



# Automated Of Molecular From First

*A Hartree–Fock Analysis of*

Alex

# Computation Properties Principles



*Small Organic Compounds*

Lin

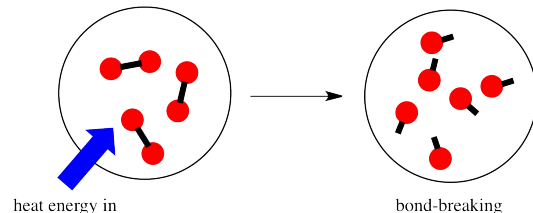
# Motivation

$$\hat{H} = -\sum_A \frac{1}{2M_A} \nabla_A^2 - \sum_i \frac{1}{2} \nabla_i^2 + \sum_{A>B} \frac{Z_A Z_B}{R_{AB}} - \sum_{Ai} \frac{Z_A}{r_{Ai}} + \sum_{i>j} \frac{1}{r_{ij}}$$

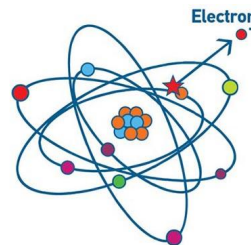
## Problem Statement

- Solving the Schrödinger equation allows us to...
  - Find many interesting molecular properties
  - Benchmark experiments with *ab initio* calculations
- But...there does not exist an analytical solution

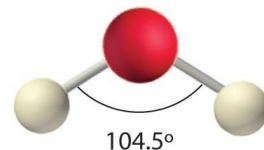
### Atomization Energy



### Ionization



### Bond Angle



# Methods: Intro to HF

## Hartree-Fock Algorithm

Approximates solutions to Schrödinger.

- Born-Oppenheimer approximation
- Wavefunction is a Slater determinant of molecular orbitals (MOs)
- Variational approach: MOs are linear combinations of atomic orbital (LCAO) basis functions (typically Gaussian)
- Mean field - e.g. no London dispersion

**Ignored** **Fixed**

$$\hat{H} = \cancel{\hat{T}_N(\mathbf{R})} + \hat{T}_e(\mathbf{r}) + \hat{V}_{NN}(\mathbf{R}) + \hat{V}_{eN}(\mathbf{r}, \mathbf{R}) + \hat{V}_{ee}(\mathbf{r})$$

$$h(i) = -\frac{1}{2}\nabla_i^2 - \sum_A \frac{Z_A}{r_{iA}} \quad v(i, j) = \frac{1}{r_{ij}}$$

$$\hat{H}_{el} = \sum_i h(i) + \sum_{i < j} v(i, j)$$

$$\Psi = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(\mathbf{x}_1) & \chi_2(\mathbf{x}_1) & \cdots & \chi_N(\mathbf{x}_1) \\ \chi_1(\mathbf{x}_2) & \chi_2(\mathbf{x}_2) & \cdots & \chi_N(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(\mathbf{x}_N) & \chi_2(\mathbf{x}_N) & \cdots & \chi_N(\mathbf{x}_N) \end{vmatrix}$$

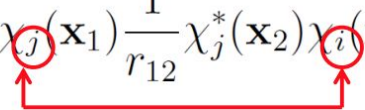
# Methods: HF Energy 1

$$E_{HF} = \sum_i^{\text{elec}} \langle i | \hat{h} | i \rangle + \sum_{i>j}^{\text{elec}} [ii|jj] - [ij|ji]$$

$$\langle i | \hat{h} | i \rangle = \int d\mathbf{x}_1 \chi_i^*(\mathbf{x}_1) \hat{h}(\mathbf{r}_1) \chi_i(\mathbf{x}_1)$$

Kinetic Energy + Nuclear Attraction  
(for every electron)

$$[ii|jj] = \int d\mathbf{x}_1 \int d\mathbf{x}_2 \chi_i^*(\mathbf{x}_1) \chi_i(\mathbf{x}_1) \frac{1}{r_{12}} \chi_j^*(\mathbf{x}_2) \chi_j(\mathbf{x}_2)$$

$$[ij|ji] = \int d\mathbf{x}_1 \int d\mathbf{x}_2 \chi_i^*(\mathbf{x}_1) \chi_j(\mathbf{x}_1) \frac{1}{r_{12}} \chi_j^*(\mathbf{x}_2) \chi_i(\mathbf{x}_2)$$


