JOURNAL OF
EMERGING INVESTIGATORS

Article

Impact of carbon number and atom number on cc-pVTZ Hartree-Fock Energy and program runtime of alkanes

Michelle Pan[1*], Vaibhav Vaiyakarnam[1*], Allen Li[2*], Larry McMahan[3]
[1] The Quarry Lane School, Dublin, California
[2] Monta Vista High School, Cupertino, California
[3] Aspiring Scholars Directed Research Program, Fremont, California
*These authors contributed equally to this work.

SUMMARY

The ground state energy of a molecule can be used in many aspects of chemistry, such as predicting energies produced or absorbed in reactions and determining the stability of the molecules and their analogs. However, different methods to compute these energies have different accuracies and speeds. We aimed to compare one of these methods, the Python-based Simulations of Chemistry Framework's Hartree-Fock (PySCF) method, with established values provided by the Computational Chemistry Comparison and Benchmark Database (CCCBDB), a reliable peer-reviewed database organized by the United States government. We also sought to assess relationships between the structure of each alkane and the runtime of the PySCF program. Molecular geometries from the CCCBDB were taken for straight-chain alkanes with 1-10 carbon atoms. Using the PySCF Hartree-Fock (HF) method, ground state energies were calculated for these alkanes using triple-zeta (cc-pVTZ) basis sets. These energies were then compared to established HF/cc-pVTZ data energies of the same alkanes in the CCCBDB. We hypothesized that the ground state energy would increase linearly and that the runtime of the program would increase quadratically as both the number of carbon atoms and total atoms increased. The data supported our hypotheses – the ground state energy had a negative linear correlation with the number of carbon atoms and the total number of atoms ($r = -1.000$), while the runtime had a quadratic correlation with the number of atoms ($R^2 = 0.9998$). The PySCF data also agreed with the CCCBDB data, indicating that PySCF is both efficient and accurate as a computational chemistry software, and can be tested in future experiments with larger organic molecules such as pharmaceutical candidates.

INTRODUCTION

Atomic energy is defined as the energy that an atom holds in its nucleus and electrons, including nuclear binding energy and potential energy, and the energy in interatomic bonds (1). Atomic energy is often used to study nuclear reactions and energy, as well as its applications in fields such as emission footprint predictions and nuclear medicine research (2). However, atomic energy calculations are highly time consuming as they involve complex calculations, since subatomic particles have both particle and wave natures, as explained by wave-particle duality (3).

Wave-particle duality delineates that instead of possessing only particle or only wave characteristics, electrons are "clouds" that have both properties. The clouds are represented by density distributions, known as a wavefunction, which describes the probability of an electron being at one position in space around the nucleus at an instant of time. The wavefunction calculations are executed via the Schrödinger equation, which also allows atomic energy to be determined (4).

When computing atomic energies for atoms and molecules, the Schrödinger equation, which has been found to be very accurate, is typically used (5). However, when multiple electrons and protons are present, the electrostatic forces make the Schrödinger equation difficult, if not impossible, to solve exactly (6). Thus, approximation methods are used to obtain a reasonably accurate solution.

The most-used approximations for atomic energies are the Quantum Monte Carlo methods, which randomly sample guesses, and the self-consistent field (SCF) methods, which begin with a guess that is repeatedly converged to a good approximation (7). We decided to use SCF methods as they are simpler while still accurate, specifically the Hartree-Fock method for its increased accuracy due to accounting for electron-electron correlation, or the interaction between multiple electron wavefunctions (8). As opposed to Density Functional Theory (DFT), which is an SCF method that only calculates the density of the electron clouds, methods in the Hartree-Fock family treat each electron as actual wavefunctions that interact and interfere with each other, leading to more accurate calculations of electronic energy.

The basis set of an SCF method is the first guess for the electron wavefunctions. There are two types of basis sets: Gaussian-type orbitals (GTOs) and Slater-type orbitals (STOs). Despite being less accurate, GTOs are faster and can be combined to estimate STOs, creating STO-nG basis sets, where n is an integer representing the size of the set (9, 10). Small values of n are typically used for calculating atomic and molecular energies because of their relatively quick runtime and still high accuracy. Using larger sets increases this accuracy, at the cost of a longer runtime (11). We used the cc-pVTZ GTO basis set, which has a size of three, offering an apt balance between accuracy and runtime.

