - 240\,(Z\,r)^2 + (Z\,r)^4)\, g_3(\mathbf{r})$	3.364×10^{-8}	3.16434×10^{-7}	7.21907×10^{-7}	0.703125 (-0.0304139)	2.7
4d (uv)	$u\, v\, Z^{7/2}(1 - Z^2r^2/144)\, g_3(\mathbf{r})$	0.000182661	0.00128575	0.00287891	0.93750 (-0.0308151)	1.4
4d ($3u^2$-r^2)	$(3u^2-r^2)Z^{7/2}(1 - Z^2r^2/144)\, g_3(\mathbf{r})$	0.0000545414	0.000738047	0.00105261	1.4320 (-0.0307423)	1.6
4d (u^2-v^2)	$(u^2-v^2)\, Z^{7/2}(1 - Z^2r^2/144)\, g_3(\mathbf{r})$	0.0000896803	0.000592578	0.00140787	0.90700 (-0.030823)	1.4
4f ($u(3v^2$-$u^2)$)	$u(3v^2-u^2)\, Z^{9/2}\, g_3(\mathbf{r})$	0.0000103102	0.0000398932	4.35202×10^{-7}	0.0420 (-0.0311751)	0.2
4f (uvw)	$uvw\, Z^{9/2}\, g_3(\mathbf{r})$	0.0000504469	0.000195675	2.15978×10^{-6}	0.0422 (-0.0311747)	0.2
4f ($u(v^2$-$w^2)$)	$u(v^2-w^2)\, Z^{9/2}\, g_3(\mathbf{r})$	0.0000252141	0.0000978487	1.08344×10^{-6}	0.04225 (-0.0311746)	0.2
4f ($u(5u^2$-$3r^2)$)	$u(5u^2-3r^2)\, Z^{9/2}\, g_3(\mathbf{r})$	6.50786×10^{-6}	0.0000252673	2.8066×10^{-7}	0.0423 (-0.0311747)	0.2
4f ($u(5v^2$-$r^2)$)	$u(5v^2-r^2)\, Z^{9/2}\, g_3(\mathbf{r})$	7.9755×10^{-6}	0.0000309659	3.41429×10^{-7}	0.0422 (-0.0311747)	0.2

$$\langle \text{STO-2G(1s)}|\ \text{STO-2G(1s)}\rangle = \langle Z^{3/2} g_2(\mathbf{r})|Z^{3/2} g_2(\mathbf{r})\rangle$$
$$= = \frac{1}{4}\left(\frac{\sqrt{2}\mathrm{A1}^2}{\mathrm{a1}^{3/2}} + \frac{8\mathrm{A1A2}}{(\mathrm{a1+a2})^{3/2}} + \frac{\sqrt{2}\mathrm{A2}^2}{\mathrm{a2}^{3/2}}\right)\pi^{3/2} \tag{17}$$

<STO-2G(1s)| STO-2G(2p$_x$)> = $\langle Z^{3/2} g_2(\mathbf{r})|x\ Z^{5/2} g_2'(\mathbf{r})\rangle = 0$ by the even/odd symmetry relation (18)

In the optimization by Equations 13 and 14, Equation 17 and the other corresponding ones have always been set to 1 as the norm. Equation 18 comes from elementary math. In relation to Equation 18 (orthogonality), for example, the <STO-2G(1s)|STO-2G(2s)> is not zero (both functions are even), only close to zero with our optimized parameters, but Z-independent via Table 1. See the right side of Table 3.

Among the energy integrals, $\langle Y_{lm}f|[h]Y_{lm}f\rangle$, used in optimization via Equations 13 and14, the analytic evaluation of a simple one is exhibited. In these energy integrals, the same one-electron wave function appears only in the braket. Notice that, the atomic number, Z, comes out as Z^2 in accordance with Equations 1-3, as well as recall Equation 12 how the $g_1(\mathbf{r})$ deconvolutes to $g_2(\mathbf{r})$ as (Equation 19)

$$\langle \text{STO-2G(1s)}|[h]\text{STO-2G(1s)}\rangle = \frac{3}{8}\left(\frac{\sqrt{2}\mathrm{A1}^2}{\sqrt{\mathrm{a1}}} + \frac{16\mathrm{a1A1a2A2}}{(\mathrm{a1+a2})^{5/2}} + \frac{\sqrt{2}\mathrm{A2}^2}{\sqrt{\mathrm{a2}}}\right)\pi^{3/2}Z^2 - \left(\frac{\mathrm{A1}^2}{\mathrm{a1}} + \mathrm{A2}\left(\frac{4\mathrm{A1}}{\mathrm{a1+a2}} + \frac{\mathrm{A2}}{\mathrm{a2}}\right)\right)\pi Z^2 \tag{19}$$