Coulomb Integral – Exchange Integral  
(for every pair of electrons)

# Methods: HF Energy 2

Minimization of HF Energy leads to complicated integral problem...

$$h(\mathbf{x}_1)\chi_i(\mathbf{x}_1) + \sum_{j \neq i} \left[ \int d\mathbf{x}_2 |\chi_j(\mathbf{x}_2)|^2 r_{12}^{-1} \right] \chi_i(\mathbf{x}_1) \\ - \sum_{j \neq i} \left[ \int d\mathbf{x}_2 \chi_j^*(\mathbf{x}_2) \chi_i(\mathbf{x}_2) r_{12}^{-1} \right] \chi_j(\mathbf{x}_1) = \sum_j \epsilon_{ij} \chi_j(\mathbf{x}_1)$$

Fock Operator

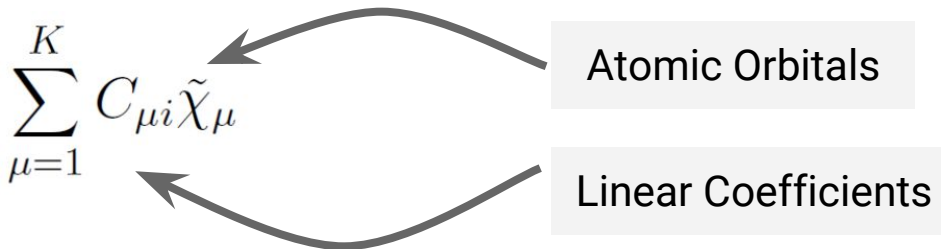
$$f(\mathbf{x}_1)\chi_i(\mathbf{x}_1) = \epsilon_i \chi_i(\mathbf{x}_1)$$

Orbital Energy



# Methods: HF Roothan Eqns

Introduce a “basis set” of atomic orbitals -- easy-to-integrate Gaussians


$$\chi_i = \sum_{\mu=1}^K C_{\mu i} \tilde{\chi}_{\mu}$$
$$\sum_{\nu} F_{\mu\nu} C_{\nu i} = \epsilon_i \sum_{\nu} S_{\mu\nu} C_{\nu i}$$
$$S_{\mu\nu} = \int d\mathbf{x}_1 \tilde{\chi}_{\mu}^*(\mathbf{x}_1) \tilde{\chi}_{\nu}(\mathbf{x}_1),$$
$$F_{\mu\nu} = \int d\mathbf{x}_1 \tilde{\chi}_{\mu}^*(\mathbf{x}_1) f(\mathbf{x}_1) \tilde{\chi}_{\nu}(\mathbf{x}_1)$$
$$\mathbf{FC} = \mathbf{SC}\epsilon$$

**Note:** But F depends on C and C depends on F - iterate until convergence

# Methods: HF Algorithm

1. Input nuclei, atomic charges, basis functions, # of electrons
2. Calculate nucleus-nucleus repulsion ( $V_{NN}$ )
3. Calculate four electron integrals - overlap ( $S$ ), kinetic energy ( $T$ ), nuclear attraction ( $V_{eN}$ ), two electron repulsion ( $V_{ee}$ )
4. Guess MO coefficients  $C$
5. Repeat until convergence
  - a. Form  $F$  from  $C, S, T, V_{eN}, V_{ee}$
  - b. Solve  $FC = SC\epsilon$  for MO coefficients  $C$  and MO energies  $\epsilon$



# Methods: HF Integration

Q: How do we find  $\mathbf{S}$ ,  $\mathbf{T}$ ,  $\mathbf{V}_{eN}$ ,  $\mathbf{V}_{ee}$ ?