The ground state energy of molecules allows insight into various other properties of the molecule and can determine

https://doi.org/10.59720/23-130

the feasibility of synthesizing a molecule. Therefore, fast and accurate methods for determining the ground state energy of molecules are called for. We used PySCF, a python library capable of running SCF methods, including our chosen Hartree-Fock method. We aimed to compare the computed energies with accepted values from the Computational Chemistry Comparison and Benchmark Database (CCCBDB), while noting the effect of the number of carbon atoms and total atoms in alkanes on the calculated energies and runtimes of the PySCF program (**Figure 1**) (12).

We hypothesized that the ground state energy would increase linearly as both the number of carbon atoms and total atoms increased because of the linear nature of normal alkanes and the relatively low contribution of atoms on opposite ends of the chain to the total energy. We also hypothesized that the runtime of the program would increase quadratically as both the number of carbon atoms and total atoms increased, because every pair of particles' contribution to the total ground state energy is considered, leading to a runtime correlating with the number of atoms squared. Our results of running the PySCF program show that regressions on both the ground state energy versus the number of atoms and runtime versus the number of atoms had very high coefficients of correlation, supporting our hypotheses.

RESULTS

The complete collection of data including Self-Consistent Field converged molecular energies for the first ten alkanes, total runtime for each trial of each alkane, and their properties and uncertainties, were tabulated (**Tables 1, 2**). Using this data, scatter plots were constructed for energy vs. number of carbon atoms, energy vs. total number of atoms, runtime vs. carbon atoms, and runtime vs. total number of atoms. Regressions were then modeled for these scatter plots.

The scatter plots of energy vs. number of carbon atoms and energy vs. total number of atoms yielded linearly correlated data (**Figures 2, 3**). To quantitatively observe this relation, linear regressions were modeled for both plots. These yielded Pearson correlation coefficients (r) of -1.000, indicating extremely strong fits. It was found that both the total number of atoms and the number of carbon atoms on the x-axis yielded linear scatter plots, corresponding to the linear increase in the number of carbon and total atoms in each molecule, and validating the functionality of PySCF. This is discussed further in the Discussion section.

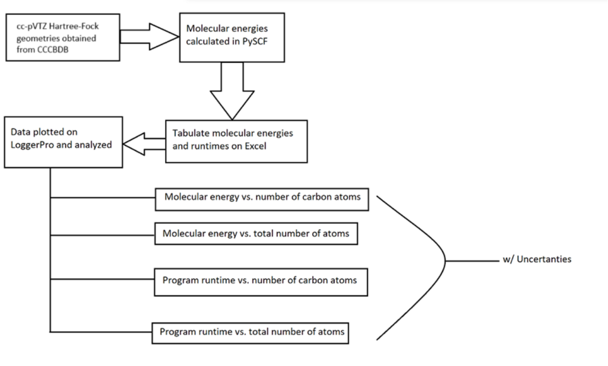


Figure 1. Method flowchart. Steps taken, software used, and databases consulted in the investigation, with specific basis sets and self-consistent field methods mentioned. The last four boxes represent the four graphs and regressions that were plotted.

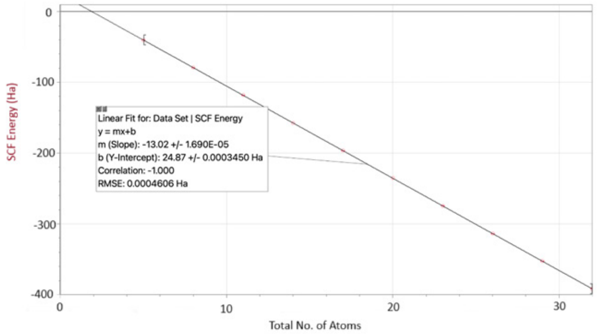


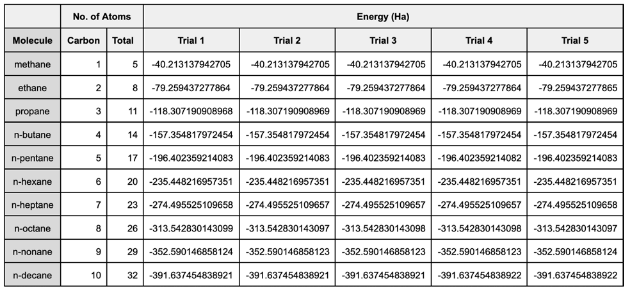
Figure 2. SCF Hartree Fock molecular energy (Hartrees) versus total number of atoms in each alkane. Average PySCF SCF Hartree Fock molecular energy versus total number of atoms. Linear fit with -1.000 correlation value. To view data in more detail, see the repository in Materials and Methods.