Table 8. STO-6G basis functions, the a_k from Tables 5-7 for k = 1, 2, 3, as well as the $a_4 = 1.0$, $a_5 = 3.0$, and $a_6 = 5.0$, have been a priori chosen via the experience from Tables 5-7. The set $\{\exp(-a_i(Z\ r/n)^2)\}$ with different $a_i>0$ values provides a linearly independent mathematical basis set for the fit, and the A_i parameters with i=1-6 have been optimized by the "energy fit" in Equations 13 and 14, $E_{n,approx}$ = <STO-6G|[h]|STO-6G> and <STO-6G|STO-6G> = 1. For u and v, see Table 1. In the optimization of a_i parameters, the expression is more and more crowded with reciprocal powers of a_i as k increases. It has been bypassed with these a priori values, and the fit has been done for the A_i parameters only. The parameters A_i are linear in the expressions of STO-kG, here as $g_6(\mathbf{r}) \equiv \Sigma_{i=1}{}^6 A_i \exp(-a_i(Z\ r/n)^2)$. It yields only a third-order algebraic system via Equation 13 (linear combination of A_iA_j and λA_iA_j) and only a second-order system via Equation 15. The latter can be solved easily by the standard algorithm [9] for parameter set $\{A_i\}$.

Orbital n	**$R_{nl}(\mathbf{r})Y_{lm}(\mathbf{r})$ by STO-6G(orbital) = $Y_{lm}f$**	**A_1, A_2**	**A_3, A_4**	**A_5, A_6**	**$E_{n,approx.}/Z^2$**	**ee%**
1s	$Z^{3/2} g_6(\mathbf{r})$	0.161247, 0.176272	0.0310744, 0.117247	-0.212035, 0.218016	-0.498513	0.3
2s	$Z^{3/2} (1 - 0.25(Z\ r)^2)\ g_6(\mathbf{r})$	0.00310493, 0.0574458	0.00746189, -0.0206425	-0.0173656, 0.0757546	-0.12218	2.3
2p (u)	$u\ Z^{5/2}\ g_6(\mathbf{r})$	0.0200731, 0.0736876	0.00181671, -0.0461477	0.0348738, -0.00414764	-0.124795	0.2
3s	$Z^{3/2} (1 - (8/27)(Z\ r)^2 + (4/729)(Z\ r)^4)\ g_6(\mathbf{r})$	0.000773401, 0.00730016	-0.0125596, 0.0253444	-0.0147536, 0.0296611	-0.0538164	3.1
3p (u)	$u\ Z^{5/2} (1 - (1/36)(Z\ r)^2) g_6(\mathbf{r})$	0.00242128, 0.00930521	-0.00399437, 0.0214922	0.00725298, 0.0150324	-0.0545467	1.8
3d (uv)	$u\ v\ Z^{7/2}\ g_6(\mathbf{r})$	0.0012682, 0.00383534	0.0000784417, -0.000830951	0.00667725, -0.00477721	-0.0554049	0.3
3d ($3u^2-r^2$)	$(3u^2-r^2)\ Z^{7/2}\ g_6(\mathbf{r})$	0.000366094, 0.00110714	0.0000226451, -0.000239816	0.00192733, -0.00137878	-0.0554047	0.3
3d (u^2-v^2)	$(u^2-v^2)\ Z^{7/2}\ g_6(\mathbf{r})$	0.000634093, 0.00191768	0.0000392218, -0.000415467	0.00333853, -0.00238848	-0.0554046	0.3
4s	$Z^{3/2} (36863.6 - 11519.9\ (Z\ r)^2 + 287.999\ (Z\ r)^4 - (Z\ r)^6)\ g_6(\mathbf{r})$	1.04423×10^{-9}, 1.27882×10^{-8}	7.5469×10^{-8}, -1.04589×10^{-8}	3.93252×10^{-7}, -2.52255×10^{-7}	-0.0304962	2.4
4p (u)	$u\ Z^{5/2} (6400 - 240\ (Zr)^2 + (Z\ r)^4)\ g_6(\mathbf{r})$	3.08537×10^{-8}, 3.97682×10^{-7}	1.21439×10^{-6}, 2.92523×10^{-7}	1.16578×10^{-6}, 1.46754×10^{-6}	-0.0307763	1.5
4d (uv)	$u\ v\ Z^{7/2}(1 - Z^2r^2/144)\ g_6(\mathbf{r})$	0.000180438, 0.00129434	0.011037, -0.00884885	0.00284549, -0.000213745	-0.0308563	1.3
4d ($3u^2-r^2$)	$(3u^2-r^2)Z^{7/2}(1 - Z^2r^2/144)\ g_6(\mathbf{r})$	0.0000543559, 0.00074156	0.00110533, 0.0000153738	-0.000491691, 0.00102955	-0.0307558	1.6
4d (u^2-v^2)	$(u^2-v^2)\ Z^{7/2}(1 - Z^2r^2/144)\ g_6(\mathbf{r})$	0.0000885443, 0.000597203	0.00406627, -0.00299872	0.0014952, -0.000200734	-0.0308655	1.2
4f ($u(3v^2-u^2)$)	$u(3v^2-u^2)\ Z^{9/2}\ g_6(\mathbf{r})$	0.0000103468, 0.0000391198	4.31517×10^{-7}, 0.0000160761	0.0000829852, -0.0000846397	-0.0311871	0.2
4f (uvw)	$uvw\ Z^{9/2}\ g_6(\mathbf{r})$	0.0000506259, 0.000191925	2.14151×10^{-6}, 0.0000778597	0.000409787, -0.000419304	-0.0311870	0.2
4f ($u(v^2-w^2)$)	$u(v^2-w^2)\ Z^{9/2}\ g_6(\mathbf{r})$	0.0000253034, 0.0000959757	1.07432×10^{-6}, 0.0000388845	0.000205072, -0.000209916	-0.0311870	0.2
4f ($u(5u^2-3r^2)$)	$u(5u^2-3r^2)\ Z^{9/2}\ g_6(\mathbf{r})$	6.53094×10^{-6}, 0.0000247842	2.78292×10^{-7}, 0.0000100297	0.0000529889, -0.0000542571	-0.0311869	0.2
4f ($u(5v^2-r^2)$)	$u(5v^2-r^2)\ Z^{9/2}\ g_6(\mathbf{r})$	8.00455×10^{-6}, 0.000030346	3.38622×10^{-7}, 0.0000123095	0.0000647982, -0.0000663061	-0.0311869	0.2