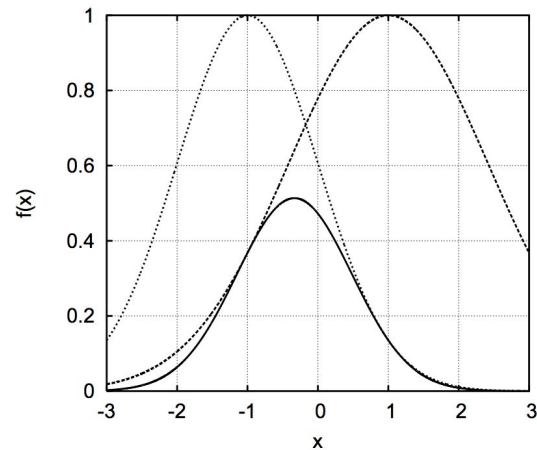
A: If atomic orbitals are Gaussian, then there exist *recurrence relations*.  
We use the **McMurchie-Davidson scheme** in our code.

Q: Why do the schemes work?

A: Product of two Gaussians = another Gaussian!

Q: How do the schemes work?

A: Complicated math -- see [3]! Feel free to ask :)



# Methods: HF Speed

Rate-determining step -- calculating two electron repulsion  $\mathbf{V}_{ee}$ .

$$f_{\mu\nu} = h_{\mu\nu} + \sum_i^{N/2} \sum_{\lambda\sigma}^K C_{\lambda i}^* C_{\sigma i} (2[\mu\nu|\lambda\sigma] - [\mu\sigma|\lambda\nu])$$

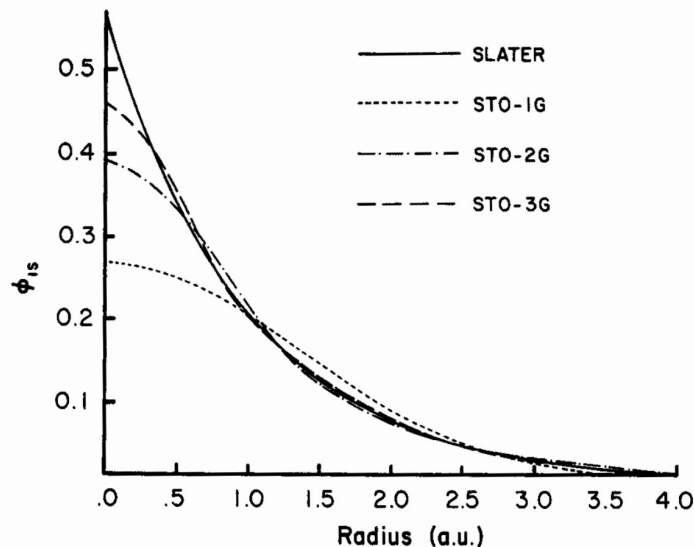
Four Atomic Orbitals  
 $O(n^4)$ -Time (i.e. really slow!)

Can use integral pre-screening to eliminate integrals between orbitals that are far away -- implemented, but insignificant for small molecules.

# Methods: STO-3G Basis

Slater-Type Orbital -- 3 Gaussians per orbital (minimal basis, not best)

Carbon Atom					
-C	0				
S	3	1.00			
		71.6168370	0.15432897		
		13.0450960	0.53532814		
		3.5305122	0.44463454		
SP	3	1.00			
		2.9412494	-0.09996723	0.15591627	
		0.6834831	0.39951283	0.60768372	
		0.2222899	0.70011547	0.39195739	
Exponents		s Coefficients		p Coefficients	