The scatter plots of runtime vs. number of carbon atoms and runtime vs. total number of atoms yielded roughly quadratically correlated data (**Figures 4, 5**). To test whether or not a quadratic regression would aptly model these scatter plots, quadratic regressions were compared with higher-order quartic regressions. For both graphs, quadratic regressions yielded correlation coefficients (R^2) of 0.9998, indicating a very high level of correlation of the data. As selecting a higher-order polynomial yielded minimal change in the correlation, quadratic regressions were selected to prevent the risk of polynomial overfitting.

Finally, the calculated PySCF energies were compared to previously computed peer-reviewed energies from the CCCBDB. The computed energies were found to be in agreement with the database energies, only differing by one-thousandth of a percent, with the highest error being 0.001396% in the energy calculation of n-hexane (**Table 3**).

DISCUSSION

Our investigation focused on three aspects: confirming PySCF's functionality, evaluating PySCF's computational accuracy, and analyzing PySCF's efficiency. This was



	No. of Atoms		Energy (Ha)				
Molecule	**Carbon**	**Total**	**Trial 1**	**Trial 2**	**Trial 3**	**Trial 4**	**Trial 5**
methane	1	5	-40.213137942705	-40.213137942705	-40.213137942705	-40.213137942705	-40.213137942705
ethane	2	8	-79.259437277864	-79.259437277864	-79.259437277864	-79.259437277864	-79.259437277865
propane	3	11	-118.307190908968	-118.307190908969	-118.307190908969	-118.307190908969	-118.307190908969
n-butane	4	14	-157.354817972454	-157.354817972454	-157.354817972454	-157.354817972454	-157.354817972454
n-pentane	5	17	-196.402359214083	-196.402359214083	-196.402359214083	-196.402359214082	-196.402359214083
n-hexane	6	20	-235.448216957351	-235.448216957351	-235.448216957351	-235.448216957351	-235.448216957351
n-heptane	7	23	-274.495525109658	-274.495525109657	-274.495525109657	-274.495525109657	-274.495525109658
n-octane	8	26	-313.542830143099	-313.542830143097	-313.542830143098	-313.542830143098	-313.542830143097
n-nonane	9	29	-352.590146858124	-352.590146858123	-352.590146858123	-352.590146858123	-352.590146858124
n-decane	10	32	-391.637454838921	-391.637454838921	-391.637454838921	-391.637454838922	-391.637454838922

Table 1. Energy vs. number of carbons/total atoms table. The energy has been measured in Hartrees. The energy is reported as negative as it corresponds to the potential energy of the electrons, as per convention.

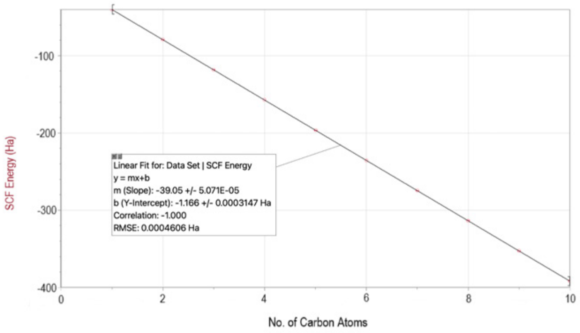


Figure 3. SCF Hartree Fock molecular energy (Hartrees) versus number of carbon atoms in each alkane. Average PySCF SCF Hartree Fock molecular energy versus number of carbon atoms. Linear fit with -1.000 correlation value - compare with total number of atoms in Figure 5. To view data in more detail, *see the repository in Materials and Methods.*

done by plotting the graph of energy vs. total number of atoms and carbon atoms, comparing the PySCF data with accepted values from the CCCBDB, and observing the runtime changes of the algorithm when the number of carbon atoms and the total number of atoms were increased in the compound. This was to ultimately understand if PySCF was a suitable candidate for running SCF calculations for larger molecules, such as organic biochemical and pharmaceutical compounds.

As indicated by the graph, the data for SCF energy versus the number of carbon atoms and SCF energy versus the total number of atoms form a linear correlation, thus supporting our hypothesis (**Figures 2, 3**). The Hartree-Fock method is size-extensive, meaning that as more electrons are present in the molecule, the computed energy should scale linearly. This size-extensivity can explain the linear trend observed in these graphs. This supports our hypothesis that the ground state energy would increase linearly due to the linear nature of normal alkanes and the low contribution of atoms on opposite ends of the chain to the total energy.