Generally, the Gaussian product theorem, $\exp(-a|\mathbf{r}-\mathbf{A}|^2)\exp(-b|\mathbf{r}-\mathbf{B}|^2) = \exp(-ab|\mathbf{A}-\mathbf{B}|^2/c)\exp(-c|\mathbf{r}-\mathbf{C}|^2)$ with $\mathbf{C} \equiv (a\mathbf{A}+b\mathbf{B})/(a+b)$ and $c \equiv a+b$, is necessary to use for analytical evaluation of overlap, kinetic, and Coulomb molecular integrals. It reduces it to a single primitive Gaussian, and finally, the $G_a(\mathbf{r},a,\mathbf{A},\mathbf{n}) \equiv (x-A_1)^{n1}(y-A_2)^{n2}(z-A_3)^{n3}\exp(-a|\mathbf{r}-\mathbf{A}|^2)$ basis functions come up, as in Equation 6 and the tables above (Table 5-8). Separation of integral variables yields, e.g., for overlap molecular integrals, the $\langle G_a|G_b\rangle = \exp(-ab|\mathbf{A}-\mathbf{B}|^2/c)\ I_x\ I_y\ I_z$ with elementary 1D integrals $I_u \equiv \int(u-A_1)^{n1}(u-B_1)^{m1}\exp(-c(u-C_1)^2)du$ and u= x, y, z. Recurrence formulae for I_u can be obtained by partial integration; the $(u-C_1)\exp(-c(u-C_1)^2)$ can be integrated, yielding Equations 20 and 21 or the $\exp(-c(u-C_1)^2)$ can be derivated, yielding Equation 22. Finally, the three versions are (Equations 20-22)