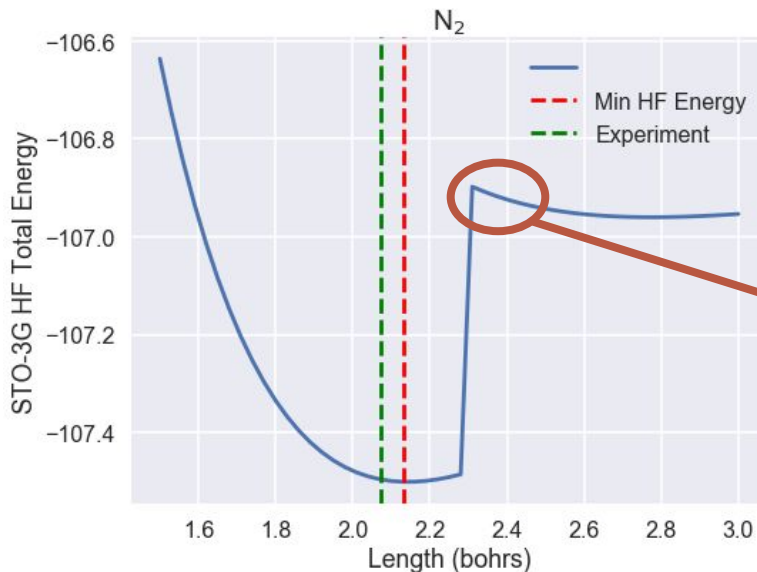
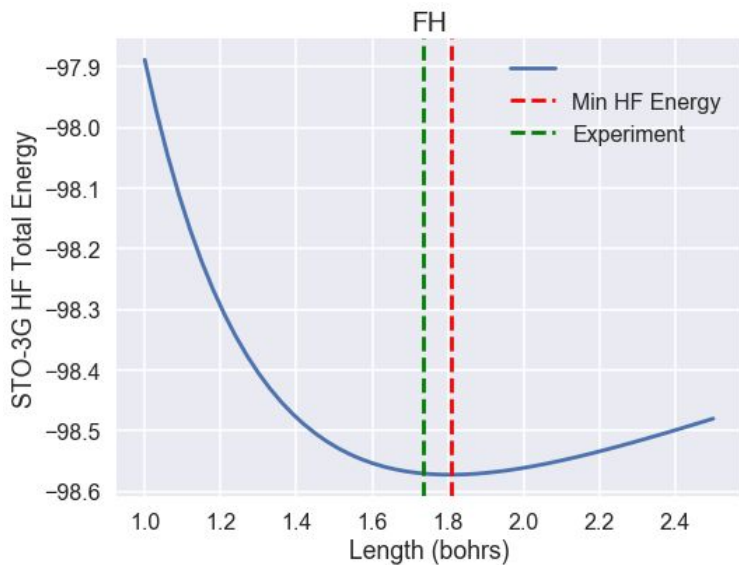


# Results: HF Code

- Implemented **restricted Hartree-Fock** from scratch in Python
  - Assumes all orbitals have 2 electrons or 0 electrons
  - Modularized code in three files -- `utils.py`, `integrals.py`, `hf.py`
- Sample calculations available
- Equilibrium bond length experiments on FH, N<sub>2</sub>
- Multi-electron atomic energy calculations
- Analysis of 59 small organic compounds (up to 4 atoms with C, N, O) -- total energy, ionization energy

# Results: Bond Lengths

Equilibrium bond length calculated as value that minimizes HF energy

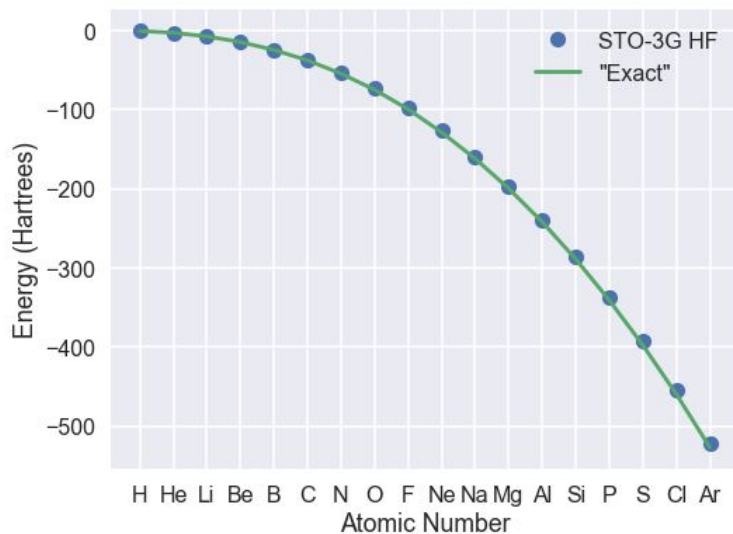


Experiment  
not too far  
from theory  
approx.