	Number of Atoms		Time (s)					
Molecule	**Carbon**	**Total**	**Trial 1**	**Trial 2**	**Trial 3**	**Trial 4**	**Trial 5**	**Average**
methane	1	5	27.49	24.12	25.72	25.1	20.57	24.60
ethane	2	8	40.58	49.96	49.32	50.89	48.91	47.93
propane	3	11	96.99	121.32	112.78	119.74	123.85	114.94
n-butane	4	14	586.09	578.42	575.56	574.86	573.27	577.64
n-pentane	5	17	1070.88	1044.53	1048.07	1045.96	1044.69	1050.83
n-hexane	6	20	1717.29	1669.33	1672.2	1672.35	1673.56	1680.95
n-heptane	7	23	2501.34	2456.85	2456.69	2458.27	2455.45	2465.72
n-octane	8	26	3434.33	3376.31	3378.68	3371.81	3381.02	3388.43
n-nonane	9	29	4548.78	4462.23	4465.41	4470.27	4456.69	4480.68
n-decane	10	32	5788.58	5680.9	5666.52	5663.37	5656.69	5691.21

Table 2. Time vs. number of carbons/total atoms. The time refers to the CPU time required to conduct the calculation, as per PySCF.

	Average Computed Energy Levels (Ha)		Uncertainties		Percent Error with CCCBDB Value
Molecule	**PySCF**	**HF/cc-pVTZ**	**E (10^{-13}·Ha)**	**t (s)**	
methane	-40.213138	-40.213466	1	3.46	0.000816%
ethane	-79.259437	-79.260035	2	5.16	0.000754%
propane	-118.307191	-118.308037	5	13.43	0.000715%
n-butane	-157.354818	-157.355913	0	6.41	0.000696%
n-pentane	-196.402359	-196.403707	5	13.18	0.000686%
n-hexane	-235.448217	-235.451504	0	23.98	0.001396%
n-heptane	-274.495525	-274.499298	5	22.95	0.001374%
n-octane	-313.542830	-313.547092	1	31.26	0.001359%
n-nonane	-352.590147	-352.594886	5	46.05	0.001344%
n-decane	-391.637455	-391.642679	5	65.95	0.001334%

Table 3. Uncertainties, and comparison between PySCF calculated data and CCCBDB HF/cc-pVTZ values. Molecular energies (in Hartrees) for each alkane from PySCF and from accepted Computational Chemistry Comparison and Benchmark DataBase (CCCBDB) HF/cc-pVTZ levels, with percent error. Maximum degree of precision–6 decimal places. Energy and time uncertainties also listed in this table to demonstrate precision of PySCF.

Because alkanes follow the general formula of C_nH_{2n+2}, as the number of carbon atoms increases by one, the molecular energy of the alkane increases by the energy offered by CH_2. This pattern continues as the value of n increases; therefore, it can be inferred that the molecular energy of CH_2 is around -39.05 Hartrees, as this is the value by which the SCF energy changes with the change from one alkane to the next. As previously described, the trends seemed to be the same whether the total number of atoms are considered or just the number of carbon atoms. This may be because of the previously mentioned property of alkanes where increasing the number of carbons increases the total number of atoms by three, which does not affect the overall trend of runtime or molecular energy.

We found the PySCF program to be highly accurate and well-suited for many common applications of ground-state energy calculations by comparing it with the peer-reviewed calculated energies from the CCCBDB. In general, as the size of the molecule increased, the magnitude of the error increased, with methane to n-pentane having differences under 0.0010%, and n-hexane to n-decane having differences under 0.0015%. However, this did not follow a clear pattern, as the minimum magnitude of error belonged to n-pentane (0.000686%), while the maximum magnitude of error belonged to the subsequent molecule n-hexane (0.001396%). These random errors can be attributed to the fluctuations in CPU processing caused by different background programs. Future investigations could carry out more trials for each tested alkane to validate or disprove this hypothesis.

Nonetheless, the errors of the PySCF calculations when compared to the CCCBDB data were negligible for the molecules tested. The convenience of integration with Python and the efficiency and high accuracy of PySCF for GTO Hartree-Fock calculations all suggest that the PySCF library is a valuable tool for molecular energy calculations. Following our validation of this technique in the context of simple compounds, future research is needed to investigate the runtimes of larger, more complex compounds.