$$I_u(n_1,m_1) = (1/(2c))\ (\ n_1\ I_u(n_1-1,m_1-1) + (m_1-1)\ I_u(n_1,m_1-2)\ ) + (C_1-B_1)\ I_u(n_1,m_1-1) \tag{20}$$

$$I_u(n_1,m_1) = (1/(2c))\ (\ (n_1-1)\ I_u(n_1-2,m_1) + m_1\ I_u(n_1-1,m_1-1)\ ) + (C_1-A_1)\ I_u(n_1-1,m_1) \tag{21}$$

$$(A_1 - B_1)\ I_u(n_1,m_1) = I_u(n_1,m_1+1) - I_u(n_1+1,m_1) \tag{22}$$

Furthermore, the SCF convergence part of the HF-SCF method can be replaced by solving the equation of the electronic energy functional with Lagrange multipliers (to keep MOs orthonormal, and which is a 4th order multivariable polynomial of the LCAO coefficients and Lagrange multipliers) directly for all its roots with algebraic geometry device. One determinant [2] approximation of ground and excited states can be calculated for molecular systems.

9. Conclusions

The algebraic formulae in Table 1 have been used in this work to create STO-kG basis sets. These formulae allow: (i). To pull out the atomic number Z, which makes unification for the periodic table. Recall that, in modern practice, all atoms in the periodic table have their own basis sets. (ii). The low-degree polynomial multipliers fix the one-electron atom orbital nodes accurately, which is better than the alternative literature used, wherein the alternating signs in the sum of Gaussians take care of this. These polynomials are functions of radial distance with positive powers as augmentation, beside the Cartesian coordinate-dependent polarization polynomials from Y_{lm}, for example, $x^*Z^{5/2}(1-(1/36)(Zr)^2)g_m(\mathbf{r})$ for $2p_x$ vs. the literature used $x^*g_m(\mathbf{r})$, wherein g_m is a linear combination of Gaussians. This allows for a decrease in the value of k in the STO-kG basis sets while achieving the same final accuracy with shorter computation time. (iii). These positive r-powers allow the analytical evaluation of molecular integrals, a fundamental property of STO-kG basis functions. See Tables 5-8 and the corresponding Figures 2-8.

Our STO-kG basis sets contain our STO-kG basis functions that correspond to the occupied orbitals in neutral ground state atoms plus one or more excited orbitals. We have shown that our 'energy fit' STO-kG basis functions reproduce the Hydrogen-like one-electron atom ground and excited state wave functions and energies better than the 'radial fit' STO-3G basis set; basically, the latter device is how the STO-kG basis sets in literature are created and widely used in today's practice of computational chemistry. In more detail, besides the shape criteria (determined by the Coulombic operator, that is, the function type exp(-r)) that the "radial fit" has, the "energy fit" satisfies one more criteria, the energy criteria (by both operators, Columbic and kinetic). Our detailed fit strategy differs from the literature, and a consequence is that the quality of our STO-3G and STO-6G basis functions is better (see their ee% values); moreover, it allows us to use the smaller STO-3G in practice for reasonable accuracy. In other words, our STO-3G basis functions are better than the literature ones; see Table 4.

Because of the limitation on the length of the article, the tests for molecules, which are the final certification, have not been reported here; they will be reported in a separate work. However, we state in advance that the 'energy fit' yields more effective basis sets in relation to computation speed and accuracy. Tables 5-8 show obvious improvements as k increases in STO-kG which have been experienced in calculations for larger molecules. As a side result, we have found via the "energy fit" for STO-kG(3d) basis functions (also in cases of higher excitations, f, g, *etc.*) that 6 technical basis functions (used in literature) are less effective than the physically more plausible 5 ones detailed in tables (Table 5-8).

Finally, the P(Z,r) polynomial part of our STO-kG model wave function, mentioned first in the Abstract, can be refined for its general form as $P(Z,r) = Z^{k+3/2} p(Z,r)$, where k is the l-quantum number in Equation 2, as well as p(Z,r) is a linear combination of powers of x, y, z, r^2, and $(Zr)^2$; see Tables 5-8. It is compatible with the right side of the literature Equation 6 and the right side of our Equation 9, necessary for analytical molecular integration. The subject of this work can be found as a preprint [10].

Acknowledgements

Thanks to Szeger Hermin for her help in typing the manuscript.

Disclosure statement

Conflict of interest: The author declare that he has no conflict of interest.
Ethical approval: All ethical guidelines have been adhered to.