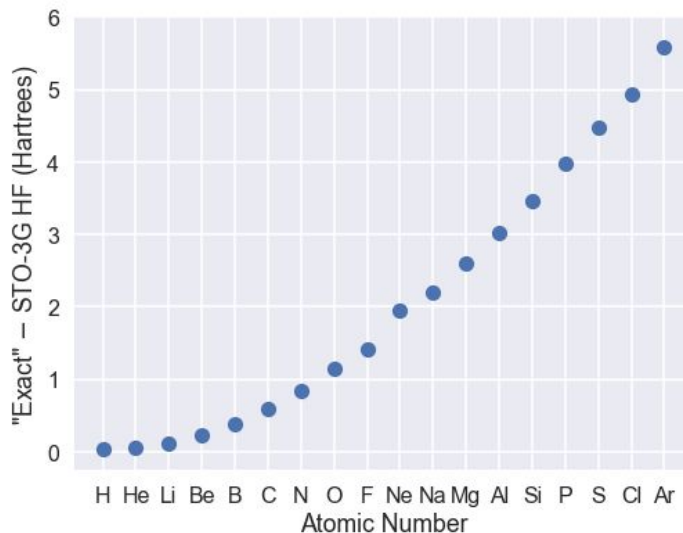
Failed to  
converge to  
global  
minimum.

# Results: Atomic Energy

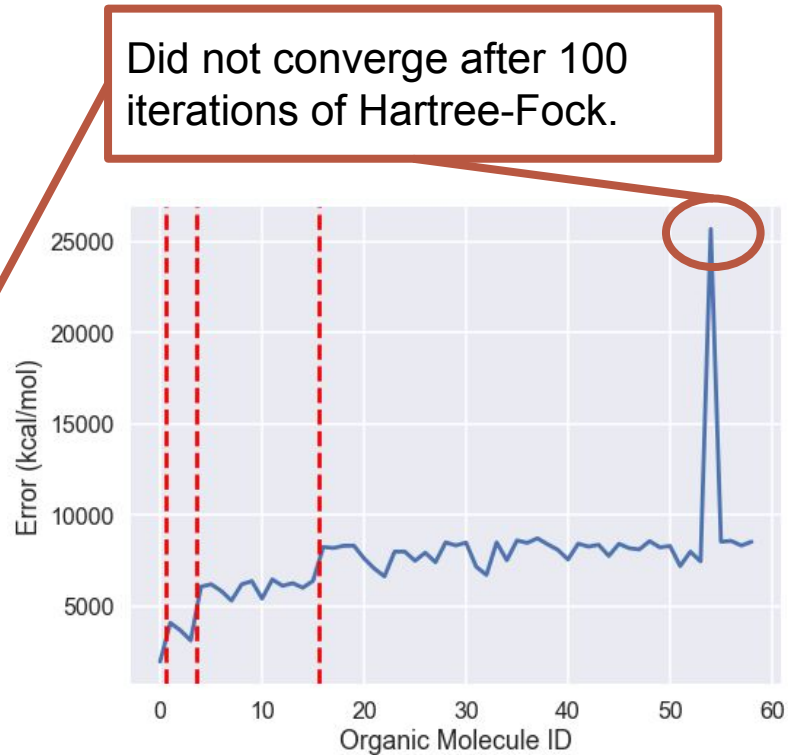
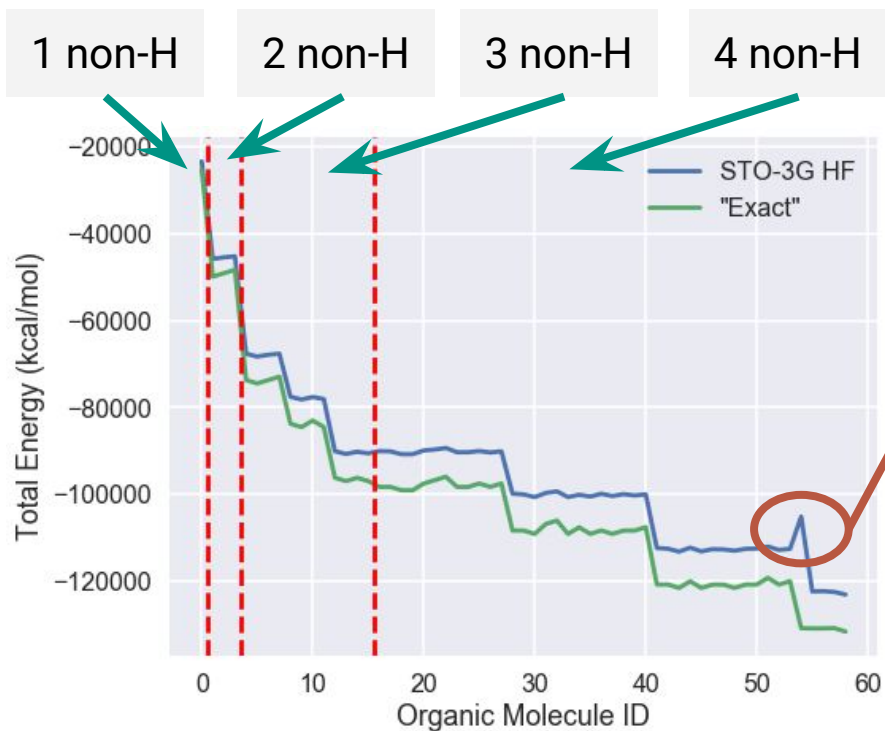
If  $X$  has *even* # of electrons: Use restricted Hartree-Fock to calculate energy of  $X$ .