To analyze the efficiency of the algorithm, the change in CPU runtime with respect to an increase in carbon and total atoms in the alkane were tested. The relationship between the total runtime of the program and the increase in the number of

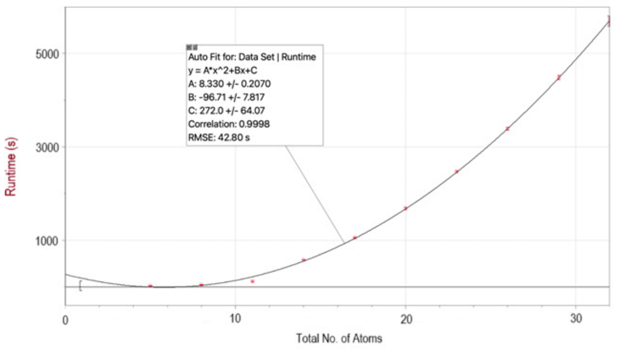


Figure 4. Program runtime using PySCF to calculate SCF HF molecular energy versus total number of atoms in each alkane. Average PySCF molecular energy program runtime versus total number of atoms. Quadratic fit with 0.9998 correlation value. To view data in more detail, see the repository in Materials and Methods.

atoms was modelled using quadratic and quartic regressions based on the shape of the general spread of data, where both regressions yielded high correlation coefficient values of 0.9998 (**Figures 4, 5**). This implies that these results may follow an O(n^4) time complexity, demonstrating the nature by which the basis sets were processed by the Python program. Because the R^2 correlation values of both graphs were 0.9998, we can infer that the quadratic and quartic regressions were a precise but not exact fit.

A potential reason for the anomaly or uncertainty in the data may have been the small changes in CPU background processes, which would have affected the memory allocation for the Python program. The greater the memory allocation for the program, the greater the amount of computational processing power available for use by the program. Additionally, in SCF methods, for every new electron accounted for, a new set of iterations between the fields in which the electron is placed must be added. The recursive nature of self-consistent field methods may have caused the quadratic runtime complexity. Future investigations can explore with more detail and profundity the aforementioned reasons or other reasons as to why the computations follow an O(n^4) time complexity and how the runtime complexity can be made more efficient. These investigations will test larger organic compounds with oxygen and nitrogen to provide a more holistic understanding of the efficiency of PySCF when the size of the molecules is increased.

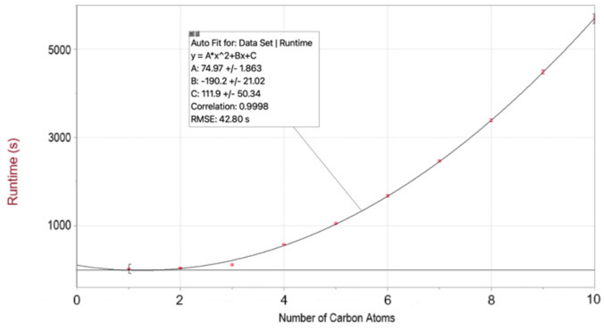


Figure 5. Program runtime using PySCF to calculate SCF HF molecular energy versus number of carbon atoms in each alkane. Average PySCF molecular energy program runtime versus number of carbon atoms. Quadratic fit with 0.9998 correlation value – compare with total number of atoms in Figure 4. To view data in more detail, see the repository in Materials and Methods.

Our hypothesis that the ground state SCF energy would increase linearly as the number of carbon atoms and the number of total atoms increased was ultimately supported by the data collected. Additionally, our prediction that the program runtime would follow a quadratic time complexity as the input (the number of carbon atoms or total atoms) increases was also supported by our data and regression modeling. However, it is of utmost importance to examine the more specific reasons behind the quadratic time complexity of the program and the ways in which the efficiency and speed of the program can be enhanced as larger molecules are tested. Our future work will involve testing longer alkanes beyond decane, and other molecules such as alkenes, alkynes, and oxo compounds, for both straight-chain and branched-chain structures. Analysis of SCF energies and runtimes for more complex molecules would enable the refinement and validation of the accuracy of our regression modeling approaches.