Abbreviations

AO = Atomic orbital(s); ASTO = accurate solution STO of Equations 1and 2, the $R_{nl}Y_{lm}$ in Tables 1 and 2; CI = Configuration interactions method; DFT = Density Functional Theory method; ee = energy error in % by Equation 4; GTO = Gaussian-type atomic orbitals, $x^{L1}y^{L2}z^{L3}$ exp(-const. r^2); HF-SCF/basis = Hartree-Fock Self Consistent Field method on a basis level; $\mathbf{r} = (x,y,z)$ spatial coordinates, its length is $r = (x^2+y^2+z^2)^{1/2}$; $R_{nl}(\mathbf{r})$ = Hydrogen-like radial wave function from solutions of Equation 1, its STO-kG approximation is $f(\mathbf{r})$; STO = Slater-type atomic orbitals, $x^{L1}y^{L2}z^{L3}$ exp(-const. r), function type coming from solutions of Equation 1, whose accurate value is $R_{nl}Y_{lm}$; s = s orbitals, not to be confused with spin, the latter comes up only in Equation 2; STO-kG = the ASTO solution is approximated with a linear combination of k GTO by 'radial fit' or 'energy fit'; $\langle f|g\rangle = \int(f^*g)ds\,d\mathbf{r}$ or $\int(f^*g)d\mathbf{r}$ according to the interpretation used, that is, integration over full 3D spatial space with or without integration over spin, without spin in this work; Y_{lm} = not-normalized spherical harmonics in their simplest form with Cartesian coordinates $Y_{lm}(x,y,z)$, see Table 1.

Funding

Financial and emotional support for this research from OTKA-K-NINCS-115733 and 119358 is kindly acknowledged.

ORCID and Email

Sandor Kristyan

kristyan.sandor@ttk.mta.hu
kristyan.sandor@ttk.hu
https://orcid.org/0000-0003-4169-2392

References

[1]. Atkins, P. W.; Friedman, R. S. Molecular Quantum Mechanics; 4th ed.; Oxford University Press: London, England, 2004.

[2]. Kristyan, S. Optimizing single Slater determinant for electronic Hamiltonian with Lagrange multipliers and Newton-Raphson methods as an alternative to ground state calculations via Hartree-Fock self consistent field. *In AIP Conference Proceedings*; AIP Publishing, **2019**, 2116, 450030. https://doi.org/10.1063/1.5114497

[3]. Kristyan, S. Evaluation of Coulomb integrals with various energy operators to estimate the correlation energy in electronic structure calculations for molecules. *Eur. J. Chem.* **2023**, *14*, 486–493.

[4]. Stewart, R. F. Small Gaussian expansions of Slater-type orbitals. *J. Chem. Phys.* **1970**, *52*, 431–438.

[5]. Stewart, B.; Hylton, D. J.; Ravi, N. A systematic approach for understanding Slater–Gaussian functions in computational chemistry. *J. Chem. Educ.* **2013**, *90*, 609–612.

[6]. Hehre, W. J.; Stewart, R. F.; Pople, J. A. Self-consistent molecular-orbital methods. I. use of Gaussian expansions of Slater-type atomic orbitals. *J. Chem. Phys.* **1969**, *51*, 2657–2664.

[7]. Pietro, W. J.; Blurock, E. S.; Hout, R. F., Jr; Hehre, W. J.; DeFrees, D. J.; Stewart, R. F. Molecular orbital theory of the properties of inorganic and organometallic compounds. 2. STO-NG basis sets for fourth-row main-group elements. *Inorg. Chem.* **1981**, *20*, 3650–3654.

[8]. Binkley, J. S.; Frisch, M. J.; Defrees, D. J.; Krishnan, R.; Whiteside, R. A.; Schlegel, H. B.; Fluder, E. M.; Pople, J. A.; Gaussian 82, Carnegie-Mellon Quantum Chemistry Publishing Unit, Pittsburgh, PA, 1982, http://www.gaussian.com (accessed November 2, 2025).

[9]. Wolfram Research, Inc., Mathematica, Version 14.3, Champaign, IL (2025). http://www.wolfram.com (accessed May 2, 2025).

[10]. Kristyan, S. Unified formulae for augmented near orthonormalized STO-kG basis sets via atomic orbital energy fit to use in molecular structure calculations. *ChemRxiv* **2025**, https://doi.org/10.26434/chemrxiv-2025-0c8hk