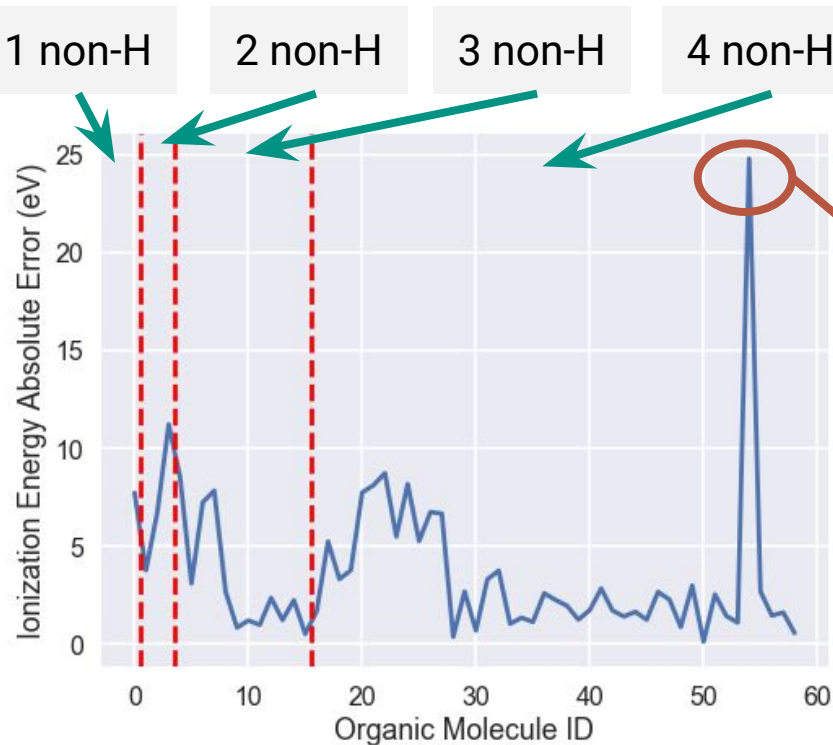
If  $X$  has *odd* # of electrons: 1) Use HF to calculate energy of  $X^-$ , 2) Subtract ionization energy to get energy of  $X$ .