Additionally, other data collection methods will be implemented in our future work. Our current method of implementing Hartree Fock calculations does not account for correlation energy, which is the extent to which electron movement in a molecule varies depending on the presence of other electrons. As the Hartree Fock wavefunction does not take this into consideration, the estimated calculations result in higher energies than the actual values, as interelectronic repulsion decreases the energy of the molecule, though the computational cost and time are reduced with our current method. Instead of using Hartree Fock methods, Post-Hartree Fock methods, such as second-order Møller-Plesset perturbation theory, configuration interaction, and coupled cluster correlation energy, could be used to improve accuracy. These three theories would consider interelectronic repulsion in different ways, thus allowing the data collected to be more accurate, though the runtime would increase as a result.

Overall, three main inferences can be extracted from the collected data. First, PySCF was identified as a suitable candidate for general molecular calculations, as it was found to be highly accurate when compared to prior peer-reviewed data from CCCBDB. Moreover, while Python algorithms tend to be slower than C++ algorithms, PySCF runs on a C++ backbone, ensuring that CPU program runtimes for hydrocarbons up to ten carbon atoms are within 100 minutes. Second, the graphs of energy versus number of atoms also supported our hypothesis that the Hartree-Fock energy corresponded linearly with the number of carbons and total atoms. Finally, the efficiency of PySCF as the number of atoms were increased was analyzed by modeling a regression for CPU runtime vs. total number of carbons and total atoms. The Hartree-Fock PySCF algorithm was found to roughly follow O(n^2) time. Future research will aim to test larger molecules with PySCF, obtain more data to see if the O(n^2) runtime trend is supported by non-hydrocarbon molecules, and move to more accurate self-consistent field techniques such as post-Hartree-Fock Methods. In the future, we aim to use PySCF to calculate electronic, vibrational, and rotational data for

https://doi.org/10.59720/23-130

larger hypothesized drug-candidate molecules to aid in drug stability analyses and pharmaceutical synthesis.

MATERIALS AND METHODS

Schrödinger Equation and Hartree-Fock Method

The monoelectronic Schrödinger equation, where the calculation of the Hamiltonian (the operator of the total energy of the molecule) is independent of time, can be written as an eigenvalue equation (Equation 1):

$$\hat{H}\psi = E\psi \quad \text{(Eqn.1)}$$

$\hat{H}$ represents the Hamiltonian of the atom, Ψ represents the wavefunction of the electron, and E represents the energy of the electron.

The Hartree-Fock method acts as an approximating computational extension to the Schrödinger equation for multielectron atoms and molecules. First, the many-electron wavefunction is approximated to be a product of the orbital wavefunction of each electron in the multielectron system (Equation 2) (13):

$$\psi(r_i, r_j) = \phi(r_i)\phi(r_j) \quad \text{(Eqn. 2)}$$

For a system with two electrons i and j, the multielectron wavefunction Ψ is the product of the orbital wavefunctions Φ of each electron.

However, this allows for the existence of two electrons of the same energy level and angular momentum in an atom, which is not possible as per the Pauli exclusion principle (14). Thus, the spin states of the electrons must also be considered, which is done by calculating the Slater determinant of the many-electron wavefunction approximation. This allows for the calculation of the electron energies and the multielectron wavefunction in terms of the spin orbitals of the electrons (15).

The spin orbitals are then calculated through the variational method, where the wavefunction of electron i is first calculated independently, then the wavefunction for j is calculated using the field of i as the average field. The process is repeated, switching between the wavefunctions of i and j until the ground state electron energy has been minimized. This calculation can be conducted for any number of electrons.

After the spin orbitals have been determined, they can be substituted into the Hartree-Fock equation to solve for the energy of one electron spin orbital (Equation 3):

$$f_i\psi_i = \varepsilon_i\psi_i \quad \text{(Eqn. 3)}$$

This equation is the analog to the Schrödinger equation for a single electron spin orbital, where Ψ_i represents the wavefunction of electron i, ε_i represents the energy of the electron, and fi represents the Fock operator for the electron, which is the analog to the Hamiltonian but including the energy of interelectronic Coulombic repulsion.

This can then be converted into a matrix equation calculable for the molecular energy, through the Roothaan-Hall equation, an analog of the Schrödinger equation for the Hartree-Fock method (Equation 4) (16):

$$FC = SC\varepsilon \quad \text{(Eqn. 4)}$$

F represents the Fock matrix, the sum of $\hat{H}$ (the core Hamiltonian matrix), and G (the interelectronic Coulombic repulsion matrix). S represents the overlap matrix, calculating the overlap in electron orbitals. C represents the orbital coefficients, which is a linear combination of the calculated spin orbitals of the electrons. ε represents the diagonal energy matrix, which stores the values of the individual ground state energies of the electrons.