# Results: Orgo Total Energy



# Results: Orgo Ionization



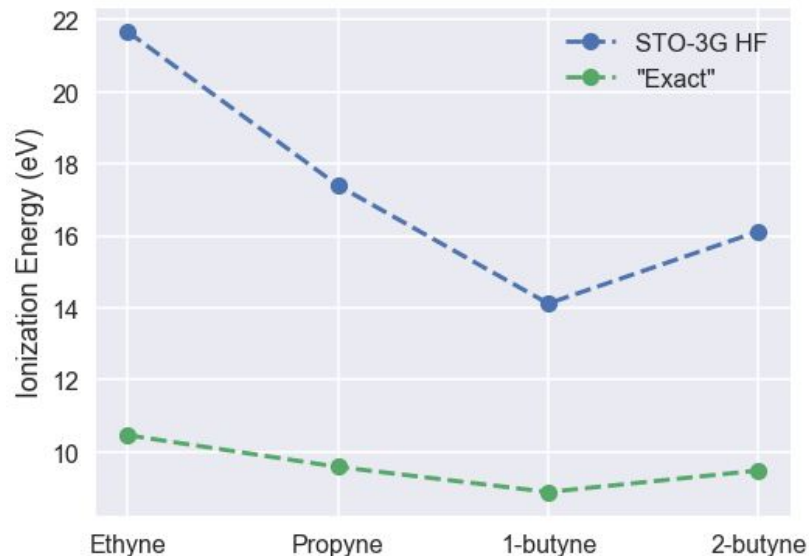
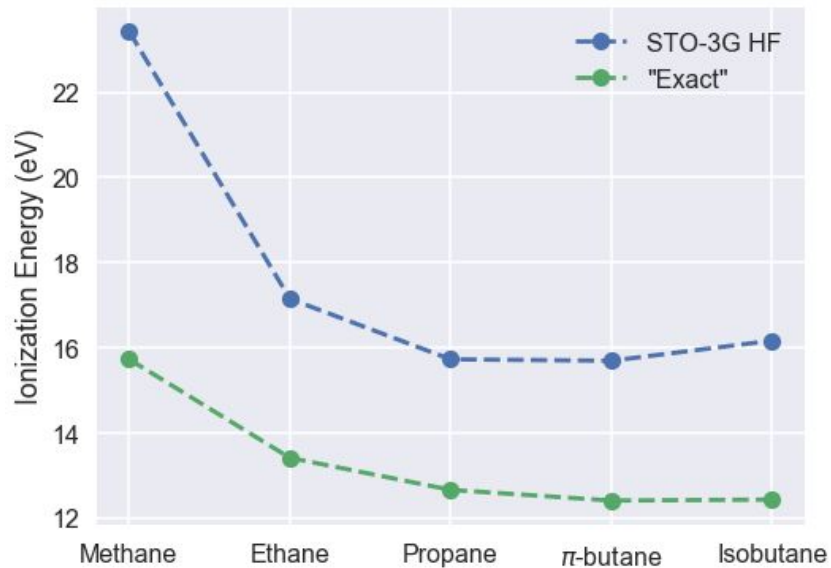
Koopmans' Theorem (1934):  
First ionization energy is equal  
to the negative HOMO energy.

Did not converge after 100  
iterations of Hartree-Fock.

Mean Abs Error: 3.6 eV (36%)  
Median Abs Error: 2.5 eV (24%)