Thus, the molecular energy can be calculated as the sum of the nuclear energies and the energy of the electrons, as calculated using the Roothaan-Hall equation. This calculation can be repeated recursively, using the results of a calculation as the basis for another Hartree-Fock calculation, until the molecular energy converges for a certain number of decimal digits. Thus, Hartree-Fock, along with other similar recursive methods, are called the Self-Consistent Field (SCF) methods.

Basis Sets

The radial component of a GTO function can be calculated using the radius between the nuclear center and the electron, the effective nuclear charge (ENC), and the angular momentum quantum number of the electron (Equation 5).

$$R = N_G r^l \exp\left(\zeta r^2\right) \quad \text{(Eqn. 5)}$$

The radial part of the GTO, R, can be calculated as a function of the electron-nucleus distance r, the angular momentum l, and a constant related to the ENC, ζ. N_G is the normalizing constant for the GTO function (17).

Slater-type orbitals, or STOs, are more accurate than GTOs, but require larger runtimes and take longer to calculate due to the normalizing constant, represented by N_S (18). The radial component of an STO function can be calculated in a similar fashion (Equation 6).

$$R = N_S r^l \exp\left(\zeta r\right) \quad \text{(Eqn. 6)}$$

PySCF and Logger Pro 3

PySCF is a peer-reviewed Python library with C optimizations facilitating Hartree-Fock and other SCF calculations using GTO basis sets. The PySCF library functioned as the foundation of our research algorithms (19). Molecular energy states were calculated using PySCF using the inherent cc-pVTZ basis set (20). A general program for a set of elements and an STO-nG basis set is structured as follows. The Hartree-Fock function is defined, taking in inputs of the molecular geometry from the CCCBDB database, basis set, and number of unpaired electrons. Then, the Hartree-Fock energy of the molecule is calculated via PySCF, and the details of the process are saved to a text file. The computed SCF energy and the SCF CPU runtime are extracted from that text file and tabulated in Excel, along with the number of carbon atoms in the alkane, and the total number of atoms in the alkane.

These Excel tables were then plotted on Logger Pro 3, a scientific data collection and analysis software designed by Vernier Science Education (21). Four graphs were made: computed SCF energy vs. number of carbon atoms, computed SCF energy vs. total number of atoms, SCF CPU runtime vs. number of carbon atoms, and SCF CPU runtime vs. total number of atoms. These were then analyzed, with lines of best fit created for the computed SCF energy graphs, and quadratic regressions modeled for the SCF CPU runtime graphs, as discussed above in the experimental report (**Figure 5**).

The uncertainty values for both molecular energy and program runtime were calculated by finding the absolute value of the difference between the maximum and minimum values from the five trials and dividing by two. Then these uncertainty values were rounded to the same decimal place as the raw data values. These can be found in the repository in the Materials and Methods section.

To run the computations, Python 2.7.18 was used on a system with 32 GB of RAM, 2 Intel Xeon x5675 CPUs, running Linux 4.15.0-210-generic. The relevant python libraries were PySCF (v.2.2.1), SciPy (v.1.10.1), and NumPy (v.1.24.3). The multithreaded nature of the system sometimes leads to slight variations in the SCF convergence calculations, leading to small variations in the calculated SCF energy. As such, to account for any non-deterministic results, five trials were conducted for each compound.

The CCCBDB database was used to obtain the geometries for the alkanes that we tested. The Computational Chemistry Comparison and Benchmark DataBase is a peer-reviewed standard reference database established by the United States National Institute of Standards and Technology (NIST). The geometries for the compounds we tested were optimized using Hartree-Fock using the cc-pVTZ results. The HF/cc-pVTZ geometries were chosen to agree with the method and basis set we used to calculate the energies, ensuring that we would get the most accurate energy calculations for that specific method and basis set.

Repository

The repository with the Logger Pro files, the tabulation of the data, and the original code is linked here: https://github.com/ZarseemDyartes/PySCF-Molecular-Calculations

ACKNOWLEDGEMENTS

We would like to thank the Aspiring Scholars Directed Research Program, which provided us with the opportunities, resources, and funding to pursue this investigation. We would also like to thank Tanvi Goyal, Ajeeth Iyer, İpek Gerz – the researchers who laid a foundation for our investigation.

Received: April 6, 2023
Accepted: October 9, 2023
Published: March 6, 2024

REFERENCES

1. Asimov, Isaac. *Atom: Journey Across the Subatomic Cosmos*, Plume, 1992.
2. Bartlett, Rodney J., and John F. Stanton. "Applications of Post—Hartree—Fock Methods: A Tutorial." *Reviews in Computational Chemistry*, edited by K.B. Lipkowitz and D.B. Boyd, vol. 5, VCH Publishers, 1994, pp. 65-169.
3. de Broglie, Louis. "Waves and Quanta." *Nature,* vol. 112, 13 Oct. 1923, p. 540. https://doi.org/10.1038/112540a0.
4. Schrödinger, Erwin. "An undulatory theory of the mechanics of atoms and molecules." *Physical Review,* vol. 28, no. 6, Dec. 1926, pp. 1049 - 1070. https://doi.org/10.1103/PhysRev.28.1049.
5. Drake, Gordon. "High precision calculations for helium." *Springer Handbook of Atomic, Molecular, and Optical Physics*, edited by G. Drake, Springer, 2006, pp. 199-219.
6. Hanson, D.M. et al. "6.7: The Helium Atom Cannot Be Solved exactly". *Chemistry LibreTexts*, 2020.
7. Hammond, Brian L., William A. Lester, and Peter James Reynolds. *Monte Carlo Methods in* Ab Initio *Quantum Chemistry*, vol. 1, World Scientific, 1994.
8. David, Grégoire, et al. "Self-Consistent Field Methods for Excited States in Strong Magnetic Fields: a Comparison between Energy-and Variance-Based Approaches." *Journal of Chemical Theory and Computation*, vol 17, no. 9, 2021, pp. 5492-5508. https://doi.org/10.1021/acs.jctc.1c00236.
9. Magalhaes, Alexandre L. "Gaussian-type orbitals versus Slater-type orbitals: a comparison." *Journal of Chemical Education*, vol. 91, no. 12, 2014, pp. 2124-2127. https://doi.org/10.1021/ed500437a.
10. Hehre, Warren J., Robert Ditchfield, and John A. Pople. "Self—consistent molecular orbital methods. XII. Further extensions of Gaussian—type basis sets for use in molecular orbital studies of organic molecules." *The Journal of Chemical Physics,* vol. 56, no.5, 1972, pp. 2257-2261. https://doi.org/10.1063/1.1677527.
11. Pietro, William J., et al. "Molecular orbital theory of the properties of inorganic and organometallic compounds. 2. STO-NG basis sets for fourth-row main-group elements." *Inorganic Chemistry,* vol. 20, no. 11, 1981, pp. 3650-3654. https://doi.org/10.1021/ic50225a013.
12. NIST Computational Chemistry Comparison and Benchmark Database, NIST Standard Reference Database Number 101, Release 22, May 2022, edited by Russell D. Johnson III. https:// doi.org/10.18434/t47c7z.
13. Hartree, Douglas R. "The wave mechanics of an atom with a non-Coulomb central field. Part I. Theory and methods." *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 24, no. 1. Cambridge University Press, 1928. https://doi.org/10.1017/S0305004100011919.
14. Pauli, Wolfgang. "Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren." *Zeitschrift für Physik*, vol. 31, 1925, pp. 765-783.
15. Fock, Vladimir. "Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems." *Zeitschrift für Physik*, vol. 61, no. 1, 1930, pp. 126-148.
16. Hall, G. G. "The molecular orbital theory of chemical valency VIII. A method of calculating ionization potentials." *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, vol. 205, no. 1083, 1951, pp. 541-552. https://doi.org/10.1098/rspa.1951.0048.
17. Gill, Peter MW. "Molecular integrals over Gaussian basis functions." *Advances in Quantum Chemistry*, vol. 25. Academic Press, 1994, pp. 141-205.
18. Slater, John C. "Atomic shielding constants." *Physical*

https://doi.org/10.59720/23-130

Review, vol. 36, no. 1, 1930, p. 57. https://doi.org/10.1103/PhysRev.36.57

19. Sun, Qiming, et al. "PySCF: the Python-based simulations of chemistry framework." *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 8, no. 1, 2018, e1340. https://doi.org/10.1098/rspa.1951.0048.
20. Sun, Qiming, et al. "Recent developments in the PySCF program package." *The Journal of Chemical Physics*, vol. 153, no. 024109, 2020. https://doi.org/10.1063/5.0006074.
21. Vernier. Vernier Computer Interface Logger Pro. 18 January 2021.