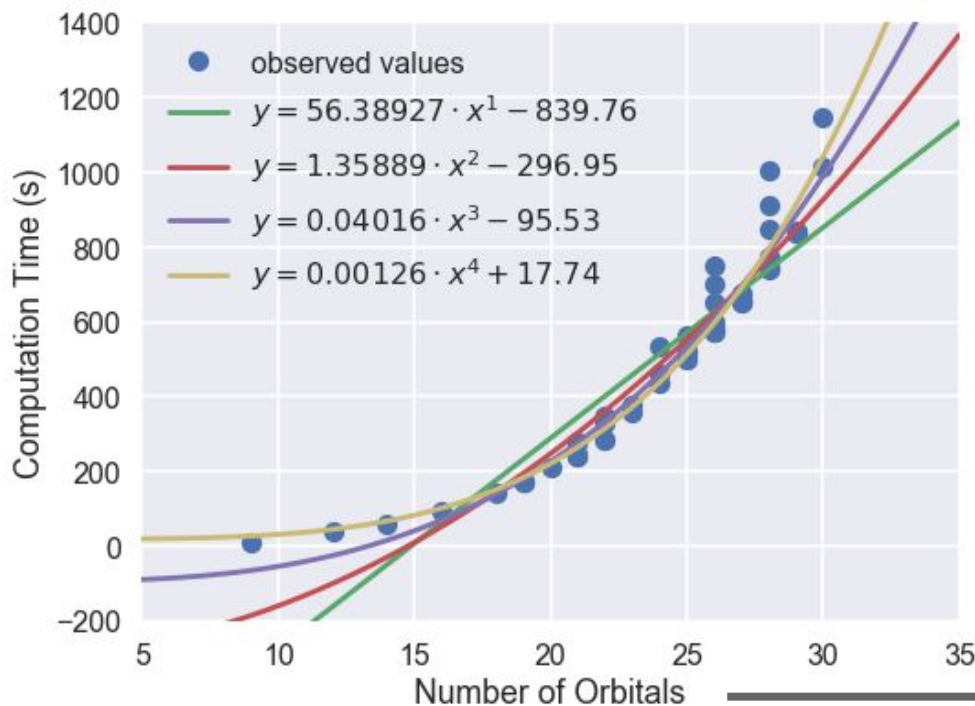


# Results: Orgo Families



Values are off, but maintains general trend as function of # of Cs

# Results: Orgo Comp Time



- Confirms that quartic run time characterizes asymptotic behavior.
- Lower powers underfit.

Equal to  $5 * [\# \text{ of C, N, O}] + [\# \text{ of H}]$

# Conclusion

## Summary

- Hartree-Fock allows for the (successful) calculation of interesting properties for small molecules
- Energy approximations are suboptimal underestimates, but still provide useful info

## Future Directions

- Unrestricted HF for single electron orbitals
- Use more complicated basis sets (e.g. 4-31G, 6-31G\*) for better calculations
- Explore coupled cluster algorithm, which takes electron correlation into account
- Use improved accuracy to examine other properties (e.g. atomization energy)
- See if non-physics-based approaches (e.g. machine learning) hold any promise [4, 5, 6]

# Acknowledgements

Huge thank you to my advisor **Dr. Michael Mavros** for his indispensable support throughout the course of this project.

Special thank you to the **Sherrill group** [7] for their extensive online support on Hartree-Fock and incredibly simple notation, which was heavily used in this poster.

Thank you to the PS 10 teaching staff for all of their hard work this semester.

# References 1

- [1] Blum, Lorenz C., and Jean-Louis Reymond. "970 million druglike small molecules for virtual screening in the chemical universe database GDB-13." *Journal of the American Chemical Society* 131, no. 25 (2009): 8732-8733.
- [2] Feller, David. "The role of databases in support of computational chemistry calculations." *Journal of computational chemistry* 17, no. 13 (1996): 1571-1586.
- [3] Helgaker, Trygve, Poul Jorgensen, and Jeppe Olsen. *Molecular electronic-structure theory*. John Wiley & Sons, 2014.
- [4] Montavon, Grégoire, Katja Hansen, Siamac Fazli, Matthias Rupp, Franziska Biegler, Andreas Ziehe, Alexandre Tkatchenko, Anatole V. Lilienfeld, and Klaus-Robert Müller. "Learning invariant representations of molecules for atomization energy prediction." In *Advances in Neural Information Processing Systems*, pp. 440-448. 2012.
- [5] Montavon, Grégoire, Matthias Rupp, Vivekanand Gobre, Alvaro Vazquez-Mayagoitia, Katja Hansen, Alexandre Tkatchenko, Klaus-Robert Müller, and O. Anatole Von Lilienfeld. "Machine learning of molecular electronic properties in chemical compound space." *New Journal of Physics* 15, no. 9 (2013): 095003.

# References 2

- [6] Rupp, Matthias, Alexandre Tkatchenko, Klaus-Robert Müller, and O. Anatole Von Lilienfeld. "Fast and accurate modeling of molecular atomization energies with machine learning." *Physical review letters* 108, no. 5 (2012): 058301.
- [7] Schuchardt, Karen L., Brett T. Didier, Todd Elsethagen, Lisong Sun, Vidhya Gurumoorthi, Jared Chase, Jun Li, and Theresa L. Windus. "Basis set exchange: a community database for computational sciences." *Journal of chemical information and modeling* 47, no. 3 (2007): 1045-1052.
- [8] Sherrill, C. David. "A Brief Review of Elementary Quantum Chemistry." *Georgia Institute of Technology, School of Chemistry and Biochemistry* (2001).
- [9] Szabo, Attila, and Neil S. Ostlund. *Modern quantum chemistry: introduction to advanced electronic structure theory*. Courier Corporation, 2012.